University of Thessaly
School of Engineering
Department of Mechanical Engineering

PhD Thesis

Bayesian uncertainty quantification and optimal experimental design in data-driven simulations of engineering systems

by

Costas Argyris

Diploma in Mechanical Engineering, University of Thessaly, 2011
M.Sc. in Mechanical Engineering, University of Thessaly, 2013

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

2017

The research done in this thesis was partially implemented under the "ARISTEIA" Action of the "Operational Program Education and Lifelong Learning" and co-funded by the European Social Fund (ESF) and Greek National Resources. Chapter 8 of this thesis was also funded by the Greek National Scholarship Foundation (IKY) within the IKYDA program framework and by the Deutscher Akademischer Austausch Dienst (DAAD, German Academic Exchange Service) within the PPP program.

Approved by the members of the seven-member examination committee :

- **First Examiner (Supervisor)**
  Dr. Costas Papadimitriou
  Professor, Department of Mechanical Engineering
  University of Thessaly

- **Second Examiner**
  Dr. Panagiotis Tsopelas
  Associate Professor, School of Applied Mathematical and Physical Sciences
  National Technical University of Athens (NTUA)

- **Third Examiner**
  Dr. Sotiris Natsiavas
  Professor, Department of Mechanical Engineering
  Aristotle University of Thessaloniki (AUTH)

- **Fourth Examiner**
  Dr. Tasos Stamatelos
  Professor, Department of Mechanical Engineering
  University of Thessaly

- **Fifth Examiner**
  Dr. Herricos Stapountzis
  Associate Professor, Department of Mechanical Engineering
  University of Thessaly

- **Sixth Examiner**
  Dr. Lambros Katafygiotis
  Professor, Department of Civil Engineering
  Hong Kong University of Science and Technology

- **Seventh Examiner**
  Dr. Dimitrios Giagopoulos
  Assistant Professor, Department of Mechanical Engineering
  University of Western Macedonia

Acknowledgments

First of all i would like to thank my advisor, Professor Costas Papadimitriou, for his valuable help and guidance during our collaboration. Also, i would like to thank the rest of the members of the examination committee for spending the time to carefully read my thesis and provide me with their helpful suggestions. Next i would like to thank Professor Tasos Stamatellos and Dr. Saeed Eftekhar Azam for their constant academic (and not only) help and support throughout the years. Above all, i am thankful to my parents, George and Sevasti, and my brother Charis, for their love, support and patience all those years. Finally, i would like to thank my friends Antonia, Luna, Dimitris and Alexandros for helping me in difficult times.

Costas Argyris

Abstract

The objective of this thesis is twofold. First, Bayesian uncertainty quantification, propagation, and model selection theories are outlined and applied to real-world problems, and second, Bayesian optimal experimental design theories are studied from a theoretical point of view and also applied to numerical case-studies. The Bayesian theory for uncertainty quantification, propagation and model selection is first presented and computational challenges are discussed. The Bayesian formulation is applied to three case-studies taken from real-world applications, using experimental data collected from field tests. It is demonstrated how the Bayesian theories can be applied to such real scale problems of model updating and model selection to draw useful conclusions about the systems under consideration. The examined problems include: 1) Model updating and model selection in a hanger of an arch bridge using modal data obtained from acceleration measurements. 2) Model updating of a bridge using a high-fidelity finite element model utilizing experimentally identified modal data. The modal data are extracted from ambient acceleration measurements using a software developed in the context of this thesis. 3) Parameter estimation of non-linear models of seismically isolated bridges using experimentally measured response time histories from the bridge. Next the theory of Information Theoretic - Bayesian optimal experimental design is developed. Two approaches are developed for the estimation of the objective function, the sampling and asymptotic approaches. Novel theoretical contributions are developed in both approaches, providing further insight into the problem of Bayesian optimal experimental design. The theories of optimal experimental design are applied to numerical case-studies using finite element models. The asymptotic approach is applied to design the optimal locations of acceleration sensors on a bridge in order to perform modal identification, while it also demonstrates the theoretical findings in the particular problem. Next a joint study is presented in both Bayesian parameter estimation and Bayesian optimal experimental design, where the newly developed sampling approach is used to find the optimal locations of strain sensors in a plate with crack under static loading.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research context

In the process of simulating the behavior of complex engineering systems, uncertainties arise mainly from the assumptions and compromises that enter into the development of mathematical models of such systems and the applied loads. Such uncertainties lead to significant uncertainties in the predictions made using simulations. Since predictions form the basis for making decisions, the knowledge and management of these uncertainties is very important.

The sources of uncertainties in engineering simulations are modeling uncertainties, loading uncertainties and numerical uncertainties. Modeling uncertainties are related to the inadequacy of the mathematical model to represent a physical system. They arise when modeling the constitutive behavior of materials, the support conditions of structures and their interaction with their environment, the interaction/coupling between substructures (fixity conditions, friction mechanisms, impact phenomena), the geometric variability due to manufacturing/construction processes, the long-term deterioration mechanisms (e.g. semi-empirical models for fatigue and corrosion), etc. The parametric uncertainties, originating from the limited knowledge about the values of the model parameters, are also considered as part of the modeling uncertainties. Loading uncertainties arise from the lack of detailed knowledge about the spatial and temporal variation of the forces (mechanical, thermal, etc) applied to engineering structures. Representative examples of loading uncertainties in structural dynamics include spatial variability of road roughness affecting the dynamics of vehicles, spatial and temporal variability of wind or earthquake-induced excitations on civil engineering structures, turbulent wind loads affecting the design and maintenance of aircrafts and variability of thermal loads affecting the design of a large class of mechanical and aerospace structures. Numerical uncertainties are related to the spatial (e.g. finite element) and temporal (numerical time integration schemes) discretization of the partial differential equations used for simulating the behavior of engineering structures, round-off errors due to computer accuracies, all affecting solution accuracy.

Those uncertainties need to be treated in a systematic manner when trying to update system models with experimental data, in order to obtain robust and reliable models capable of making accurate predictions. The Bayesian statistical framework represents a general, rational and powerful tool for managing uncertainties that is capable of handling the above stated difficulties (Beck and Katafygiotis, 1998; Beck, 2010; Yuen, 2010). Within this framework, probability is interpreted as the degree of plausibility of a hypothesis based on the conditioning information (Cox interpretation of probability), where the hypothesis may refer to the model parameters but also to the model itself. This interpretation makes it possible to extend the application of probability theory to fields where the frequentist interpretation may not be directly intuitive, as it is the case for one-of-a-kind structures, where no ensemble exists, and also in the case of limited data, where classical statistics is of limited applicability. Probability density functions (pdf) are used to describe the variations in the plausibility of the different possible values of the model parameters. Initial knowledge about the parameters is taken into account through the prior pdf which is defined by the researcher. The Bayesian approach updates the prior pdf to the posterior pdf which represents our complete updated knowledge about the parameters in light of the experimental data. In contrast to deterministic methods, probabilistic methods give the answer in terms of a pdf instead of a point estimate, which reflects the uncertainty present in our inference due to the various sources of uncertainty described above.

Probability density functions are also used to model various sources of uncertainty such as model error and measurement noise which are both unknown quantities but enter the formulation nevertheless and need to be taken explicitly into account. In the Bayesian framework, probabilities need not only express the plausibility of a range of parameter values and model uncertainty, but can also be used as measures of how good a model is in general compared with other competing models of the same system, in light of the experimental data. This is known as Bayesian model selection (Beck and Yuen, 2004).

The measured data collected from system component tests or system operation through monitoring can provide valuable information for improving the mathematical models and the probability models of uncertainties of both the system and loads. Incorporating these data-driven updated models in simulations will yield updated or posterior robust predictions, constituting improved and more reliable estimates of the system performance. The Bayesian framework exploits the available measured data and any prior information based on engineering experience, to perform the following tasks:

- Identify and select the most probable mathematical models among a competitive family of mathematical models (linear vs nonlinear models; elastic vs hysteretic models; friction/impact models; correlation structure of a spatially varying quantities such as modulus of elasticity) introduced to represent the behavior of mechanical components.
- Identify probabilistic models that best account in predictions for the mismatch between model-based predictions and measurements, manifested due to the inadequacy/imperfections of the mechanical models used.
- Calibrate the parametric uncertainties involved in mechanical and prediction error models.

- Propagate uncertainties in simulations for updating robust predictions taking into account the validated models and the calibrated uncertainties, as well as rationally weight the effect of one or more highly probable models promoted by the Bayesian methodology.

The second research area of this thesis is concerned with the optimal design of experiments. Model-based optimal experimental design has to do with finding which is the best way to perform an experiment so that a specific purpose is achieved, using a model of the system as a guide. Common purposes include model parameter inference and making predictions using the model. In this thesis model parameter inference was set as our goal. That is, we seek experimental designs which will result in experimental data that are the most useful for identifying the parameters of the model. The model parameters are uncertain, as are the experimental data, since no experiment has taken place at the time of the design. These uncertainties are also treated with the Bayesian framework for uncertainty quantification, by assigning a prior pdf for the parameters and a probabilistic model for the difference between the model predictions and the data, known as the prediction error.

The problem of finding the optimal design is formulated as an optimization problem with respect to the design variables which describe the various possible experimental designs. The objective function is rooted in Information Theory and is the expected information gain (KL-divergence) between the posterior and prior pdfs of the model parameters. This measure quantifies how good a specific design is for model parameter inference and can also be seen as a special case of a more general 'expected utility' framework for optimal experimental design (Lindley, 1956). However, the analytical calculation of the objective function may not be possible even for simple models and therefore two approaches are followed: approximating the integrals with sums using samples (Ryan, 2003; Huan and Marzouk, 2013), and simplifying the integrals by introducing asymptotic approximations (Argyris et al., 2016).

1.2 Organization of this Thesis

The research work of this thesis is organized into 9 chapters as follows:

- Chapter 2 presents the Bayesian theory for uncertainty quantification, propagation and model selection using experimental data. Also the computational challenges and tools required for Bayesian analysis are discussed. Chapters 3, 4 and 5 include applications of the Bayesian theories developed in Chapter 2 in real-world problems with measured experimental data obtained from actual field measurements.

- Chapter 3 is a case-study in a hanger of an arch bridge located in Greece. The objective is to study the tension developed in the hanger and the effect of its supports using measured modal frequencies obtained from acceleration measurements. For that purpose, several model classes are introduced to model the hanger including high-fidelity finite element models and simple analytic models. Bayesian model selection is used to rank the models based on the experimental data.

Bayesian parameter estimation and uncertainty propagation is used to find the plausible values for the model parameters (hanger tension and boundary condition parameters) and propagate their uncertainties to make robust predictions.

- Chapter 4 is a case-study in a motorway bridge located in Greece. Field measurements are executed and acceleration time histories are collected from multiple sensor setups covering the entire length of the bridge. Bayesian modal identification is used to extract the modal frequencies and mode shapes from the acceleration measurements. The software used to perform the modal identification is presented in the Appendix. The experimentally identified modal data are used to perform model updating in a high-fidelity detailed finite element model of the bridge. Various aspects of modeling are examined such as the effect of the soil flexibility and the use of mode shapes as vector or scalar data (using the MAC value). All of those different modeling assumptions result in different models of the structure and their effect in Bayesian model updating is investigated.

- Chapter 5 is a case-study in a model of a seismically isolated bridge. Bayesian inference is used to quantify and propagate uncertainties in the parameters of a nonlinear model of a seismically isolated structure equipped with nonlinear hysteretic devices. The ultimate goal is to build high fidelity models of the system components to simulate the behavior (performance and reliability) of the combined system. The structural parameter calibration and uncertainty quantification is performed at the system level for the nonlinear hysteretic isolation devices using controlled laboratory tests performed on a shake table. Uncertainty models are identified using measurements of system tests, consisting of displacement, acceleration and restoring force response time histories. The model uncertainty analyses resulted in building a high fidelity model for the system that can be used for performing reliable robust performance predictions that properly take into account model uncertainties.

- Chapter 6 presents the Information Theoretic - Bayesian theory for model-based optimal experimental design. This problem is formulated as an optimization problem with respect to the design variables which describe the various experimental design possibilities such as sensor locations and excitation characteristics. The goal of the experiment is to perform Bayesian parameter inference about the parameters of the model. Two approaches are discussed in the approximation of the objective function, the sampling approach (derivative-free but computationally expensive) and the asymptotic approach (requires first derivatives but is computationally cheaper). The theory for both methods is developed. Novel theoretical findings are presented for both methods that provide insight into the problem of optimal experimental design. For the sampling method, a new improved formula for estimating the objective function is presented, followed by a novel interpretation. For the asymptotic method, a novel theorem is presented that explains the effect of the Gaussian prior in the optimal design. Chapter 6 ends with simple applications using analytic models that demonstrate both approaches and their differences. Numerical case-studies of optimal experimental design using

both the sampling and asymptotic methods with finite element models are carried out in Chapters 7 and 8.

- Chapter 7 is a numerical case-study in performing asymptotic optimal experimental design to find the optimal locations of acceleration sensors in order to do modal identification in a bridge. The case-study is carried out using the same high-fidelity detailed finite element model of the bridge used for model updating in Chapter 4. Due to the linear nature of the problem, the asymptotic method is chosen and its use is demonstrated to find the optimal locations for up to 20 sensors. The novel theoretical result regarding the effect of the Gaussian prior developed in Chapter 6 is demonstrated and verified numerically.

- Chapter 8 is a numerical case-study in both parameter estimation and optimal experimental design. The examined problem is a plate with a crack and the goals are to perform: 1) optimal strain sensor location in the plate for the most efficient and reliable crack identification and 2) crack identification using simulated experimental strain data. The same model is used for both applications, which is a parametrized finite element model of a plate with a crack, where the parameters describe the crack location, length and orientation. Due to the non-availability of derivatives of the strains with respect to the crack parameters, the sampling method is used for the optimal sensor location part. The strain sensors are placed in regular grid configurations of different types which are compared in order to find which is the best, using the new formula for the objective function estimator introduced in Chapter 6. The design variables describe the location and size of the grid for a particular number of strain sensors. The crack identification problem is treated as a Bayesian parameter estimation problem according to the theory of Chapter 2, where the parameters are those describing the crack. Due to the lack of real experimental data, simulated strain data were created in order to demonstrate the methodology. Emphasis is given in the effect of the number of strain sensors and the prediction error magnitude in the results.

- Chapter 9 summarizes the work done in this thesis and suggests directions for potential future work.

Chapter 2

Bayesian theory for uncertainty quantification, propagation and model selection

2.1 Introduction

A Bayesian framework for uncertainty quantification, propagation and model selection in complex structural dynamics simulations using experimental measurements is presented. The framework covers uncertainty quantification techniques for parameter estimation and model selection, as well as uncertainty propagation techniques for robust prediction of output quantities of interest in reliability and safety of the structural systems analyzed. Bayesian computational tools (sampling algorithms) are presented. The Bayesian framework and the computational tools are implemented in the next Chapters for linear and nonlinear models in structural dynamics using either identified modal properties or measured response time histories. High performance computing techniques that drastically reduce the excessive computational demands that arise from the large number of system simulations are outlined.

2.2 Parameter estimation

Consider a parameterized class M_m of structural dynamics models used to predict various output quantities of interest (QoI) of a system. Let $\underline{\theta}_m \in R^{N_m}$ be a set of parameters in this model class that need to be estimated using experimental data and $\underline{f}(\underline{\theta}_m|M_m)$ be model predictions of output QoI given a value of the parameter set $\underline{\theta}_m$. Probability density functions (PDF) are used to quantify the uncertainty in the parameters $\underline{\theta}_m$. The probability distribution of the parameter set $\underline{\theta}_m$ quantifies how plausible is each possible value of the model parameters. The user may assign a prior probability distribution

$\pi_m(\underline{\theta}_m)$ to the model parameters to incorporate prior information on the values of the model parameters. The structural model and uncertainty propagation algorithms can be used to identify the uncertainty in the prediction of the output QoI. However, the probability distribution $\pi_m(\underline{\theta}_m)$ is subjective based on previous knowledge and user experience.

In Bayesian inference, the interest lies in updating the probability distribution of the model parameters $\underline{\theta}_m$ based on measurements and then propagate these uncertainties through the structural dynamics model to quantify the uncertainty in the output QoI. For this, let $D \equiv \underline{\hat{y}} = \{\hat{y}_r \in R^{N_0}, r = 1, \cdots, m\}$ be a set of observations available from experiments, where N_0 is the number of observations. The Bayesian formulation starts by building a probabilistic model that characterizes the discrepancy between the model predictions $\underline{f}(\underline{\theta}_m|M_m)$ obtained from a particular value of the model parameters $\underline{\theta}_m$ and the corresponding data $\underline{\hat{y}}$. This discrepancy always exists due to measurement, model and computational errors. An error term $\underline{e}$ is introduced to denote this discrepancy. The observation data and the model predictions satisfies the prediction error equation

$$\underline{\hat{y}} = \underline{f}(\underline{\theta}_m|M_m) + \underline{e} \tag{2.1}$$

A probabilistic structure for the prediction error should be defined to proceed with the Bayesian calibration. Let M_e be a family of probability model classes for the error term $\underline{e}$. This model class depend on a set of prediction error parameters $\underline{\theta}_e$ to be determined using the experimental data. Similarly to the structural model parameters $\underline{\theta}_m$, the probability distribution $\pi_e(\underline{\theta}_e)$ is also assigned to quantify the possible values of the prediction error parameters $\underline{\theta}_e$.

The Bayesian approach (Beck and Katafygiotis, 1998; Beck, 2010) to model calibration is used for updating the values of the combined set $\underline{\theta} = (\underline{\theta}_m, \underline{\theta}_e)$ associated with the structural and the prediction error parameters. The parameters $\underline{\theta}_m$ and $\underline{\theta}_e$ can be considered to be independent with prior probability distribution for the combined set given by $\pi(\underline{\theta}|M) = \pi_m(\underline{\theta}_m|M_m)\pi_e(\underline{\theta}_e|M_e)$, where $M = \{M_m, M_e\}$ includes the structural and prediction error model classes. The updated PDF $p(\underline{\theta}|D, M)$ of the parameters $\underline{\theta}$, given the data D and the model class M, results from the application of the Bayes theorem

$$p(\underline{\theta}|D, M) = \frac{p(D|\underline{\theta}, M)\ \pi(\underline{\theta}|M)}{p(D|M)} \tag{2.2}$$

where $p(D|\underline{\theta}, M)$ is the likelihood of observing the data from the model class and $p(D|M)$ is the evidence of the model, given by the multi-dimensional integral

$$p(D|M) = \int_{\Theta} p(D|\underline{\theta}, M)\ \pi(\underline{\theta}|M)\ d\underline{\theta} \tag{2.3}$$

over the space of the uncertain model parameters.

The updated probability distribution of the model parameters depends on the selection of the prediction error $\underline{e}$. Invoking the maximum entropy principle, a normal distribution $\underline{e} \sim N(\underline{\mu}, \Sigma)$, where $\underline{\mu}$ is the mean and Σ is the covariance matrix, is a reasonable choice for the error since the normal distribution is the least informative among all distributions with specified the lowest two moments. The structure

imposed on the mean vector $\underline{\mu}$ and the covariance matrix Σ affect the uncertainty in the model parameter estimates. A zero mean model error is usually assumed so that $\underline{\mu} = \underline{0}$. However, to take into account the bias in the model predictions of the various response quantities involved in $\underline{f}(\underline{\theta}_m|M_m)$ and try to reconcile conflicting predictions, one could introduce a shift in the predictions by taking $\underline{\mu} \neq \underline{0}$. In this case the parameters defining the structure of $\underline{\mu}$ are part of the unknowns in $\underline{\theta}_e$ to be determined by the Bayesian technique. A diagonal matrix is a reasonable choice for the covariance matrix in the case where the components of the prediction error can be considered to be uncorrelated. This holds in the case of uncorrelated measurements in $\underline{\hat{y}}$ and independent components in the prediction vector $\underline{f}(\underline{\theta}_m|M_m)$. As a results, the covariance matrix takes the form $\Sigma = diag(\sigma_r^2 \hat{y}_r^2)$, where the variance parameters σ_r^2 are part of the unknown constants in $\underline{\theta}_e$ to be determined by the Bayesian calibration. In structural dynamics, the effect of prediction error correlation has been investigated and found to affect the results of the model calibration when the sensors are closely located (Simoen et al., 2013b). Depending on the nature of the simulated QoI, alternative prediction error models can also be used.

Using the prediction error equation (2.1), the measured quantities follow the normal distribution $\underline{\hat{y}} \sim N(\underline{f}(\underline{\theta}_m|D) + \underline{\mu}(\underline{\theta}_e), \Sigma(\underline{\theta}_e))$, where the explicit dependence of $\underline{\mu}(\underline{\theta}_e)$ and $\Sigma(\underline{\theta}_e)$ on $\underline{\theta}_e$ is introduced to point out that the mean and the covariance of the overall normal prediction error model depends only on the model prediction error parameters $\underline{\theta}_e$ and is independent of the structural parameters $\underline{\theta}_m$. Consequently, the likelihood $p(D|\underline{\theta}, M)$ of observing the data follows the multi-variable normal distribution given by

$$p(D|\underline{\theta}, M) = \frac{|\Sigma(\underline{\theta}_e)|^{-1/2}}{(2\pi)^{m/2}} \exp\left[-\frac{1}{2} J(\underline{\theta}; M)\right] \tag{2.4}$$

where

$$J(\underline{\theta}; M) = [\underline{\hat{y}} - \underline{f}(\underline{\theta}_m|M) - \underline{\mu}(\underline{\theta}_e)]^T \Sigma^{-1}(\underline{\theta}_e)[\underline{\hat{y}} - \underline{f}(\underline{\theta}_m|M) - \underline{\mu}(\underline{\theta}_e)] \tag{2.5}$$

The selection of the prior distribution affects the posterior distribution of the model parameters for the case of relatively small number of data. Usually a non-informative prior can be used. For example a uniform distribution of the model parameters does not give any preference to the values of the model parameters given the data. For cases of large number of model parameters where unidentifiability issues may occur, a Gaussian prior can avoid unidentifiability issues and enable the estimation of the model parameters using Bayesian numerical analysis tools, avoiding convergence problems of the gradient and stochastic optimization techniques used in Bayesian asymptotic approximations.

2.3 Model selection

The Bayesian probabilistic framework can also be used to compare two or more competing model classes and select the optimal model class based on the available data. Consider a family $M_{Fam} = \{M_i, i = 1, \cdots, \kappa\}$, of κ alternative, competing, parameterized FE and prediction error model classes, and let $\underline{\theta}_i \in R^{N_{\theta_i}}$ be the free parameters of the model class M_i. The posterior probability $Pr(M_i|D)$ of the i-th

model class given the data D is (Beck and Yuen, 2004; Yuen, 2010)

$$Pr(M_i|D) = \frac{p(D|M_i)\ Pr(M_i)}{p(D|M_{Fam})} \tag{2.6}$$

where $P(M_i)$ is the prior probability and $p(D|M_i)$ is the evidence of the model class M_i. The optimal model class M_{best} is selected as the one that maximizes $P(M_i|D)$ given by (2.6). Model class selection is used to compare between alternative model classes and select the best model class (Muto and Beck, 2008). The model class selection can also be used to identify the location and severity of damage (Ntotsios et al., 2009).

2.4 Uncertainty propagation for robust prior and posterior predictions

Let $q(\underline{\theta})$ be a scalar output QoI of the system. Prior robust predictions, before the availability of measured data, are derived by propagating the prior uncertainties in the model parameters quantified by the prior PDF $\pi(\underline{\theta}|M)$. Posterior robust predictions of $q(\underline{\theta})$ are obtained by taking into account the updated uncertainties in the model parameters given the measurements D. Let $p(q|\underline{\theta}, M)$ be the conditional probability distribution of q given the values of the parameters. Using the total probability theorem, the prior and posterior robust probability distribution $p(q|M)$ of q, taking into account the model M, is given by (Papadimitriou et al., 2001; Beck and Taflanidis, 2013)

$$p(q|M) = \int p(q|\underline{\theta}, M)\ p(\underline{\theta}|M)\ d\underline{\theta} \tag{2.7}$$

as an average of the conditional probability distribution $p(q|\underline{\theta}, M)$ weighted by the PDF $p(\underline{\theta}|M)$ of the model parameters, where $p(\underline{\theta}|M) \equiv \pi(\underline{\theta}|M)$ for prior estimate in the absence of data, or $p(\underline{\theta}|M) \equiv p(\underline{\theta}|D, M)$ for posterior estimate given the data D, respectively.

Let $G(q;\underline{\theta})$ be a performance measure of the system which depends on the deterministic output QoI $q(\underline{\theta})$. The prior robust performance measure $E[G(q;\underline{\theta})|M)] \equiv E_\pi[G(q;\underline{\theta})|M)]$ or the posterior robust performance measure $E[G(q;\underline{\theta})|M)] \equiv E_p[G(q;\underline{\theta})|D, M)]$ given the data D is

$$E[G(q;\underline{\theta})|M)] = \int G(q;\underline{\theta})\ p(\underline{\theta}|M)\ d\underline{\theta} \tag{2.8}$$

where $p(\underline{\theta}|M)$ is either the prior PDF $\pi(\underline{\theta}|M)$ or the posterior PDF $p(\underline{\theta}|D, M)$, respectively.

Robust predictions of $q(\underline{\theta})$ that account for the uncertainty in $\underline{\theta}$ are also obtained by simplified measures such as mean and variance ${\sigma_q}^2 = E[q^2(\underline{\theta})] - {m_1}^2 = {m_2}^2 - {m_1}^2$ with respect to $\underline{\theta}$, derived from the first two moments m_k of $q(\underline{\theta})$, $k = 1, 2$, given by the multi-dimensional integrals

$$m_k(D, M) = \int [q(\underline{\theta})]^k\ p(\underline{\theta}|M)\ d\underline{\theta} \tag{2.9}$$

over the uncertain parameter space. The integral (2.9) is a special case of (2.8) by selecting $G(q;\underline{\theta}) = [q(\underline{\theta})]^k$.

2.5 Sampling algorithms

In the general case, the model of the system is not a simple analytical model where the dependence on the parameters is explicit, and needs to be treated as a black-box system where we can only provide the inputs and get the outputs. In this case the calculation of the multi-dimensional integrals required for the estimation of the evidence of a model or for uncertainty propagation must be performed numerically. This is also true for the posterior distribution itself which can not be calculated as a closed-form expression of the parameters, since the dependence on the parameters is not explicit. The approach used throughout this thesis is to use sampling algorithms to deal with all of the above mentioned problems. A detailed description of the state-of-the-art computational tools for Bayesian inference can be found in Papadimitriou (2016).

Sampling algorithms are non-local methods capable of providing accurate representations for the posterior PDF and accurate robust predictions of output QoI. Sampling algorithms, such as Markov Chain Monte Carlo (MCMC) (Metropolis et al., 1953; Hastings, 1970; Cheung and Beck, 2009) are often used to generate samples $\underline{\theta}^{(i)}$, $i = 1, \cdots, N$, for populating the posterior PDF, estimating the model evidence and computing the uncertainties in output QoI. Among the stochastic simulation algorithms available, the transitional MCMC algorithm (TMCMC) (Ching and Chen, 2007) is one of the most promising algorithms for finding and populating with samples the importance region of interest of the posterior probability distribution, even in challenging unidentifiable cases and multi-modal posterior distributions. Approximate methods based on Kernels are then used to estimate marginal distributions of the parameters. In addition, the TMCMC method yields an estimate of the evidence in (2.3) of the model class M_i based on the samples already generated by the algorithm.

Sampling methods can be conveniently used to estimate the multi-dimensional integrals (2.7) and (2.8) from the samples $\underline{\theta}^{(i)}$, $i = 1, \cdots, N$, generated from the posterior probability distribution $p(\underline{\theta}|D, M)$. In this case, the integrals can be approximated by the sample estimates

$$p(q|M) \approx \frac{1}{N} \sum_{i=1}^{N} p(q|\theta^{(i)}, M) \tag{2.10}$$

$$E[G(q;\underline{\theta})|M)] \approx \frac{1}{N} \sum_{i=1}^{N} G(q;\underline{\theta}^{(i)}) \tag{2.11}$$

respectively. The simplified measures of uncertainties given in (2.9) are also given by the sample estimate (2.11) with $G(q;\underline{\theta}) = [q(\underline{\theta})]^k$. The sample estimates (2.10) and (2.11) require independent forward system simulations that can be executed in a perfectly parallel fashion.

2.6 Computational issues and tools

When the Bayesian theories are applied to real-world engineering problems, the models used in the simulations of those systems tend to be very complex in nature. In most practical engineering problems the model of the actual system is a finite element model. The finite element model complexity can be due to several reasons. Common sources of complexity include: non-linearities (geometrical or of other type), complicated geometry, and modeling of very special regions such as cracks. Non-linearities can appear in several forms including geometric, material, or equation nonlinearities. Geometric non-linearities require a more elaborate and sophisticated procedure to take place in order to solve the model, where several steps need to be taken to account for the changes in the geometry during the solution. A case with a geometrically non-linear finite element model is handled in Chapter 3. A complicated geometry requires in general much more finite elements in order to be discretized accurately than a simple geometry. This in turn increases the number of nodes and degrees of freedom of the model, resulting in an increase of computational time-to-solution. A case with a complicated geometry of an actual bridge is handled in Chapter 4. Finally, the modeling of very fine regions such as cracks using the finite element method requires an increased number of elements near the crack tip, which increases the number of nodes and degrees of freedom of the model as well, resulting in a higher computational time-to-solution. A case with a cracked structure is handled in Chapter 8.

Three common methods for reducing the computational cost are: model reduction, surrogate modeling and parallelization. Model reduction techniques are used to reduce the size of the system matrices (stiffness and mass matrices in finite element formulations) so that they can be solved faster, while retaining sufficient accuracy in the model predictions. A model reduction technique is used in order to reduce the size of a large high-fidelity finite element model of a bridge in Chapter 4. Surrogate modeling works by approximating the model prediction at a given parameter value based on its output at a set of fixed parameter values. In that way, surrogate modeling acts as an interpolating function for the model output, and replaces actual model runs with interpolation approximations. Model reduction and surrogate modeling techniques act on the model level by making it easier and cheaper to evaluate. On the hardware level, parallelization is a technique that is commonly used in order to exploit parallel computing environments by performing many model runs simultaneously in multiple CPU cores. In this way, the hundreds or thousands of required model runs are distributed among multiple CPU cores which in turn reduces the computational time. The TMCMC algorithm is very-well suited for parallel implementation in a computer cluster. Details of the parallel implementation are given in (Angelikopoulos et al., 2012; Hadjidoukas et al., 2015) and the algorithm is applied in Chapters 3, 4, 5 and 8. Specifically, a parallel implementation algorithm is activated at every stage of the TMCMC algorithm exploiting the large number of short, variable length, chains that need to be generated at the particular TMCMC stage. Dynamic scheduling schemes can be conveniently used to optimally distribute these chains in a multi-host configuration of complete heterogeneous computer workers. The dynamic scheduling scheme ensures an efficient balancing of the loads per computer worker in the case of variable run time of likelihood

function evaluations and unknown number of surrogates activated during estimation. Specifically, each worker is periodically interrogated at regular time intervals by the master computer about its availability and samples from TMCMC chains are submitted to the workers on a first come first serve basis to perform the likelihood function evaluations so that the idle time of the multiple workers is minimized. It should be noted that uncertainty propagation using sampling algorithms is highly parallelizable. For infinite computing resources, the time to solution for making robust prediction of a number of response QoI can be of the order of the time to solution for one simulation run.

Chapter 3

Bayesian estimation of tension in bridge hangers using modal frequency measurements

Original publication:
Papadimitriou, C., Giakoumi, K., Argyris, C., Spyrou, L. and Panetsos, P. "Bayesian estimation of tension in bridge hangers using modal frequency measurements". In: *Journal of Structural Monitoring and Maintenance*, 3(4): 349-375, (2016).

Abstract

The tension of an arch bridge hanger is estimated using a number of experimentally identified modal frequencies. The hanger is connected through metallic plates to the bridge deck and arch. Two different categories of model classes are considered to simulate the vibrations of the hanger: an analytical model based on the Euler-Bernoulli beam theory, and a high-fidelity finite element (FE) model. A Bayesian parameter estimation and model selection method is used to discriminate between models, select the best model, and estimate the hanger tension and its uncertainty. It is demonstrated that the end plate connections and boundary conditions of the hanger due to the flexibility of the deck/arch significantly affect the estimate of the axial load and its uncertainty. A fixed-end high fidelity FE model of the hanger underestimates the hanger tension by more than 20 compared to a baseline FE model with flexible supports. Simplified beam models can give fairly accurate results, close to the ones obtained from the high fidelity FE model with flexible support conditions, provided that the concept of equivalent length is introduced and/or end rotational springs are included to simulate the flexibility of the hanger ends. The effect of the number of experimentally identified modal frequencies on the estimates of the hanger

tension and its uncertainty is investigated.

3.1 Introduction

Hangers are used as deck support elements in arch bridges. Methods to monitor the axial loads in hangers are important for identifying the structural integrity of arch bridges. Large enough axial loads in the hangers affect the hanger modal frequencies due to the stiffness increase, so that the estimation of the axial load can be based on comparing model predictions with the experimentally estimated modal frequencies. The axial load is then estimated as the one that gives model predictions of the modal frequencies that matches the experimentally identified modal frequency values. The model predictions of the modal frequencies are affected by the flexural rigidity, end connection details and boundary conditions at the end supports of the hangers. Connections of a circular cross-section hanger to the bridge deck and arch may often be made through metallic guides or plates and give rise to boundary conditions that are not well defined, complicating the selection of the appropriate boundary conditions in a modeling procedure (Lagomarsino and Calderini, 2005). Depending on the connections, the arch or the deck may be flexible in different direction of motion. To obtain reliable predictions of the hanger axial loads these factors should be taken into account in the modeling.

Methods based on string and beam theory (Bokaian, 1990; Barcilon, 1976) have been developed for estimating the cable tension from experimentally identified modal frequencies. Exact formulas requiring the iterative solution of the characteristic equation as well as simplified practical formulas have been proposed for the estimation of the tension (Shinke et al., 1996; Fang and Wang, 2012; Ren et al., 2005; Nam and Nghia, 2011; Huang et al., 2015) taking into account the bending stiffness and sag-extensibility. To account for the boundary conditions, Ceballos and Prato (2008) introduced rotational springs at the cable ends. Techniques were introduced to approximately determine the rotational spring stiffnesses and the axial force and cable bending stiffness were then adjusted to fit the experimental values of the modal frequencies. The problem of tension estimation in tie-rods using beam theory and rotational springs to simulate unknown boundary conditions was also discussed in Lagomarsino and Calderini (2005). Bellino et al. (2010) has introduced the concept of equivalent cable length to account for the unknown boundary conditions. They proposed a method to estimate the cable tension by means of vibration response and moving mass technique. The same authors (Bellino et al., 2011) introduced the modal length concept and developed a method for estimating the cable tension using supplementary measurements from a cable with an added mass. Belleri and Moaveni (2015) have also used the concept of equivalent length and rotational springs to account for uncertain boundary conditions, providing reliable estimates of tensile loads in tie rods using measurements of the first modeshape in addition to the modal frequency measurements.

FE modeling of the hanger was also used to accurately identify the hanger tension, flexural and axial rigidity. A FE model can be used as a baseline model for identifying the cable tension (Kim and Park, 2007). The importance of boundary conditions on the reliable estimates of the modal frequencies from a FE model was pointed out in Park et al. (2015). Ni et al. (2002) used FE modeling to estimate the

cable tension from multimodal measurements, concluding that for long-span large-diameter cables the tension is reliably estimated when accurately accounting for all effects such as bending stiffness, boundary conditions as well as other constraints. The construction of a finite element model requires special software and resources and may be a time consuming procedure, especially when solving the inverse problem of estimating the cable tension and boundary conditions. Simple models based on beam theory that include the effect of flexural rigidity, flexibility due to end connections and boundary conditions can provide computationally inexpensive estimates of the hanger tension.

The objective of this work is to estimate the hanger tension in relatively large-diameter hangers of arch bridges that are connected to the bridge deck and arch through sizable metallic end plates, asymmetrically oriented at bottom and top hanger ends, making difficult the specification of the cable length, hanger-plate assembly flexibility and boundary conditions. The estimation is based on modal frequencies of the cable, in both the longitudinal and transverse direction of the bridge, identified from modal tests. Two different categories of model classes are introduced to represent the dynamics of the hanger. Different model classes in both categories are used to predict the hanger modal frequencies given the axial load in the hanger. The first category is based on a conventional analytical beam model formulation resulting from Euler-Bernoulli beam theory, used to predict the modal frequencies of the hanger with various boundary conditions, given the tension on the hanger. Two sets of boundary conditions are considered, one with fixed ends and the other with flexible ends, quantified by rotational springs attached at the ends. The concept of the equivalent length that is left free to be estimated together with the hanger tension so that predictions match the measurements is also considered. Analytical transcendental equations are developed and numerically solved to obtain the modal frequencies.

The second category is based on high-fidelity three-dimensional baseline FE models developed in Abaqus for the hanger, including the end plate connections. Two different types of boundary conditions are implemented and the effect on the hanger tension estimation is evaluated. The first type assumes fixed ends, while the second type models the end surfaces that connect to the deck and the arch as flexible with respect to the rotation about axes perpendicular to the hanger axis. The flexibility of the end supports arises from the flexibility of the arch and the deck at the connection ends and is appropriately modeled by attaching on the boundary plate ends a set of distributed springs along the direction of hanger axis to simulate the rotation along the two axes perpendicular to the hanger axis. The procedure for predicting the modal frequencies from the FE models given the hanger tension and the boundary conditions is outlined.

Bayesian inference (Beck and Katafygiotis, 1998; Yuen, 2010) for parameter estimation and model selection is used for estimating the hanger tension based on the different model classes introduced and the experimentally identified modal frequencies. The Bayesian model selection method (Beck and Yuen, 2004) is used to select the best model class for representing the dynamics of the hangers. In contrast to existing methods, the present work uses Bayesian inference for the first time to discriminate between model classes, select the best model class out of a series of increasingly complex models, as well as estimate the axial force and its uncertainty. The present investigation includes comparison of results

and conclusions related to the estimate of the hanger tension along with its uncertainty, the effect of the number of experimental frequencies available from modal tests, the effect of end plate connections and boundary conditions on the prediction of the hanger tension, as well as the adequacy of each one of the model class to represent the dynamics of the hanger.

3.2 Description of hanger and experimental data

The hanger under investigation is hanger 3 of the 20 hangers used to support the deck of an arch bridge located in Athens, Greece. The arch bridge is shown schematically in Figure 3.1. The hanger geometry, along with the geometry of the connections of the hangers to the deck and the arch, is shown in Figure 3.2a. The hangers are made out of steel and they are connected to the deck and the arch substructures with edge plates as shown in Figure 3.2b. The edge plates are welded to the hangers and deck or the arch. The connections of the hanger with the deck or the arch with end plates are approximately 1 m long and affect the hanger flexibility at the two ends. The plate that connects the hanger with the deck has its orientation along the longitudinal direction of the bridge deck, while the plate that connects the hanger to the arch has its orientation along the transverse direction of the bridge deck. The boundary conditions at the plate surface that connects to the arch or the deck substructures depend on the flexibility of the deck and the arch.

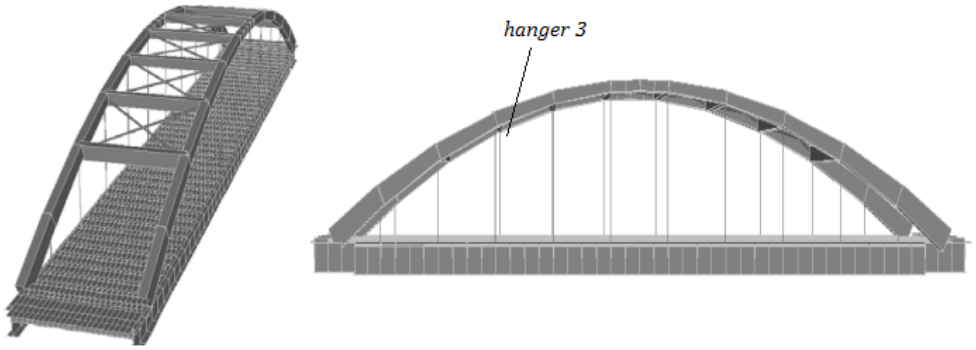
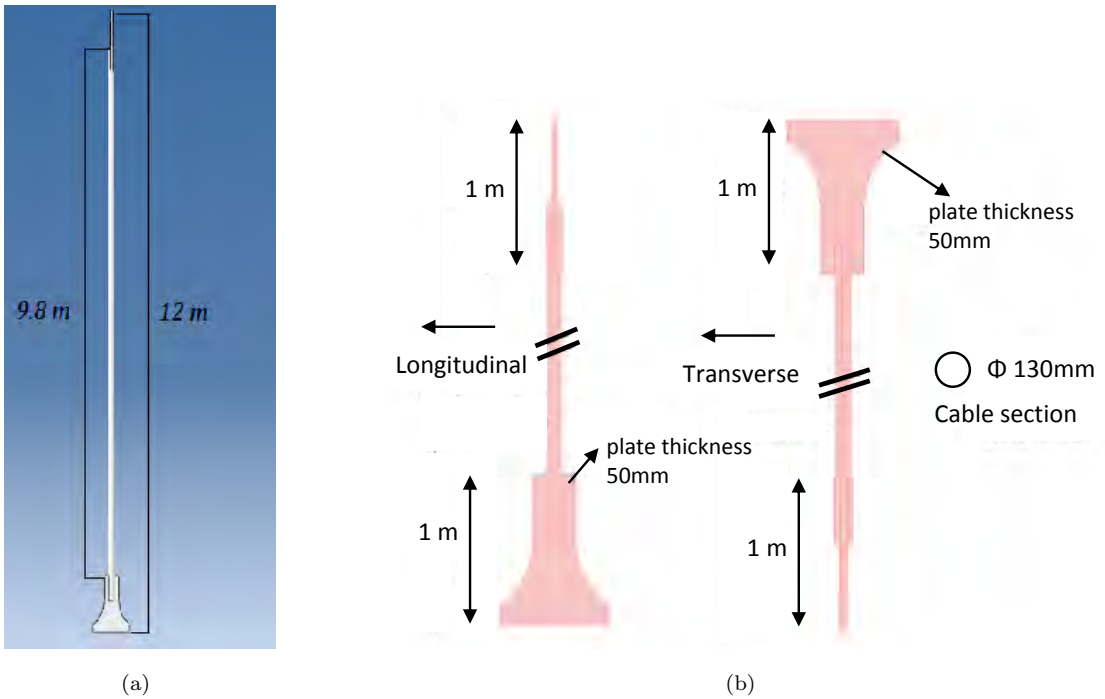


Figure 3.1: The arch bridge

The hangers are made of steel with modulus of elasticity $E = 200Gpa$, mass density 7800 kg/m^3 and Poisson ratio 0.3. The total length of the hanger 3 including the connections is 12 m, while the clear length of the circular section of the hanger is 9.817 m. The diameter of the circular section of the hanger is 0.13 m. It should be noted that the hanger differs from conventional cables which are often assumed as string elements. The bending stiffness of the present hanger cannot be ignored in predicting the modal frequencies. In addition the plate elements installed in the hangers to connect the circular section to

the bridge arch and deck also affect the bending stiffness. Based on the design plans, the geometry of the hanger, its material properties and the connection details of the two edge plates of the hanger are identical. The only difference is the orientation of each edge plate. Assuming that the end conditions of the edge plates are fixed, the modal frequencies of vibration of the hanger along the longitudinal and transverse direction bridge are expected to be identical.



(a) (b)

Figure 3.2: (a) The geometry of hanger 3, (b) The geometry of the plates connecting the circular hanger to the arch and the deck of the bridge

Impulse hammer tests performed on the hanger 3 of the bridge are used to estimate the modal frequencies of the hanger. The tests are performed using two acceleration sensors placed on the hanger at distances approximately 1.9m to 2.3m from its bottom edge. The sensor configuration is such that one acceleration sensor measures along the longitudinal direction and the other along the transverse direction of the bridge. To avoid placing sensors close to a node of a mode shape, two sensor setups are used, one with both sensors placed at approximately 2.3 m from the bottom edge of the hanger and the other with both sensors placed at approximately 0.4 m lower than the first location. For each set up, two impulse hammer tests are performed, one striking the hammer on the hanger along the longitudinal and the other striking the hanger along the transverse direction of the bridge, ensuring that the hanger modes along the longitudinal and transverse directions are excited, respectively. The impact locations of the hammer are chosen to be in close proximity to the location of the sensors. For each setup and test case, the measurements consist of the time histories of the hanger force and the two accelerations.

Two sets of measurements are available to identify each mode. For each set of measurements, the optimal values of the lowest six modal frequencies in either the longitudinal or the transverse directions are estimated using nonlinear least-squares frequency domain methods, minimizing the Euclidean norm of the difference between the experimentally identified frequency response functions estimated by the measured time histories and the ones predicted by a modal model of the hanger (Heylen et al., 1998; Ewins, 2000). Illustrative forced vibration acceleration time histories can be found in Appendix B.

Table 3.1: Experimentally identified modal frequencies (Hz) of hanger 3 in the transverse (trans) and longitudinal (long) directions

Mode #	trans	long	% difference
1	5.82	6.09	2.3
2	13.85	14.8	3.3
3	26.17	27.0	1.6
4	40.47	41.8	1.6
5	59.3	61.5	2.8
6	81.3	83.68	1.4

Table 3.1 shows the optimal values of the experimentally identified modal frequencies of hanger 3 in the transverse and longitudinal directions. Note that the uncertainty in the experimentally identified values of the modal frequencies are not reported due to the small number of experimental tests available. However, they are expected to be considerably less than 1%. It is observed that the modal frequencies differ in the two longitudinal and transverse directions. The percentage differences range from 1.4% to 3.3% and cannot be justified by material or geometric variability of the hanger. Due to the symmetry of the hanger and the connection plates, this is a strong indication that the boundary conditions at the end of the hangers are responsible for such differences. Thus the hanger end conditions cannot be assumed to be fixed. This study investigates the effects of the boundary conditions on the estimation of the axial force and provides evidence, based on Bayesian inference, that fixed boundary conditions assumed for the ends may result in misleading estimates of the axial hanger loads.

3.3 Modal frequency predictions based on beam theory

The prediction of the modal frequencies of the hanger subjected to an axial load is based on the Euler-Bernoulli beam theory. The equation of motion of a beam subjected to axial tension T along the beam axis z is given by (William, 1996):

$$EI\frac{\partial^4 u}{\partial^4 z} - T\frac{\partial^2 u}{\partial^2 z} + \rho A\left(\frac{\partial^2 u}{\partial t^2}\right) = 0 \tag{3.1}$$

where $u \equiv u(z,t)$ is the deflection of the beam in a direction y in the (y,z) plane, ρ is the density, E is the modulus of elasticity, I is the moment of inertia of the circular cross section about the x and y axes,

and A is the area of the cross-section of the beam. All geometrical and material properties are assumed constant along the length of the beam.

Two models are introduced that differ on the boundary conditions considered. In the first model the ends of the beam are fixed, whereas in the second model the ends are flexible. The flexibility in rotation of the ends is simulated using rotational springs.

3.3.1 Beam with fixed ends

For fixed-end supports, the boundary conditions are $u(0,t) = 0$, $u(L,t) = 0$, $u'(0,t) = 0$, $u'(L,t) = 0$. Following the usual eigenvalue analysis, the modal frequencies are obtained by solving the characteristic equation

$$h(\gamma,\zeta) = det \begin{pmatrix} 0 & 1 & 0 & 1 \\ bl & -c_{22}bl & al & c_{24}al \\ sinh(bl) & cosh(bl) & sin(al) & cos(al) \\ c_{41}blcosh(bl) & c_{42}blsinh(bl) & c_{43}alcos(al) & -c_{44}alsin(al) \end{pmatrix} = 0 \tag{3.2}$$

where al and bl are given as

$$al = l\sqrt{\sqrt{\zeta^4+\gamma^4}-\zeta^2}, \qquad bl = l\sqrt{\sqrt{\zeta^4+\gamma^4}+\zeta^2} \tag{3.3}$$

with respect to the two parameters γ and ζ defined by

$$\gamma^4 = \frac{m\omega^2}{EI}, \qquad \zeta = \sqrt{\frac{T}{2EI}} \tag{3.4}$$

and $m = \rho AL$ is the mass of the beam. The elements c_{ij} are given in this case by $c_{22} = c_{24} = 0$ and $c_{41} = c_{42} = c_{43} = c_{44} = 1$. Equation (3.2) can also be written as (Shinke et al., 1996):

$$2(al)(bl)[1 - cos(al)cos(bl)] + [(bl)^2 - (al)^2]sin(al)sinh(bl) = 0 \tag{3.5}$$

The problem of estimating the modal frequency ω in (3.2) or (3.5) given the axial force T is turned into the problem of estimating γ given the value of ζ. The values of γ can be obtained by the numerical solution of (3.2) or (3.5).

To proceed, the following dimensionless parameter

$$\xi = \sqrt{\frac{T}{EI}}l \tag{3.6}$$

is introduced due to its significant role (Shinke et al., 1996) in the dynamic behavior of the beam. For large values of ξ ($\xi \geq 20$) the dynamic characteristics of the beam are similar to those of a string. For small values of ξ ($\xi < 20$), the characteristics of a hanger are similar to those of a beam. Two different parameterizations have been proposed (Shinke et al., 1996) depending on the range of ξ values.

For large values of ξ ($\xi \geq 20$), the dimensionless parameter

$$\eta_n = \frac{f}{f_n^s} \tag{3.7}$$

is introduced, where

$$f_n^s = \frac{n}{2l}\sqrt{\frac{Tg}{w}} \tag{3.8}$$

is the theoretical values of the n-th order natural frequency ($\omega = 2\pi f$) of a string (Humar, 2012). In this case al and bl in (3.3) take the form:

$$al = \frac{\xi}{\sqrt{2}}\sqrt{-1+\sqrt{1+\left(\frac{2n\pi\eta_n}{\xi}\right)^2}} \qquad bl = \frac{\xi}{\sqrt{2}}\sqrt{1+\sqrt{1+\left(\frac{2n\pi\eta_n}{\xi}\right)^2}} \tag{3.9}$$

while the characteristic equation (3.5) becomes

$$g(\eta_n,\xi) = 2n\pi\eta_n(1-cos(al)cosh(bl)) + \xi sin(al)sinh(bl) = 0 \tag{3.10}$$

where, using equations (3.4), (3.6), (3.7) and (3.8), the function $g(\eta,\xi) = h(\sqrt{n\pi\xi\eta}/l, \xi/(l\sqrt{2}))$ with $h(\gamma,\zeta)$ defined in (3.2). The normalized modal frequencies η_n are obtained by solving the characteristic equation (3.10) for a given value of ξ.

For small values of ξ ($\xi < 20$), the dimensionless parameter

$$\phi_n = \frac{f}{f_n^b} \tag{3.11}$$

is introduced, where

$$f_n^b = \frac{a_n^2}{2\pi l^2}\sqrt{\frac{EIg}{w}} \tag{3.12}$$

is the theoretical values of the n-th order natural frequency of a beam fixed at both ends (Humar, 2012). The values of a_n are the solutions of $cos(a)cosh(a) = 1$. The first six solutions are given in Table 3.2.

Table 3.2: The first six roots of $cos(a)cosh(a) = 1$

a_1	a_2	a_3	a_4	a_5	a_6
4.73	7.8532	10.9956	14.1372	17.2788	20.4204

In this case al and bl in (3.3) are transformed into

$$al = \frac{\xi}{\sqrt{2}}\sqrt{-1+\sqrt{1+\left(\frac{2\alpha_n^2}{\xi^2}\phi_n\right)^2}}, \qquad bl = \frac{\xi}{\sqrt{2}}\sqrt{1+\sqrt{1+\left(\frac{2\alpha_n^2}{\xi^2}\phi_n\right)^2}} \tag{3.13}$$

while the characteristic equation takes the form

$$g(\phi_n,\xi) = 2a_n^2\phi_n(1-cos(al)cosh(bl)) + \xi^2 sin(al)sinh(bl) = 0 \tag{3.14}$$

where, using equations (3.4), (3.6), (3.11) and (3.12), the function $g(\eta,\xi) = h(\alpha_n\sqrt{\eta}/l, \xi/(l\sqrt{2}))$ with $h(\gamma,\zeta)$ defined in (3.2). The normalized modal frequencies ϕ_n are obtained by solving (3.14) for a given value of ξ. When the axial force approaches zero ($\xi = 0$) then ϕ_n tends to 1.

(a) Transverse y

(b) Longitudinal x

Figure 3.3: The rotational springs resisting rotation of ends about the transverse y and longitudinal x directions

3.3.2 Beam with flexible ends

For flexible supports, modeled by rotational springs at the two ends as shown in Figure 3.3, the boundary conditions for beam deflections in the (y, z) plane are $u(0,t) = 0$, $EIu''(0,t) - k_1u'(0,t) = 0$, $u(L,t) = 0$ and $EIu''(L,t) + k_3u'(L,t) = 0$, where k_1 and k_3 are the rotational springs applied at the ends (Figure 3.3a) to resist rotation around x direction due to bending. Similar support conditions hold for beam displacements in the (x, z) plane with k_2 and k_4 introduced as rotational springs applied at the ends (Figure 3.3b) to resist rotations with respect to the y direction. The prediction of the axial force is obtained by solving the characteristic equation for the flexible boundary conditions represented by rotational springs (Lagomarsino and Calderini, 2005). Ceballos and Prato (2008) have derived explicit expressions which are required to be solved iteratively. Next we briefly state the characteristic equation.

Introducing the following dimensionless parameters for the spring constants:

$$k_a = \frac{k_1 l}{EI} \qquad \text{and} \qquad k_c = \frac{k_3 l}{EI} \tag{3.15}$$

and following the usual eigenvalue analysis, the modal frequencies in the (y, z) plane are obtained by solving the characteristic equation (3.2), where the elements c_{ij} are given in this case by

$$c_{22} = \frac{bl}{k_a}, \qquad c_{24} = \frac{al}{k_a} \tag{3.16a}$$

$$c_{41} = 1 + \frac{(bl)}{k_c} tanh(bl), \qquad c_{42} = 1 + \frac{(bl)}{k_c} \frac{1}{tanh(bl)} \tag{3.16b}$$

$$c_{43} = 1 - \frac{(al)}{k_c} tan(al), \qquad c_{44} = 1 + \frac{(al)}{k_c} \frac{1}{tan(al)} \tag{3.16c}$$

Note that for very large values of k_a and k_c, such that the following conditions hold

$$\frac{al}{k_a} << 1, \quad \frac{bl}{k_a} << 1, \quad \frac{al}{k_c}[tan(al)]^{\pm 1} << 1, \quad \text{and} \quad \frac{bl}{k_c}[tanh(bl)]^{\pm 1} << 1 \tag{3.17}$$

the corresponding elements $c_{22} = c_{24} = 0$ and $c_{41} = c_{42} = c_{43} = c_{44} = 1$ and the characteristic equation tends to the one for the beam with fixed supports. Moreover, for large values of ξ ($\xi >> 20$), say $\xi > 200$ the hanger behaves as a string which means that the flexural stiffness of the hanger is not important and thus the flexibility of the end supports does not affect the modal frequencies.

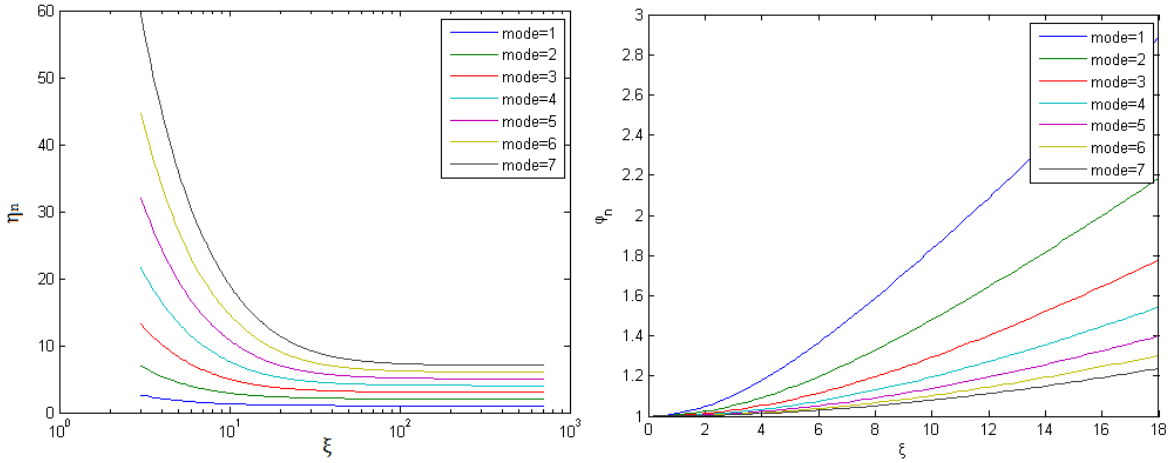
Similar expression holds for estimating the modal frequencies in the (x, z) plane with k_a and k_c replaced by $k_b = k_2 l/EI$ and $k_d = k_4 l/EI$.

3.3.3 Estimation of modal frequencies

The lowest seven dimensionless frequencies η_n and ϕ_n for beam with fixed ends, calculated from the equations (3.10) and (3.14) for large and small values of ξ, are given in Figures 3.4a and 3.4b, respectively. Equations (3.10) and (3.14) are transcendental equations and for their solution an iterative method can be used such as the Newton-Raphson algorithm. Herein the "fzero" function in Matlab is used.

In order to minimize or eliminate the probability of missing the correct solutions for a given ξ value, the procedure of estimating modal frequencies starts from large values of ξ ($\xi > 200$), where the solution approaches the known values given, due to equations (3.7) and (3.8), by $\eta_n = n$, $n = 1, 2, \ldots$. Subsequently, the modal frequencies for smaller values $\xi - \delta\xi$ are obtained iteratively using the previous solutions for ξ as an initial estimate to find the zeros of the function $g(\eta, \xi - \delta\xi)$ close to the solutions at $\eta(\xi)$. Using the method described above, equations (3.10) and (3.14) have been solved and the values of η or ϕ are tabulated for values of ξ ranging from [0,700] to be further used for estimating the modal frequencies at the intermediate values of ξ for either fixed or flexible supports as follows.

First, it should be noted that the solution of (3.10) and (3.14) for fixed ends are upper bounds of the solution of (3.10) with flexible ends, since the flexibility at the supports is expected to reduce the values of the modal frequencies for all modes. This observation can be used to construct an algorithm for reliably obtaining the solutions of the modal frequencies of the beam with fixed or flexible supports for any value of $\xi \in [0, 700]$. These solutions are obtained using fsolve or fzero in Matlab by providing the intervals where the solution lies or starting values very close to the solutions. This has to be done with a 100% reliability since this procedure is automated to be used within the Bayesian framework to compute the solutions for a large number of sample points ξ. To find the lowest k solutions of the transcendental equation $g(\eta, \xi)$ for a fixed value of ξ, the following steps are performed:



(a) (b)

Figure 3.4: The dimensionless solutions (a) $\eta_n = n\eta$ versus ξ for large values of ξ, and (b) ϕ ($\phi_n = a_n^2\phi$) versus ξ for small values of ξ for beam with fixed ends

1. Given ξ, find i such that $\xi \in [\xi_i, \xi_{i+1}]$ using the tabulated ξ_i values.
2. Estimate $\eta_i = \eta(\xi_i)$ and $\eta_{i+1} = \eta(\xi_{i+1})$ and use linear interpolation to estimate $\eta = \eta(\xi) = \eta(\xi_i) + [\eta(\xi_{i+1}) - \eta(\xi_i)](\xi - \xi_i)/(\xi_{i+1} - \xi_i)$.
3. Divide the interval $[0, (1+a)\eta]$, where $a << 1$ is selected by the user, into N equal intervals of size $\Delta\eta = (1+a)\eta/N$. Compute $g_j = g(j\Delta\eta, \xi)$, $j = 1 \ldots, N$.
4. Find the values of η among the $\eta_j = j\Delta\eta$ for which a sign change occurs in the function $g(\eta, \xi)$. The Matlab command η(find(diff(sign(g))$\sim$= 0) can be used. Let ℓ be the number of sign changes.
   - If $\ell >= k$ select the lowest k values η_j, $j = 1, \ldots, k$ and use each one of these k values as starting value to find the zeros of the function $g(\eta, \xi)$.
   - If $\ell < k$ then go to 3, set $N \leftarrow 2N$ and repeat the steps until an $\ell >= k$.

The algorithm guarantees that the lowest k modal frequencies can be found with a high reliability, provided that $\Delta\eta$ is significantly less than the minimum distance between consecutive roots of $g(\eta, \xi)$. The cost of this procedure is that it requires N function evaluations of the determinant in (3.2). However, this cost is unavoidable in order to guarantee that a zero of the function is not missed.

3.3.4 Beam model classes

Two families of model classes are introduced based on the Euler-Bernoulli beam theory. The objective is to examine which of the introduced model classes are most appropriate to model the hanger and produce reliable estimates of the hanger tension. Model class *Bfix* is the beam model with fixed ends, while model

class *Bflex* is a beam model with flexible ends simulated by rotational springs, two at each side of the beam. For the three-dimensional beam, the vibrations of the beam in the (x, z) and (y, z) planes are considered uncoupled for both model classes. Thus, for a given value of ξ, the modal frequencies are computed by solving 2 two-dimensional beam problems, considering the boundary conditions for each plane motion, one in the (x, z) plane with spring constants k_1, k_3 and the other in the (y, z) plane with spring constants k_2, k_4. The concept of effective length (Bellino et al., 2010) is also introduced to take into account the flexibility of the end conditions. Beam models with either one or two effective lengths (one per direction of deflection) are used to account for the different flexibilities per direction of the end connections due to different orientation of the connecting plate at the bottom and top of the hanger and the unknown boundary conditions arising from the deck and arch flexibility. The model classes are flexible to predict different modal frequencies along the transverse and longitudinal direction of the beam by using different effective lengths per direction of hanger deflection and/or applying different rotational spring constants. Each one of the introduced families of model classes contain models that are further classified in Section 6 depending on the number and type of parameters they include for estimation.

3.4 Modal frequency predictions based on finite element models

A high fidelity FE model is also used to predict the modal frequencies of the hanger shown in Figure 3.2, under different boundary conditions. The FE modeling and analyses are carried out using the ABAQUS general purpose FE program. Second-order ten-node tetrahedral elements (C3D10) are used to create the FE mesh in the $10m$ beam with circular cross-section, as well as in the two $1m$ end plate connections of the circular beam with the rest of the bridge structure. The model consists of about 71,500 nodes, 43,000 elements, and a total of 215,000 nodal DOF. Mesh sensitivity studies were carried out to select the optimal mesh size that ensures convergent numerical calculations providing accurate predictions of the lowest 12 modal frequencies.

3.4.1 Types of boundary conditions and FE model classes

Two different types of boundary conditions are considered. The first type corresponds to fixed ends at the boundaries, implemented by constraining the motion of the DOFs at the bottom and top plate surfaces (Figure 3.5), connecting with the deck and the arch, to be zero. The fixed-end finite element model is denoted by $FEfix$ and involves the hanger tension as parameter to be estimated using the measured data. The second type permits only the rotation of the bottom and top plate surfaces (connecting with the deck and the arch) about the two axes x and y along the longitudinal and transverse directions of the bridge deck, respectively. These boundary conditions are implemented by constraining the motion of the midpoint of the boundary plate surface to zero along all three x, y and z directions, constraining the movement of the side nodes of the edge surface along the x and y directions to zero, and adding springs along the z direction of the side nodes of the edge surface, restraining their motion along the z

direction of the hanger according to the spring constants. Such springs provide resistance to rotations of the edge surfaces with respect to the x and y axes. Two independent sets of springs are added to simulate the rotational resistance with respect to the x and y axes. Each set is uniformly distributed along the opposite sides of the edge surface. The distributed stiffness values are denoted by $k_{rx,b} = k_1$ for the springs along the sides 1-2 and 3-4 of the bottom edge surface (Figure 3.5a), $k_{rx,t} = k_3$ for the springs along the sides 2-3 and 4-1 of the top edge surface (Figure 3.5b), $k_{ry,b} = k_2$ for the springs along the sides 2-3 and 4-1 of the bottom edge surfaces (Figure 3.5a), and $k_{ry,t} = k_4$ for the springs along the sides 1-2 and 3-4 of the top edge surfaces (Figure 3.5b). The flexible end FE model is denoted by $FEflex$ and involves five parameters to be estimated using the measured data: the hanger tension and the four distributed spring constants k_1, k_2, k_3 and k_4.

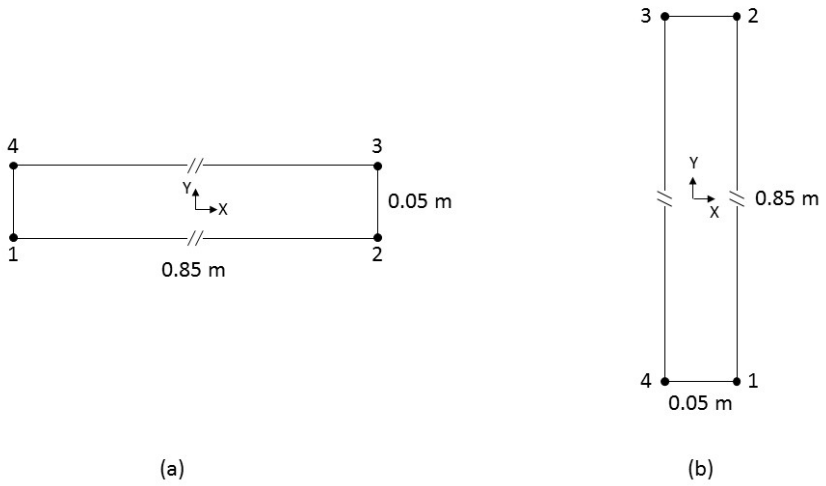


Figure 3.5: Plate surface connecting with (a) deck (bottom edge of hanger), and (b) arch (top edge of hanger)

3.4.2 Estimation of modal frequencies

The high tension of the hanger affects the modal frequencies due to increase in the tangent stiffness. In order to predict the tangent stiffness due to hanger tension and subsequently the effect on the modal frequencies of the hanger, a geometrically nonlinear analysis of the hanger has to be performed. The evaluation of the modal frequencies in Abaqus that takes into account the stiffness increase due to large hanger tension consists of a certain sequence of actions. The aim is to obtain the modal frequencies of the system after the hanger tension has been applied.

Action 1: This is a static deformation step for estimating the tangent stiffness matrix under the application of the hanger tension. The hanger tension is applied at one edge (e.g. bottom edge in Figure

3.5a). The boundary conditions at the bottom edge are deactivated to allow the hanger to deform. The axial load T is applied as a pressure $p = T/(bh)$ uniformly distributed through the bottom face of the connection. The geometrically nonlinear static analysis is performed and the tangent stiffness matrix and the mass matrix are then extracted from the Abaqus model.

Action 2: After the hanger has been elongated in Action 1, the correct boundary conditions are activated at the bottom edge and the hanger is kept in its deformed state. Specifically, for fixed-end hanger conditions, all displacement DOFs at the bottom edge of the hanger are restrained. For flexible supports, springs are added at the bottom edge DOFs so that rotational stiffness conditions are simulated. The resulting mass matrix as well as the stiffness matrix which consist of the tangent stiffness matrix and the stiffness due to the spring constants is extracted from Abaqus.

Action 3: Using the mass matrix and the tangent stiffness matrix obtained from Action 2, the eigenvalue analysis is performed to obtain the modal frequencies and the mode shapes of the hanger. It should be noted that due to the non-circular cross-section of the hanger along its total length, arising from the 1m plate connections, a mode shape deforms both in (x, z) and (y, z) planes, in contrast to the in-plane deformation predicted by the Euler-Bernoulli beam theory when circular cross-section is assumed. The mode shapes are used to distinguish the type of the closely spaced modes by monitoring the mode shape deformation in the (x, z) and the (y, z) planes. In this way mode switching which may happen as the hanger tension varies in the Bayesian inverse formulation outlined in Section 5 can be monitored so that the correct modes with dominant hanger deflection in the longitudinal or transverse direction of the bridge are associated to the experimentally identified modes.

3.5 Bayesian inference

The Bayesian framework for parameter estimation and model selection (Beck and Katafygiotis, 1998; Yuen, 2010; Vanik et al., 2000) is used to estimate the hanger axial load based on the model classes introduced in the previous sections. The inference is based on the lowest m experimentally identified modal frequencies $D = \{\hat{\omega}_r, r = 1, ..., m\}$ of the hanger. Consider a parameterized model class M_m and let $\underline{\theta} \in R^{N_\theta}$ be the set of free structural model parameters to be identified. Depending on the model class used, the model parameters may include the hanger force, the equivalent beam length(s) and the rotational stiffness of the boundary conditions of the hanger. Let also $\omega_r(\underline{\theta})$ be the predictions of the modal frequencies obtained for a particular value of the parameter set.

Probability density functions (PDF) are used to quantify uncertainties in the parameter set $\underline{\theta}$. Using Bayes theorem, the posterior PDF $p(\theta|D, M_m)$ of the model parameters $\underline{\theta}$ based on the measured data D and the modeling assumptions M_m, is obtained as follows:

$$p(\underline{\theta}|D, M_m) = \frac{p(D|\underline{\theta}, M_m)p(\underline{\theta}|M_m)}{p(D|M_m)} \tag{3.18}$$

where $p(D|\underline{\theta}, M_m)$ is the probability of observing the data (likelihood function) from a model M_m corresponding to a particular value of the parameter set $\underline{\theta}$, $p(\underline{\theta}|M_m)$ is the prior PDF of the model parameters, and $p(D|M_m)$ is the evidence of the model class given by

$$p(D|M_m) = \int_{\Theta} p(D|\underline{\theta}, M_m) p(\underline{\theta}|M_m) d\underline{\theta} \tag{3.19}$$

where Θ is the domain of variation of the model parameters.

The likelihood $p(D|\underline{\theta}, M_m)$ is derived by using a probability model for the prediction error e_r, $r = 1, \ldots, m$, for the modal frequencies defined as the fractional difference between the measured modal frequencies and the corresponding modal frequencies predicted from the model class M_m using a particular value of the parameter set $\underline{\theta}$. Specifically, e_r satisfies the prediction error equation

$$\hat{\omega}_r = \omega_r(\underline{\theta}) + \hat{\omega}_r e_r \tag{3.20}$$

for all modes $r = 1, ..., m$. Modeling the prediction errors as zero-mean Gaussian variables, $e_r \sim N(0, \sigma^2)$, with standard deviation σ, assuming that the prediction errors are independent, and including the prediction error parameter σ into the uncertain parameter set $\underline{\theta}$, the likelihood $p(D|\underline{\theta})$ takes the form

$$p(D|\underline{\theta}, M_m) \sim \frac{1}{(\sqrt{2\pi})^m \sigma^m} exp\left[-\frac{m}{2\sigma^2} J(\underline{\theta})\right] \tag{3.21}$$

where $J(\underline{\theta})$ given by

$$J(\underline{\theta}) = \frac{1}{m} \sum_{r=1}^{m} \frac{[\omega_r(\underline{\theta}) - \hat{\omega}_r]^2}{[\hat{\omega}_r]^2} \tag{3.22}$$

represents the measure of fit between the measured modal frequencies and the modal frequencies predicted by the model.

The Bayesian framework can also be used to select the best model class among a family of alternative competitive model classes $M_1, \ldots, M_\mu$ (Beck and Yuen, 2004), used to represent the dynamics of the hanger. Using the Bayes theorem, the posterior probability $P(M_i|D)$ of the model class M_i given the data D is obtained from

$$P(M_i|D) = \frac{p(D|M_i)P(M_i)}{p(D)} \tag{3.23}$$

where $p(D|M_i)$ is the evidence of M_i, $P(M_i)$ is the prior probability of M_i and $p(D) = \sum_{i=1}^{\mu} p(D|M_i)P(M_i)$ is a normalizing constant that guaranties that the sum of the probabilities over all model classes considered in the selection equals one. Assuming that the model classes are equally probable prior to the use of the data, then the most probable model class based on the data corresponds to the model class with the highest evidence.

Bayesian computational tools are used to estimate the uncertainty in the model parameters, select the best model class and propagate uncertainty. Herein we use the TMCMC (Ching and Chen, 2007) and its parallelized extended version (Angelikopoulos et al., 2012) in order to sample from the posterior PDF of each model class, estimate uncertainties in the model parameters such as axial hanger load and

equivalent length(s), as well as propagate uncertainties to compute output quantities of interest such as modal frequencies. One more merit of using the TMCMC algorithm for Bayesian purposes is the calculation of the evidence as a by-product of the algorithm (Ching and Chen, 2007).

Note that for identifiable cases, the Gaussian approximation of the posterior PDF of the model parameters (Yuen and Mu, 2015, 2010) can also be used to save computational effort for the hanger problem. This usually requires the availability of the first and second derivatives of the minus the logarithm of the posterior PDF with respect to the model parameters. However, analytical formulations for such derivatives based on direct differentiation or adjoint methods are not readily available and such approximate formulation has not being pursued further in this work. Instead the TMCMC is very suitable for the considered model classes due to the model non-intrusiveness and the absence of analytical derivatives of the output quantities of interest with respect to model parameters (Hadjidoukas et al., 2015). The most probable values of the parameters in $\underline{\theta}$ are also obtained by minimizing the $-\log p(\underline{\theta}|D, M_m)$ using the CMA-ES algorithm (Hansen et al., 2003).

3.6 Results

The axial load is estimated using the FE model classes and the simple beam model classes introduced in the previous sections. Estimation is based on the lowest twelve experimentally identified modal frequencies of the hanger (see Table 3.1), six along the transverse and six along the longitudinal direction of the bridge deck. The objective is to estimate the value of the hanger tension and its uncertainty, to explore the effect of the end hanger conditions and the number of measured modal frequencies on the accuracy of the hanger tension estimates, and to select the best model classes based on simplified beam theory that are adequate representations of the hanger behavior.

The parameter estimation and the model class selection is performed for the model classes reported in Table 3.3. The FE model classes $FEfix$ and $FEflex$ are used to identify whether or not the flexibility in the end conditions of the hanger, arising from the flexibility of the arch and the deck, is important. The simple beam model classes $Bfix$ and $Bflex$ are used for the purpose of identifying which one of them is capable of predicting adequately the hanger tension based on the results obtained from the FE model classes which are considered to contain the more accurate information for the hanger tension. Depending on the type and number of parameters they are left free to be inferred by the Bayesian formulation, the beam model classes are further classified as follows.

- $Bfix(T, L)$: two-parameter model class, with parameters the hanger tension T and the equivalent hanger length L. This model uses the equivalent length concept (Bellino et al., 2010) so that it can adjust the length of the uniform cross-section of the beam to fit the modal frequency data, thus accounting for the flexibility of the end plate connections and the boundary conditions between the hanger and the arch or the deck.

- $Bfix(T, L_t, L_l)$: three-parameter model class, with parameters the hanger tension T and equivalent

hanger lengths L_t and L_l assumed to be different for deflections in the (x, z) (longitudinal direction) and (y, z) (transverse direction) planes. This model has freedom to provide different beam flexibilities along the transverse and longitudinal directions of motion and thus better fit the different values of the experimental modal frequencies observed along these two directions (see Table 3.1).

- $Bfix(T|L_i)$: a family of μ one-parameter model classes with parameter the hanger tension given the value of the hanger length to be L_i, $i = 1, \ldots, \mu$. Each model class is defined by the different value L_i of the beam length. The aim is to choose the best model class or, equivalently, the best equivalent hanger length L_i that best represents the observed data.

A similar classification is introduced for $Bflex$ model classes resulting in

- $Bflex(T, L, \underline{k})$: three- to six-parameter models, with parameters the hanger tension, the equivalent length, as well as the rotational stiffnesses at the two ends. The rotational stiffness to be included in $\underline{k}$ may vary from one to four. For example, $Bflex(T, L, k_1, k_3)$ denotes the model class with the rotational stiffness k_1 and k_3 (see Figure 3.3) used as free parameters to be determined, while the k_2 and k_4 are set fixed to 10^{15} to simulate rigid supports at the respective directions.

The nominal value of the axial load is considered to be $T_0 = 922KN$ corresponding to the most probable value of the fixed-end beam model based on the lowest two experimental frequencies, one in the transverse and the other in the longitudinal direction. The nominal value of the hanger length is $L_0 = 12m$, corresponding to the total length that includes the approximately $2m$ length of the plate connectors (Figure 3.2) at the two ends. The parameters θ_T and θ_L in the set $\underline{\theta}$, introduced for the axial load T and the length L of the beam, respectively, scale the nominal values so that the axial load is $T = \theta_T T_0$ and the length is $L = \theta_L L_0$. The nominal values of the spring stiffnesses used to simulate boundary conditions are taken to be $k_{nom} = 10^{10}$. For a spring stiffness k, the corresponding parameter θ_k in the set $\underline{\theta}$ is introduced so that $k = (k_{nom})^{\theta} = 10^{10\theta_k}$.

The prior distributions of all parameters are selected to be uniform. The bounds are selected to be $\theta_T \in [0.5, 1.5]$ for the hanger tension parameter, $\theta_L \in [0.817, 1.2]$ for the length of the beam model, and $\theta_k \in [0.1, 1.5]$ for all the spring stiffness parameters. Note that the lower bound 0.817 for θ_L corresponds to the clear length $L_{cl} = 0.817 * 12 = 9.817$m of the circular part of the hanger. The range of variation of the prediction error parameter σ is $\sigma \in [0.001, 0.1]$. The TMCMC algorithm (Ching and Chen, 2007; Angelikopoulos et al., 2012) is used to sample the posterior PDF of each model class, compute the uncertainties in the model parameters, estimate the evidence of each model class, and propagate uncertainties to predictions of the modal frequencies. The values of the TMCMC parameters are selected to be $\beta^2 = 0.2$ and $TolCov = 1.0$. The most probable values of the model parameters are obtained using the CMA-ES algorithm (Hansen et al., 2003) with the search domain to be the one defined by the support of the uniform priors assumed for the model parameters.

Table 3.3 presents the most probable value (MPV), the mean, the standard deviation, and the 5%, 50% and 95% quantiles of the hanger tension estimated from all model classes. The number of samples per TMCMC stage is 1000 for $FEflex$, 500 for $FEfix$, 5000 for $Bfix$ and $Bflex$ model classes.

The different number of samples per TMCMC stage used for each model class is chosen as a trade-off between computational effort and solution accuracy. A large number of samples is used for the analytical beam model due to the small time-to-solution, of the order of a few minutes, in relation to the time-to-solution required for estimating the hanger force using the finite element model classes. A smaller number of 500 samples per TMCMC stage is used for the FEfix model class due to a small number of two parameters involved as opposed to six parameters involved in the FEflex model class. Representative times-to-solution obtained in a dual 4-core computer using the parallelized versions of TMCMC algorithm are of the order of 60 hours for $BEflex$, 12 hours for $BEfix$ and several minutes for the $Bfix$ and $Bflex$ model classes. The number of TMCMC stages, which affect the time-to-solution is approximately 4 to 7, depending on the model class and the individual run.

3.6.1 Hanger force estimation based on FE model classes

Comparing the log evidence for the FE model classes $FEflex$ and $FEfix$ it can be seen that the $FEflex$ is clearly the preferred model since the resulting relative probabilities of the two models are $Pr(FEflex|D) = a/(1+a) = 0.86$ and $Pr(FEfix|D) = 1/(1+a) = 0.14$, where $a = exp(24.03 - 22.24)$ is evaluated from the log evidence values in Table 3.3. The $FEflex$ gives a significantly better fit to the data which is equal to 1.03% as compared to the fit 2.63% for the $FEfix$ model. The two FE models give completely different predictions of the hanger tension. Model $FEflex$ predicts the most probable value at 0.91 with uncertainty as quantified by the quantiles to be in the range $[0.75, 1.04]$ (std=0.090), while the $FEfix$ predicts the most probable value to be 22% lower at 0.70 with uncertainty to be in the range $[0.60, 0.80]$ (std=0.063).

To identify the source of such differences the results of the models $FEfix(long)$ and $FEfix(trans)$ are used. The $FEfix(long)$, which is based on fitting the six modal frequencies with dominant hanger deflections along the longitudinal direction, give predictions of the hanger tension that are closer to those obtained by the $FEflex$. In contrast, the $FEfix(trans)$ which is based on fitting the six modal frequencies with dominant hanger deflections along the transverse direction gives predictions that are approximately 35% lower than the $FEflex$ model predictions and 15% lower than the $FEfix$ model predictions. These differences in predictions are due to the fact that the experimentally identified modal frequencies of the hanger in the transverse direction are consistently lower than the modal frequencies of the hanger in the longitudinal directions (see Table 1) which, due to the symmetry of the hanger and the orientation of the top and bottom identical plate connections in the transverse and longitudinal directions, it can only be explained by the flexibility of the arch and/or the deck when the hanger vibrates in the transverse direction. This support flexibility was accounted in the model class $FEflex$ while in the model class $FEfix(trans)$ it was ignored resulting in a significantly lower hanger tension in order to compensate and match the lower modal frequencies. The $FEfix$ models class which did not allow for support flexibility in both directions also resulted in lower hanger tension, trading-off the fit of the modal frequencies in the longitudinal and transverse direction.

The most probable values of the rotational stiffnesses are estimated from the $BEflex$ model class to be $k_1 = 10^8$, $k_2 = 10^{11}$, $k_3 = 10^4$ and $k_4 = 10^{13}$, with the value of k_3 being significantly smaller than the other 3 much stiffer rotational springs, indicating that the main source of support flexibility is at the top end of the hanger in the transverse direction and it is due to the flexibility of the bridge arch. In conclusion, the 5-parameter model $FEflex$ is the most preferred model class as compared to the one-parameter $FEfix$ model class and predicts more reliably the hanger force, while the $FEfix$ underestimates the hanger force by approximately 22%.

Table 3.3: Log evidence and estimates of hanger tension for all model classes

Model class	$N_{\underline{\theta}}$	Log Evidence	Tension quantiles [5%, 50%, 95%]	Tension Mean	Tension Standard Dev.	Tension MPV	Fit
$FEflex$	5	24.03	[0.75, 0.92, 1.04]	0.909	0.0897	0.822	0.0103
$FEfix$	1	22.24	[0.59, 0.70, 0.80]	0.701	0.0627	0.701	0.0263
$FEfix(long)$	1	–	[0.75, 0.83, 0.92]	0.836	0.0498	0.831	0.0102
$FEfix(trans)$	1	–	[0.52, 0.60, 0.75]	0.616	0.0714	0.576	0.0237
$Bfix(T, L)$	2	20.72	[0.68, 0.86, 1.06]	0.867	0.114	0.854	0.0248
$Bfix(T, L_t, L_l)$	3	22.49	[0.73, 0.86, 0.98]	0.856	0.0731	0.856	0.0115
$Bflex(T, L, k_3)$	3	20.90	[0.70, 0.88, 1.08]	0.883	0.1215	0.888	0.0112
$Bflex(T, L, k_1, k_3)$	4	20.10	[0.70, 0.94, 1.35]	0.974	0.1905	0.888	0.0112
$Bflex(T, L, \underline{k})$	6	19.23	[0.77, 1.19, 1.44]	1.144	0.2196	0.968	0.0115

From the values of the standard deviation and the 5% and 95% quantiles of the hanger tension predicted from the $FEflex$ model, it can also be concluded that the uncertainty in the hanger tension is of the order of 10%. This uncertainty is due to the inability of the model class $FEflex$ to fit exactly all 12 modal frequencies. This uncertainty in the hanger force prediction should be taken into account when using the values of the hanger tension to infer structural safety. In contrast to existing studies that report a single value of the hanger force, the Bayesian inference framework also estimates the uncertainty in such a value. This uncertainty may affect the safety margins of the hanger and the bridge. From the engineering point of view, using the nominal hanger tension value of $T_0 = 922KN$ and the hanger circular cross-sectional area of diameter $D = 0.130m$, one has that the stress prediction in the hanger within 5% and 95% credible intervals ranges between $[0.16, 0.22]\sigma_y$ values, where $\sigma_y = 330MPa$ is the yield stress of this specific hanger, resulting in a relative high safety factor (well within the safe domain when failure is assumed to occur due to stresses in the hanger exceeding material yield or fracture stresses).

3.6.2 Hanger force estimation based on simple beam model classes

The ability of the simplified beam model classes to predict the hanger tension is next examined. From the results in Table 3.3 it can be observed that almost all simplified beam models make predictions of

the value of the hanger tension and its uncertainty that are significantly closer to the predictions of the $FEflex$ model than the predictions of the $FEfix$ model. This is very promising for using such simplified models for hanger tension predictions.

Comparing the log evidence values for the fixed support model classes $Bfix(T, L)$ and $Bfix(T, L_t, L_l)$, it can be clearly seen that the 3-parameter model class $Bfix(T, L_t, L_l)$ has higher preference than the two-parameter model class $Bfix(T, L)$ with $Pr(Bfix(T, L_t, L_l)|D) = 0.85$ and $Pr(Bfix(T, L)|D) = 0.15$. The error for the most probable parameter values (last column in Table 3.3) obtained by CMA is 1.1% for the $Bfix(T, L_t, L_l)$ model which should be compared to the 2.6% error for the $Bfix(T, L)$ model. Both models provide a sample mean estimate of the hanger force which is approximately 5% less than the mean estimate of the $FEflex$ model class. The 5% and 95% credible intervals quantifying the uncertainty in the hanger force are predicted by both beam models to be close to the corresponding uncertainty bounds predicted by the $FEflex$, with the $Bfix(T, L)$ to slightly underestimate the lower bound, while the $Bfix(T, L_t, L_l)$ to slightly underestimate the upper bound.

To assess the correlation between the hanger tension and the equivalent beam length, the projection of the posterior samples obtained using the TMCMC algorithm in the two-dimensional space (θ_T, θ_L) of the model parameters is presented in Figure 3.6 for the model class $Bfix(T, L)$. We note that the length of the beam has a positive correlation with the axial load of the beam which is expected since when the length is increased, the frequencies tend to decrease. As a result the hanger tension and thus the predicted modal frequency must be increased to compensate this decrease, maintaining the fit with the values of the experimental modal frequencies.

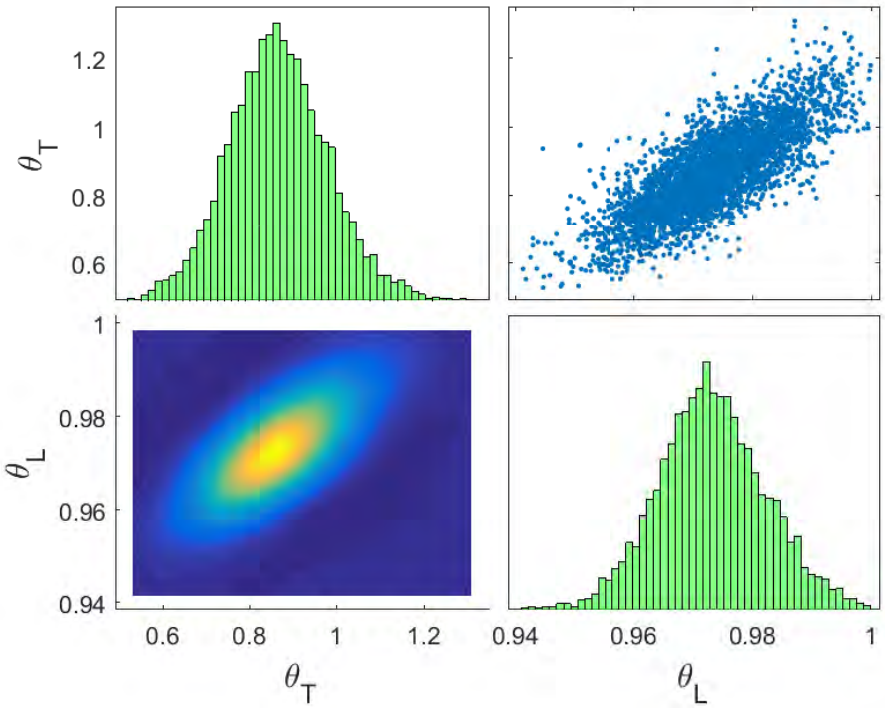


Figure 3.6: Parameter estimation for model class $Bfix(T, L)$. Diagonal: Marginal distributions. Above-diagonal: sample projections in (θ_T, θ_L) parameter space. Below Diagonal: Contour plots in (θ_T, θ_L) parameter space

Concerning the flexible-end beam model classes $Bflex(T, L, k_3)$, $Bflex(T, L, k_1, k_3)$ and $Bflex(T, L, \underline{k})$, the log evidence values suggest that the least-parameter model class $Bflex(T, L, k_3)$ is preferred to the 4-parameter and 6-parameter model classes since the addition of the extra parameters in the model is penalized by the Bayesian formulation, considered it as overfitting (Beck and Yuen, 2004). The 3-parameter model class $Bflex(T, L, k_3)$ provides reasonable uncertainty bounds of $[0.70, 1.08]$ of the hanger force, that are closer to the ones provided by the two-parameter model $Bfix(T, L)$, while the error in the fit is 1.1% which is closer to the error provided by the $Bfix(T, L_t, L_l)$ model. This means that the $Bflex(T, L, k_3)$ fits better the experimental data than the $Bfix(T, L)$ model. The models $Bflex(T, L, k_1, k_3)$ and $Bflex(T, L, \underline{k})$, which are less preferred mainly due to overparameterization, predict much higher mean hanger tension and uncertainty bounds. This can be attributed to the flexibility that this model has to provide a reasonable fit to the data by trading-off the hanger force values with the flexibility of the springs at the end supports.

Comparing the $Bfix$ and $Bflex$ model classes, the most preferred model corresponding to the largest log evidence value is $Bfix(T, L_l, L_t)$ with the second and third preferred model classes to be $Bflex(T, L, k_3)$ and $Bfix(T, L)$, respectively. Comparing with the $FEflex$ model results, the most pre-

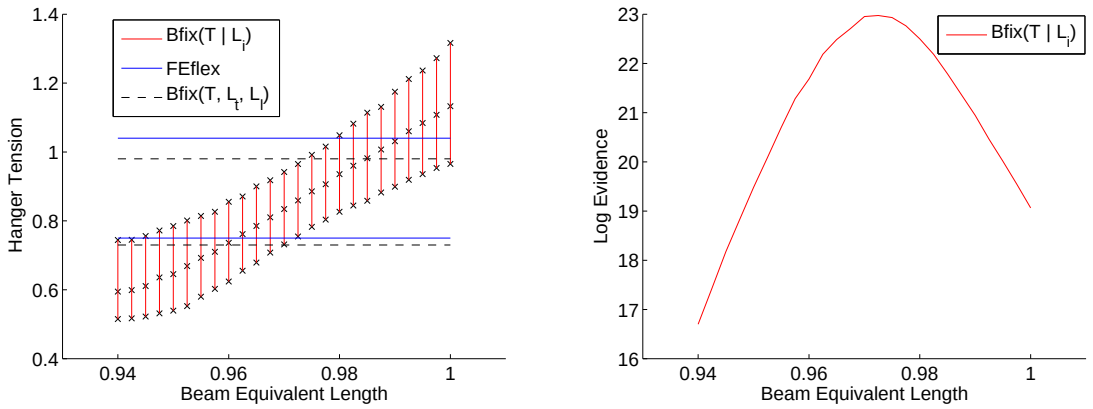
ferred $Bfix(T, L_l, L_t)$ model gives slightly tighter uncertainty bounds for the hanger force. The worst fit in the experimental frequencies is accomplished by the two-parameter $Bfix(T, L)$ model due to its less flexibility with a single equivalent hanger length to simultaneously fit the modal frequencies in the transverse and longitudinal directions. In contrast, model $Bflex(T, L, k_3)$ is flexible to simultaneously fit the longitudinal modal frequencies by adjusting the beam length, and the transverse modal frequencies by adjusting the spring stiffness k_3.

Table 3.4 presents the 5% and 95% quantiles, the mean and the standard deviation of the equivalent length estimated from each simple beam model class introduced in Table 3.3. It can be seen that the identified uncertainty in the equivalent length values is very narrow. Comparing the fixed-support model classes, $Bfix(T, L)$ gives uncertainty bound of $[0.959, 0.988]L_0$, while $Bfix(T, L_t, L_l)$ gives lower values $[0.949, 0.972]L_0$ for L_l and higher values $[0.972, 995]L_0$ for L_t to make the beam more flexible to fit the lower modal frequencies in the transverse direction. The coefficient of variation (cov = standard deviation over mean) for the $Bfix(T, L)$ model is 0.9%, while for the $Bfix(T, L_t, L_l)$ model it is approximately 0.7% for both L_t and L_l. All these uncertainties are quite small compared to the uncertainty of 10% predicted for the hanger tension.

Table 3.4: Estimates of the hanger equivalent length for all beam model classes

Model class	Length quantiles [5%, 50%, 95%]	Length Mean	Length Standard Dev.
$Bfix(T, L)$	[0.959, 0.973, 0.988]	0.975	0.0090
$Bfix(T, L_t, L_l)$ $-L_t$	[0.972, 0.984, 0.995]	0.984	0.0070
$-L_l$	[0.949, 0.960, 0.972]	0.960	0.0071
$Bflex(T, L, k_1)$	[0.953, 0.970, 0.986]	0.970	0.0101
$Bflex(T, L, k_1, k_2)$	[0.938, 0.965, 0.985]	0.964	0.0138
$Bflex(T, L, \underline{k})$	[0.923, 0.950, 0.980]	0.951	0.0178

Figure 3.7a presents the results of the estimation of the hanger force by using the family of models $Bfix(T|L_i)$ with L_i varying within the uncertainty bound computed by the $Bfix(T, L)$ model. Figure 3.7b gives the evidence of each model in the family. It is clear that the best model class is the one that corresponds to length $L_{best} = 0.973$ and predicts hanger tensions that are consistent with the ones predicted by $Bfix(T, L_t, L_l)$, with narrower uncertainties since it does not take into account the uncertainty in the equivalent length. However, as one moves away from the best values of the equivalent length, the predictions of hanger tension values and uncertainties depart considerably (underpredicting or overpredicting) from the predictions of the baseline FE model $FEflex$ or $Bfix(T, L_t, L_l)$. This means that arbitrary guesses of the equivalent length to carry out the identification of the hanger tension are highly likely to give erroneous estimates. The proposed Bayesian method is a rational framework to provide the correct estimates of the hanger tension and equivalent lengths, as well as their uncertainties.



(a) (b)

Figure 3.7: (a) Comparison of hanger tension 5%, 50% and 95% quantiles estimated by the model classes $Bfix(T|L_i)$ with 5% and 95% quantiles estimated by the model classes $FEflex$ and $Bfix(T, L_t, L_l)$. (b) Log evidence values of model classes $Bfix(T|L_i)$ for different L_i values

Finally, the estimates of the equivalent length L provided by the $Bflex(T, L, k_3)$ model are closer and slightly larger than the estimates of L_l for the $Bfix(T, L_t, L_l)$ model so that the experimentally identified longitudinal modal frequencies are matched, while the stiffness k_3 of the $Bflex(T, L, k_3)$ model is adjusted to match the modal frequencies in the transverse direction. The uncertainty in the equivalent length L values is higher (cov=1.0%). The uncertainties in the equivalent length L predicted by the four parameter model $Bflex(T, L, k_1, k_3)$ and the seven-parameter model $Bflex(T, L, \underline{k})$ are even higher (1.5% and 1.9%, respectively), due to the flexibility of these models to compensate the change in the equivalent length by a change in the rotational stiffness values in order to fit the identified modal frequencies. In any case even the highest uncertainty of 1.9% is small, indicating the narrow range of values that the equivalent length can take in order to fit the measured data.

Figure 3.8 presents the propagation of the uncertainty in the model parameters to the lowest twelve modal frequencies. The results in this Figure correspond to predicted modal frequencies normalized with respect to the experimental frequencies. Thus the distance of these values from one is a measure of how close the values of the model predicted modal frequencies are to the experimentally identified modal frequencies. We can clearly see the impact of the flexible end supports of the $FEflex$ model in predicting the modal frequencies compared to the fixed-end model class $FEfix$. The fixed-end FE model class cannot reliably predict the measured modal frequencies, missing at least seven of them by as much as 4%. The flexible-end FE model class reliably predicts all measured modal frequencies, since the line equal to 1 is within the uncertainty bounds of the predictions for most modal frequencies. We also notice that the 3-parameter beam model classes $Bfix(T, L_t, L_l)$ and $Bfix(T, L, k_3)$ also give good predictions of the modal frequencies. The predictions of the $Bfix(T, L_t, L_l)$ model class based on the two equivalent lengths, independent in the transverse and longitudinal directions, give very similar predictions of the

modal frequencies and their uncertainties to those obtained from the $FEflex$ model class for several of the modal frequencies. The $Bfix(T, L, k_3)$ also gives similar predictions but with higher uncertainty. The predictions of the 2-parameter model class $Bfix(T, L)$ are closer to those of the fixed-end $FEfix$ model than to $FEflex$ model, with much higher uncertainty so that a number of measured modal frequencies are contained within the uncertainty bounds. From the simple beam model classes, the best predictions are obtained from model class $Bfix(T, L_t, L_l)$.

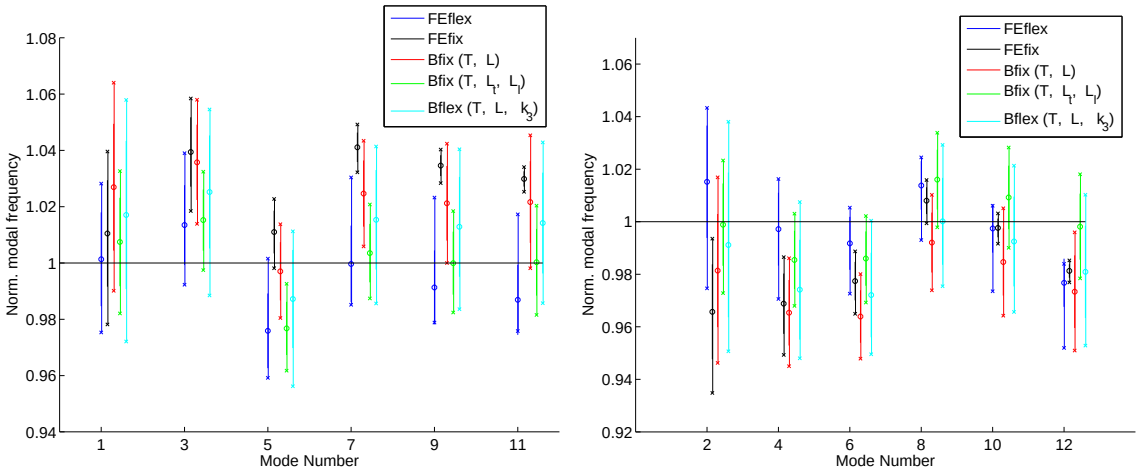
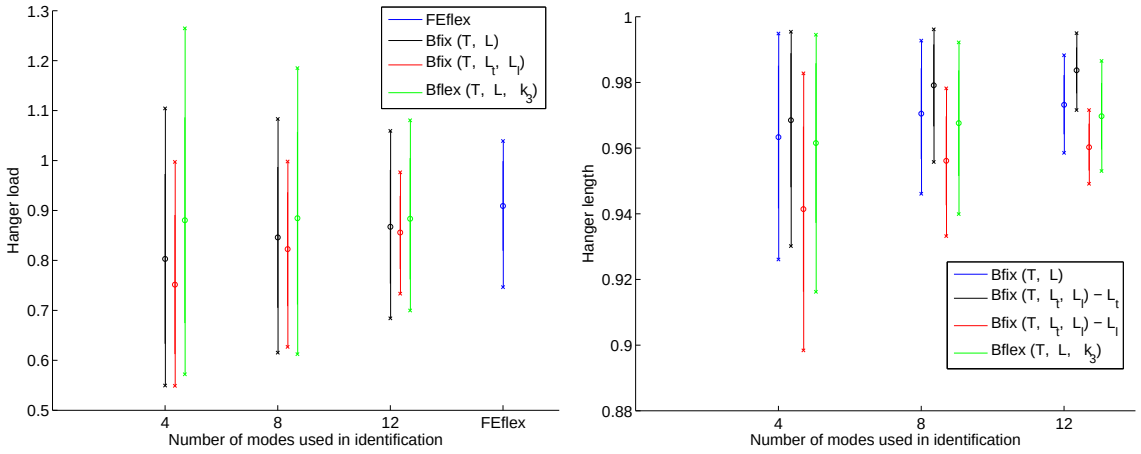


Figure 3.8: Uncertainty propagation to the output frequencies for selected model classes

To study the effect of the number of modal frequencies used in the identification, Figure 3.9a presents results for the hanger force (mean and quantiles values) estimated from the model classes $Bfix(T, L)$, $Bfix(T, L_t, L_l)$ and $Bflex(T, L, k_3)$ using the lowest $m = 4$ or 8 or 12 identified modal frequencies. Given m modal frequencies, $m/2$ correspond to the longitudinal direction and $m/2$ to the transverse direction. Results are also compared to the hanger force mean and quantile values estimated by the $FEflex$ model class. Similar results for the effective length mean and quantile values are presented in Figure 3.9b. We can clearly see the impact of the number of modes on the predictions of the hanger force uncertainties and structural reliability. As the number of modes used in the Bayesian identification decreases from 12 to 8 or 4, the uncertainty in the hanger tension and equivalent length values increases substantially for all beam model classes. This uncertainty increase is due to inadequacy of the small number of measured modal frequencies to identify with certainty the hanger tension and the equivalent lengths. When the hanger tension uncertainty is further used for structural safety estimation, the smaller number of measured modes leads to less reliable estimates of structural safety corresponding to higher failure probability than it is actually obtained from the identification that is based on a higher number of measured modal frequencies.



(a) (b)

Figure 3.9: (a) Hanger tension mean and $[5\%, 95\%]$ quantiles and (b) hanger equivalent length mean and $[5\%, 95\%]$ quantiles for different model classes for 4, 8 and 12 modes used for identification

3.7 Conclusions

Bayesian inference is used to quantify and calibrate the uncertainties in the tension of bridge hangers by integrating the information from a structural model and the experimentally identified modal frequencies. A number of competitive structural model classes used for representing the vibrational characteristics of hangers are investigated for their accuracy, including simple Euler-Bernoulli beam models as well as high-fidelity FE models. The effect of the hanger end connection details and boundary conditions due to flexibility of arch and deck substructures on the hanger tension predictions is examined.

The high fidelity FE model with flexible supports, expected to give the most reliable estimates, is able to predict the lowest 12 experimentally identified modal frequencies in both transverse and longitudinal directions. The uncertainty in the hanger force is of the order of 10% of its mean value and should be taken into account in structural safety considerations. The FE model with fixed supports fails to fit the experimentally identified modal frequencies, underpredicting the hanger tension value and its uncertainty by more than 20%. Results suggest that the connection details and boundary conditions due to the flexibility of deck and arch substructures affect significantly the dynamics of the considered hanger. The flexibility arises mainly from the flexibility of the arch due to transverse deflections of the hanger.

The dynamics of the hanger in the longitudinal and transverse directions of the bridge seem to be reasonably approximated by simplified beam models using the concept of two independent equivalent lengths in the transverse and longitudinal direction, with values left free to be estimated by the Bayesian approach. The fixed-end beam model with the two independent equivalent lengths is selected by the Bayesian framework as the most preferred model class. It provides accurate estimates of the hanger tension and the uncertainties, close to the ones provided by the baseline FE model with flexible supports.

The beam model with a single equivalent length fails to give adequate estimates of the measured modal frequencies for the specific hanger due to different modal frequencies arising from the asymmetry of the boundary conditions. The simplified beam model with a single equivalent length and end rotational springs is able to adequately represent the dynamics of the hanger, giving fairly accurate results with higher uncertainty in the values of the hanger tension, an indication that the model is less preferred than the fixed-end model with two equivalent lengths. A model with fixed length value selected arbitrarily is highly unlikely to fit the measured frequency data due to the fact that the dynamics of the beam is significantly affected by the selected value of the length of the beam. There is only a very narrow range of length values that fit the measured modal frequencies and give accurate predictions of the hanger tension.

The use of equivalent length and/or rotational springs to simulate hanger support flexibilities is deemed important in the beam modeling. Results suggest that the fixed-end models are substantially less accurate than the ones that take flexibility into account, either in terms of two equivalent lengths introduced to independently model flexibilities in transverse and longitudinal directions or in terms of a combination of an equivalent length and rotational springs. The Bayesian approach applied on the simplified beam models provides sufficiently accurate estimates of the hanger tension and its uncertainty, requiring three orders of magnitude less computational effort than the high-fidelity FE models. As the number of available identified modal frequencies increases, the prediction accuracy of the simple beam models is improved, while the uncertainty in the hanger tension is reduced. In contrast to inverse methods based on estimating a single value of the hanger tension, it is demonstrated in this study that the uncertainties in the hanger tension can be significant and should be considered in inverse methods since they affect predictions of structural reliability and safety.

Chapter 4

Model updating of Metsovo bridge

4.1 Introduction

The evaluation of the actual dynamic characteristics of bridges, such as modal frequencies, modal damping ratios and mode shapes, through vibration measurements, as well as the development of high-fidelity finite element (FE) models, has been attracting an increasing research effort worldwide. Measured response data of bridges mainly under ambient vibrations offer an opportunity to study quantitatively and qualitatively their dynamic behavior. These vibration measurements can be used for estimating the modal properties of the bridges, as well as for updating the corresponding FE models used to simulate their behavior (Friswell and Mottershead, 1995; Sehgal and Kumar, 2015). The information for the calibrated FE models and their associated uncertainties is useful for checking design assumptions, for validating the assumptions used in model development, for improving modeling and exploring the adequacy of the different classes of FE models, for identifying possible soil-structure interaction effects, and for carrying out more accurate robust predictions of structural response. These models are representative of the initial structural condition of the bridge and can be further used for structural health monitoring purposes (Vanik and Beck, 1998; Vanik et al., 2000; Yuen, 2010; Yuen et al., 2006a).

Bayesian methods for ambient (operational) modal identification (Yuen et al., 2002; Au, 2012a,b; Au et al., 2012; Au and Zhang, 2012; Au et al., 2013; Zhang and Au, 2016; Zhang et al., 2016; Ni et al., 2016, 2017) and structural model updating (Beck and Au, 2002; Beck and Katafygiotis, 1998; Katafygiotis and Beck, 1998; Katafygiotis et al., 1998; Beck, 2010; Yuen et al., 2006b; Cheung and Beck, 2009; Muto and Beck, 2008; Goller and Schueller, 2011; Yuen, 2012; Lam et al., 2015; Yan and Katafygiotis, 2015) are used to develop high fidelity FE models of bridges using modal properties identified from ambient vibration measurements. Due to the large size of the bridge, the mode shapes of the structure are assembled from a number of sensor configurations that include optimally placed reference sensors as well as moving sensors (Au, 2011). The modal properties are then integrated within Bayesian model updating formulations to calibrate the parameters of large scale FE models as well as their associated uncertainty. The goal is to

develop accurate and reliable models of the actual structures that are proven to closely simulate their behavior.

As far as the computational part is concerned, for complex posterior distributions, stochastic simulation algorithms such as the Transitional Markov Chain Monte Carlo (TMCMC) (Ching and Chen, 2007) can be conveniently used to sample from the posterior distribution for parameter estimation, model selection and uncertainty propagation purposes. These methods require a large number of forward model runs which can increase the computational effort to excessive levels if one simulation for a high-fidelity large-order FE model requires several minutes or even hours to complete. For that purpose, fast and accurate component mode synthesis (CMS) techniques, consistent with the FE model parameterization (Papadimitriou and Papadioti, 2013; Jensen et al., 2014), are used to achieve drastic reductions in computational effort. Further computational savings are achieved by adopting a parallelized version of the TMCMC algorithm to efficiently distribute the computations in available multi-core CPUs (Angelikopoulos et al., 2012; Hadjidoukas et al., 2015).

The use of the Bayesian uncertainty quantification (model updating), propagation and model selection methodologies is demonstrated by updating the parameters of a high fidelity FE model developed for the Metsovo bridge, using modal properties experimentally identified from ambient vibration data. The FE model is parametrized with respect to the stiffnesses of various structural components. Ambient acceleration time histories from multiple points in the structure are used to extract the modal properties of the bridge experimentally, and the identified modal properties are used as data in the Bayesian model updating methodologies in order to perform inference about the model parameters. Two classes of models are examined and compared using Bayesian model selection (Beck and Yuen, 2004; Muto and Beck, 2008). The models explore the effect of the soil-structure interaction. The updated model is validated using a subset of the experimental data that was not used in the updating procedure in order to test its accuracy and reliability as a model of the actual structure.

4.2 Description of the bridge

The ravine bridge of Metsovo in section 3.2 (Anthohori-Anilio tunnel) of Egnatia Motorway is crossing the deep ravine of Metsovitikos river, 150m over the riverbed. This is the highest bridge of the Egnatia Motorway, with the height of the tallest pier equal to 110m. The total length of the bridge is 537m. The bridge has 4 spans of length 44.78m, 117.87m, 235m, 140m and 3 piers of which M1(45m) supports the boxbeam superstructure through pot bearings (movable in both horizontal directions), while M2(110m) and M3(35m) piers connect monolithically to the structure. A sketch of the bridge is shown in Figure 4.1 while photos of the actual bridge are shown in Figures 4.2 and 4.3.

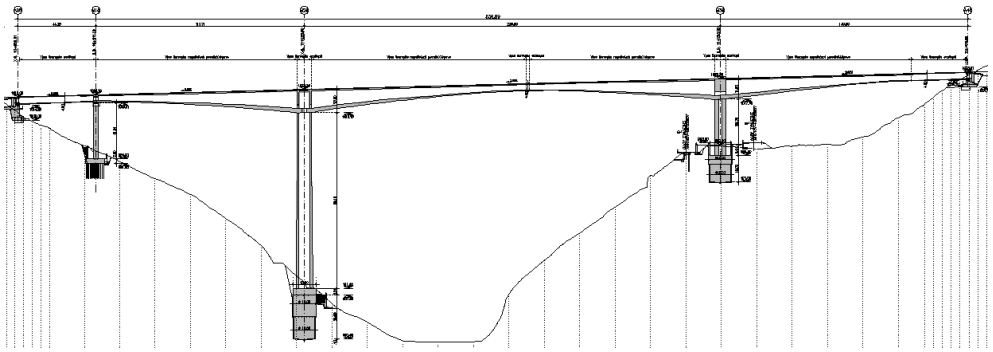


Figure 4.1: Longitudinal view of the Metsovo ravine bridge

Figure 4.2: The ravine bridge of Metsovo

Figure 4.3: The ravine bridge of Metsovo

The detailed geometry of the bridge is complicated because the piers have variable cross-sections and the deck is also inclined. A detailed design of the bridge was done by Kelantonis (2010) where one can find detailed instructions on how to design the 3D model of the Metsovo bridge in SolidWorks from 2D drawings. In Figures 4.4 and 4.5 the final 3D model is shown.

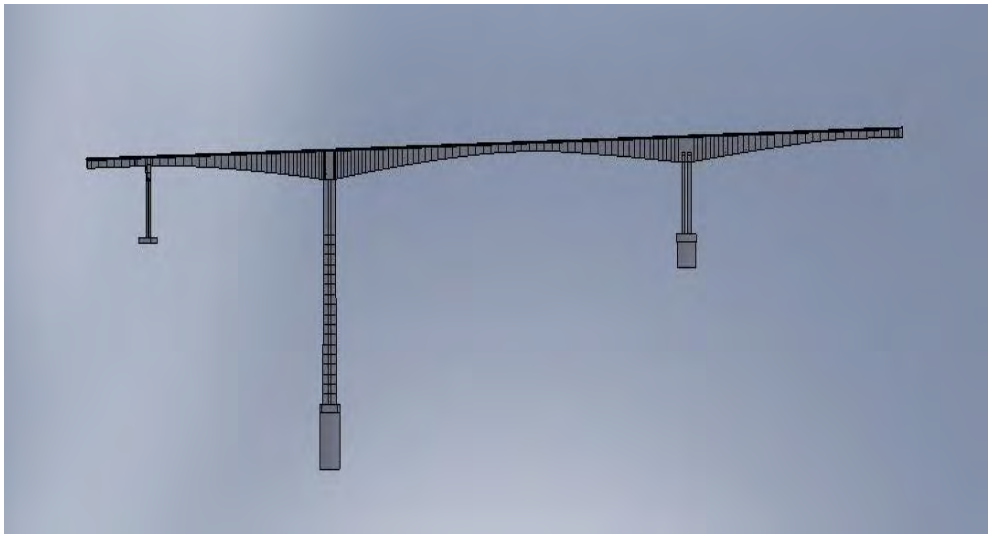


Figure 4.4: 3D model in SolidWorks

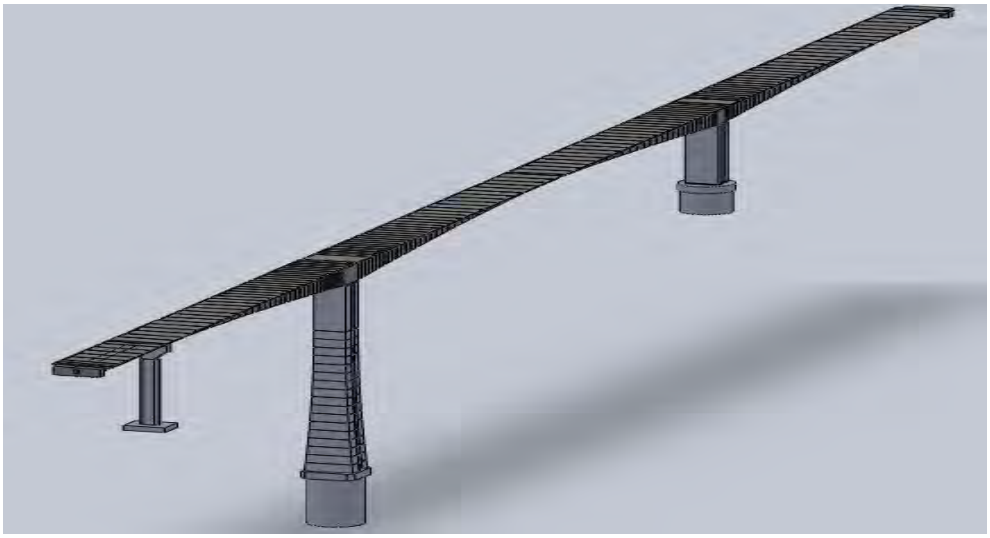


Figure 4.5: 3D model in SolidWorks

The 3D SolidWorks model is then imported into COMSOL Multiphysics for finite element analysis, and specifically for eigenfrequency analysis.

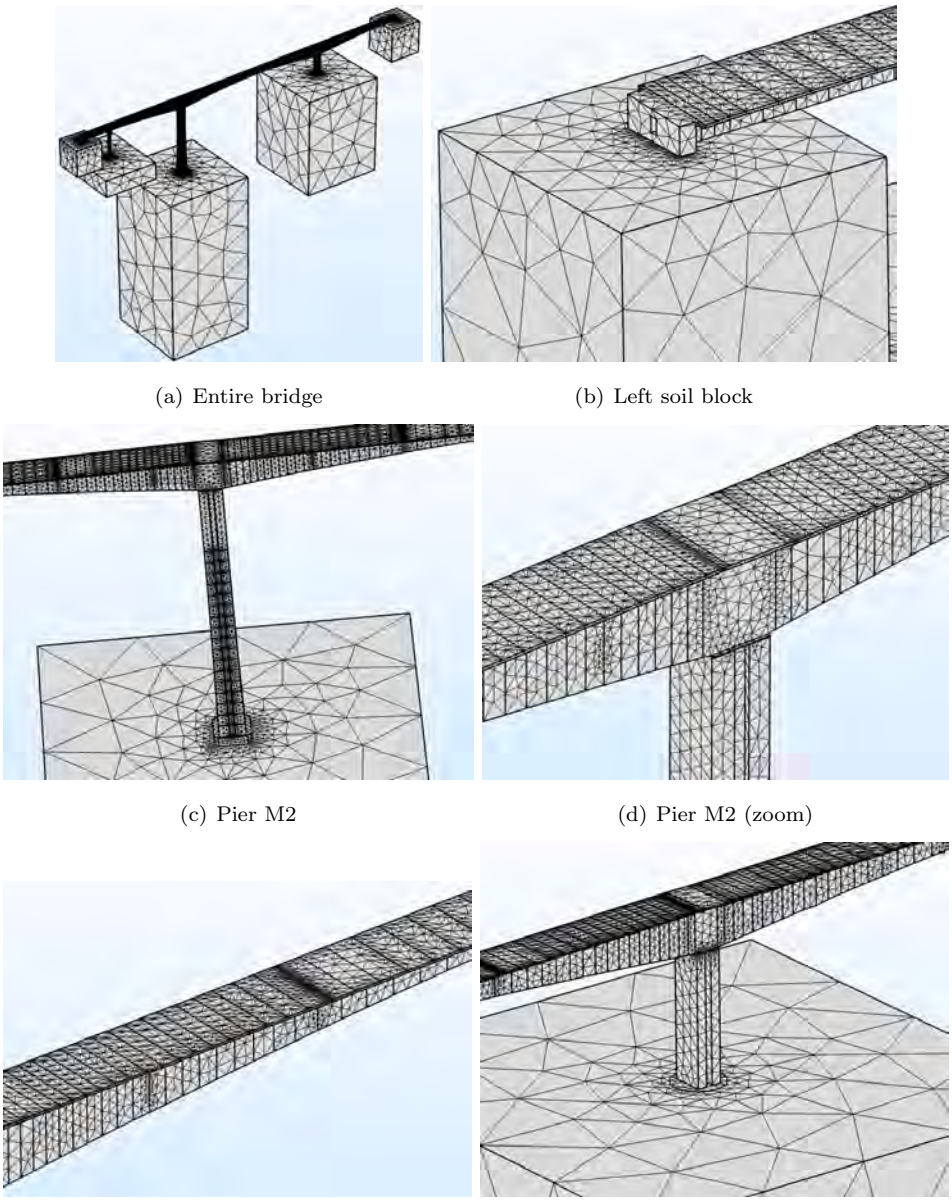
4.3 Computational model of the bridge

4.3.1 Finite element model

The FE model of the bridge is created using 3-dimensional tetrahedral quadratic Lagrange finite elements. The model takes into account the potential soil-structure interaction by modeling the soil with large blocks of material and embedding the piers and abutments into these blocks. The 3 parameters of the model are associated with the modulus of elasticity of the deck, piers and soil respectively. Specifically, the parameters multiply the nominal values of the corresponding moduli of elasticity. The nominal values of the moduli of elasticity of the deck, piers and soil were selected to be the values used in design, which are $37GPa$ for the deck, $34GPa$ for the piers and $1GPa$ for the soil. Therefore, the updated model parameters are expected to lie close to 1 for the deck and pier parameters. The updated value of the soil parameter will show if the soil-structure interaction is significant (the soil is flexible) or if the soil is much more stiffer than the bridge and behaves almost as a rigid body. In the later case, a large value of the soil parameter is expected which would indicate that the soil completely restrains the bridge movement.

A coarse FE mesh is chosen to predict the first 20 modal frequencies and mode shapes of the bridge with sufficient accuracy. The largest size of the elements in the mesh is of the order of the thickness of the deck cross-section. Several mesh sizes were tried, and an accuracy analysis was performed in order to find a reasonable trade-off between the number of degrees of freedom (DOF) of the model and the accuracy in the predicted modal frequencies. By trying different mesh sizes in the deck, piers, and soil

blocks, a mesh of 830.115 DOFs was kept for the bridge-soil model. This mesh was found to cause errors of the order of $0.1 - 0.5\%$ in the first 20 modal frequencies, compared to the smallest possible mesh sizes which had approximately 3 million DOFs. In that way, the model was optimized with respect to the number of DOFs using a variable element size in each part of the bridge. This can be noted especially in Figure 4.6 where it is evident that the size of the elements grows larger in the soil blocks.



(a) Entire bridge

(b) Left soil block

(c) Pier M2

(d) Pier M2 (zoom)

(e) Deck

(f) Pier M3

Figure 4.6: Finite element mesh of the bridge with soil blocks.

The time required for a complete run of the FE model is approximately 2 minutes on a 8-core 3.20 GHz computer. Due to the thousands of forward model runs for different values of the model parameters that are required by the Bayesian computational tools, it is necessary to reduce the time required for a single model run to just a couple of seconds in order to bring the computational effort to manageable levels.

4.3.2 Model reduction using CMS

Model reduction is used to reduce the model size and thus the computational effort to manageable levels. Specifically, the parameterization-consistent CMS technique (Papadimitriou and Papadioti, 2013; Jensen et al., 2014) is applied in the bridge-soil FE model.

Let ω_c be the cut-off frequency which represents the highest modal frequency that is of interest in FE model updating. Herein, the cut-off frequency is selected to be equal to the 20th modal frequency of the nominal FE model, i.e. $\omega_c = 3.51Hz$. The bridge is divided into 16 physical components with 15 interfaces between the components as shown in Figure 4.7. The DOFs of components are called internal DOFs whereas the DOFs of interfaces are called interface DOFs. Note that components 6 and 10 correspond to the left and right abutments respectively, which connect the bridge deck with the soil blocks. For each component, it is selected to retain all modes that have frequency less than $\omega_{\max} = \rho\omega_c$, where the ρ values affect the computational efficiency and accuracy of the CMS technique. The total number of DOFs before the model reduction is applied and the number of DOFs retained for $\rho = 5$ are given in Table 4.1. A total of 170 internal DOFs out of the 814.080 are retained for all 16 components. The total number of DOFs of the reduced model is 16.205 which also includes 16.035 interface DOFs. It is clear that more than an order of magnitude reduction in the number of DOFs is achieved using CMS. Table 4.1 also shows the largest fractional error between the modal frequencies computed using the complete FE model and the ones computed using the CMS technique for $\rho = 5$. A very good accuracy is achieved where it is seen that the error falls below 0.2%.

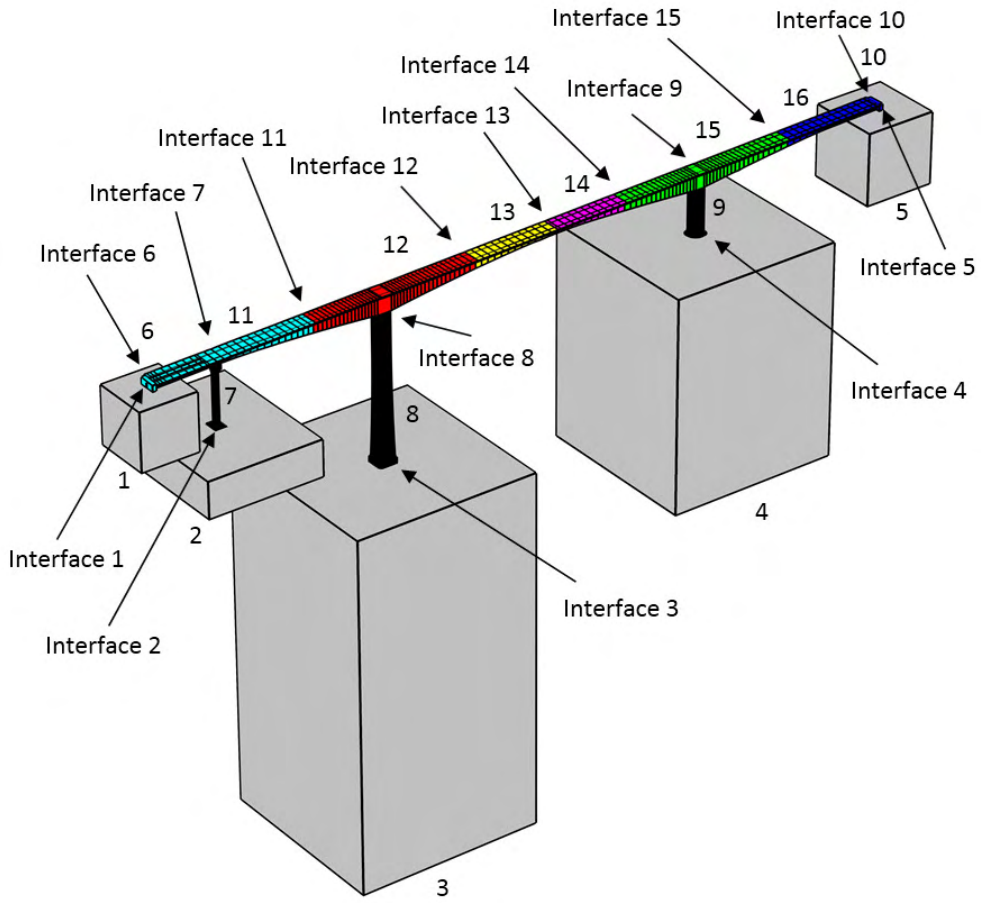


Figure 4.7: Components of the FE model of the bridge-soil system.

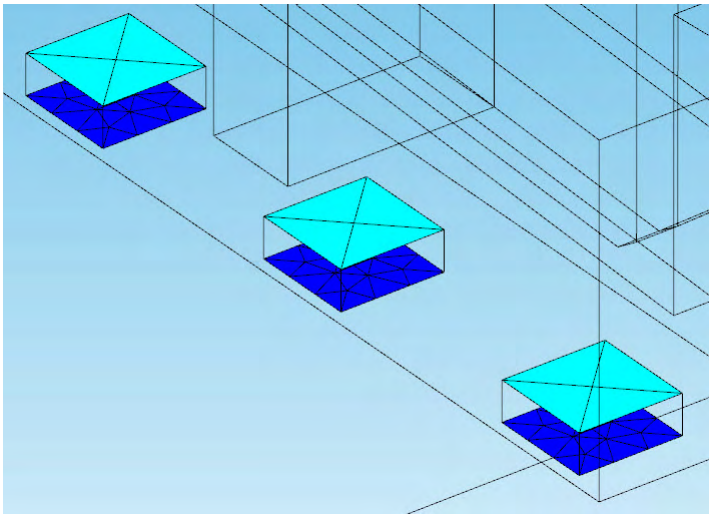
Table 4.1: Number of DOF and percentage modal frequency error for the full (unreduced) and reduced models.

DOF	Full model	$\rho = 5$	$\rho = 5,\ \nu = 200$
Internal	814.080	170	170
Interface	16.035	16.035	1.721
Total	830.115	16.205	1.891
Highest Percentage Error %	0.00	0.18	0.23

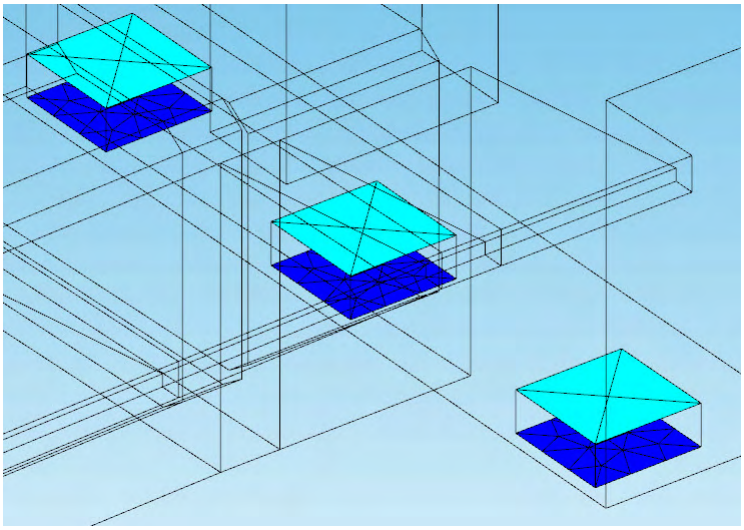
The large number of the interface DOFs can be reduced by retaining only a fraction of the constrained interface modes (Papadimitriou and Papadioti, 2013). For each interface, only the modes that have

frequency less than $\omega_{\max} = \nu\omega_c$ are retained, where ν is user and problem dependent. Results for $\nu = 200$ are given in Table 4.1. It can be seen that the largest fractional error for the lowest 20 modes of the structure falls below 0.25%. In particular, for $\nu = 200$ and $\rho = 5$ the reduced system has 1.891 DOFs from which 170 generalized coordinates are fixed-interface modes for all components and the rest 1.721 generalized coordinates are constrained interface modes (Papadimitriou and Papadioti, 2013). The error in this case is 0.23%.

Details about the reduction achieved on the internal and interface DOFs of each component separately are shown in Table 4.2. The values of the parameters ρ and ν are not specified by the CMS methodology used, but they rather have to be specified by the user depending on the specific problem. Different values of the parameters for the various components and interfaces will lead to different sizes of reduced models and corresponding accuracies in the predicted frequencies. A trade-off was made between reducing the model as much as possible (fewer kept DOFs) and keeping the accuracy of the predicted modal frequencies as close as possible to those of the unreduced model. It can be seen that a value of $\rho = 1$ was used for the 5 soil components instead of the value of $\rho = 5$ that was used for all the other components. This enabled retaining less modes from the soil components which led to fewer kept DOFs, without compromising the accuracy of the predicted modal frequencies. On the other hand, a larger value of $\nu = 5000$ was necessary to be used for the 4 interfaces of the abutments (shown in Figure 4.8) compared to the value of $\nu = 200$ used for the rest of the interfaces.



(a) Interface 1 (blue) and 6 (cyan).



(b) Interface 5 (blue) and 10 (cyan).

Figure 4.8: Interfaces of components 1, 5, 6 and 10 along with their mesh.

The different sizes of the elements in the top and bottom sides of the abutments are due to the different FE mesh sizes used to discretize the soil and deck. The deck was meshed with a coarser mesh than the soil in order to reduce the DOFs of the model. This was the result of an analysis done for several mesh cases, and was chosen to provide the optimal trade-off between the number of DOFs and accuracy in the predicted frequencies of the full FE model. Therefore, the top side of the abutments which touches the deck has fewer elements than the bottom side which touches the soil. This results in the number of DOFs seen in Table 4.2, which is 621 DOFs for interfaces 1 and 5 (bottom of abutments) and 117 DOFs for interfaces 6 and 10 (top of abutments). For the selected value of $\nu = 5000$ for these interfaces, this

results in keeping 400 of the 621 DOFs for interfaces 1 and 5 and all 117 DOFs for interfaces 6 and 10. This value of $\nu = 5000$ was found to be necessary in order to keep the required accuracy in the predicted frequencies of the reduced model.

Table 4.2: Detailed information about model reduction using CMS. (*809220 internal DOFs + 4860 fixed DOFs = 814080 DOFs)

Component	Full model		ρ in ()		ν in ()		Reduction percentage (%)
	Internal	Interface	Internal	Interface	Internal	Interface	
1	26058	621	1 (1)	621	1	400 (5000)	1.5
2	12306	1125	1 (1)	1125	1	100 (200)	0.75
3	29271	3567	20 (1)	3567	20	100 (200)	0.36
4	19443	2469	10 (1)	2469	10	100 (200)	0.5
5	27012	621	1 (1)	621	1	400 (5000)	1.45
6	729	117	1 (5)	117	1	117 (5000)	13.95
7	30672	702	7 (5)	702	7	22 (200)	0.09
8	94092	456	15 (5)	456	15	60 (200)	0.08
9	11367	690	9 (5)	690	9	55 (200)	0.53
10	729	117	1 (5)	117	1	117 (5000)	13.95
11	106485	1311	21 (5)	1311	21	50 (200)	0.06
12	164154	1221	22 (5)	1221	22	50 (200)	0.04
13	49266	945	12 (5)	945	12	50 (200)	0.12
14	43467	1056	11 (5)	1056	11	50 (200)	0.13
15	134544	1017	24 (5)	1017	24	50 (200)	0.05
16	59625	0	14 (5)	0	14	0 (200)	0.02
Sum	809220*	16035	170	16035	170	1721	0.23
Total	830115		16205		1891		

Table 4.1 shows the largest error in the first 20 modal frequencies between the unreduced and reduced models. Table 4.3 shows all 20 modal frequencies predicted by the unreduced and reduced models, together with their fractional difference with respect to the unreduced model times 1000.

Table 4.3: Comparison of modal frequencies between the unreduced and reduced models together with fractional difference times 1000.

Frequency	COMSOL 830.115 DOFs	Interior reduction 16.205 DOFs	Interface reduction 1.891 DOFs
1	0.29229	0.2923 (-0.0342)	0.293 (-2.4291)
2	0.57285	0.5729 (-0.0873)	0.5738 (-1.6584)
3	0.61889	0.6189 (-0.0162)	0.6192 (-0.5009)
4	0.84749	0.8475 (-0.0118)	0.8489 (-1.6637)
5	1.04847	1.0485 (-0.0286)	1.0495 (-0.9824)
6	1.06912	1.0692 (-0.0748)	1.0702 (-1.0102)
7	1.38448	1.3848 (-0.2311)	1.388 (-2.5425)
8	1.57685	1.5771 (-0.1585)	1.5784 (-0.983)
9	1.68234	1.6828 (-0.2734)	1.6895 (-4.256)
10	1.96412	1.9647 (-0.2953)	1.9664 (-1.1608)
11	2.15487	2.1553 (-0.1995)	2.1562 (-0.6172)
12	2.31451	2.315 (-0.2117)	2.3165 (-0.8598)
13	2.49529	2.4994 (-1.6471)	2.4996 (-1.7273)
14	2.74198	2.7429 (-0.3355)	2.7445 (-0.919)
15	2.81053	2.8141 (-1.2702)	2.8149 (-1.5549)
16	2.87191	2.8753 (-1.1804)	2.8761 (-1.459)
17	2.94499	2.9463 (-0.4448)	2.9499 (-1.6672)
18	3.31186	3.318 (-1.8539)	3.3195 (-2.3069)
19	3.37427	3.38 (-1.6981)	3.3809 (-1.9649)
20	3.5128	3.5205 (-2.192)	3.5205 (-2.192)

Thus, using CMS a drastic reduction in the number of DOFs is obtained which can exceed two orders of magnitude, without sacrificing the accuracy with which the lowest 20 modal frequencies are computed. The time to solution for one run of the reduced model is of the order of a few seconds which should be compared to approximately 2 minutes required for solving the unreduced FE model.

4.4 Experimental modal identification

4.4.1 Instrumentation

The measurement system consisted of 5 triaxial and 3 uniaxial accelerometers paired with a 24-bit data recording system, a GPS module for synchronization between sensors, and a battery pack. The system is wireless (everything is powered from the battery) and can be easily moved from one location in the structure to another. The recorder can easily connect with a laptop through wired (Ethernet) or wireless (Wi-Fi) connection to be set up in the desired way (sampling rate, recording duration, repeater recordings etc) or view the measurements while they are being recorded for quality checking. The instrumentation is shown in Figures 4.9, 4.10 and 4.11. The recorded responses are mainly due to road traffic, which ranged from light vehicles to heavy trucks, and environmental excitation such as wind loading, which classifies this case as ambient (operational) modal identification. An important aspect of this measurement system is the fact that it is wireless, since this allowed for multiple sets of repeated measurements that had to be performed for accurate mode shape identification, given the limited number of sensors and the large length of the deck. Specifically, the entire length of the deck was covered in 13 sensor configurations, shown in Figure 4.12 with each configuration recording for 20 minutes at a sampling rate of 100 Hz. One triaxial and three uniaxial sensors (one vertical and two horizontal) remained in the same position throughout the measurements, in order to provide common measurement points amongst different configurations such as to enable the assembling of the total mode shape from partial mode shape components measured from the different configurations (Au, 2011; Yan and Katafygiotis, 2015). The locations of the reference sensors were obtained by minimizing the information entropy using an optimal sensor location theory (Papadimitriou and Lombaert, 2012; Yuen and Kuok, 2015) such as to provide the highest information content for identifying the modal parameters of the structure. Indicative sensor configurations are shown in Figures 4.13 and 4.14 where green points correspond to reference sensors and blue points correspond to the moving sensors of the specific sensor configuration. The three numbers above each point correspond to the three measured DOFs in the three directions measured by the triaxial sensors.

Figure 4.9: Junction box and unit including recorder, battery pack and GPS module

Figure 4.10: Uniaxial accelerometer, junction box, and unit including recorder, battery pack and GPS module

Figure 4.11: Triaxial accelerometer

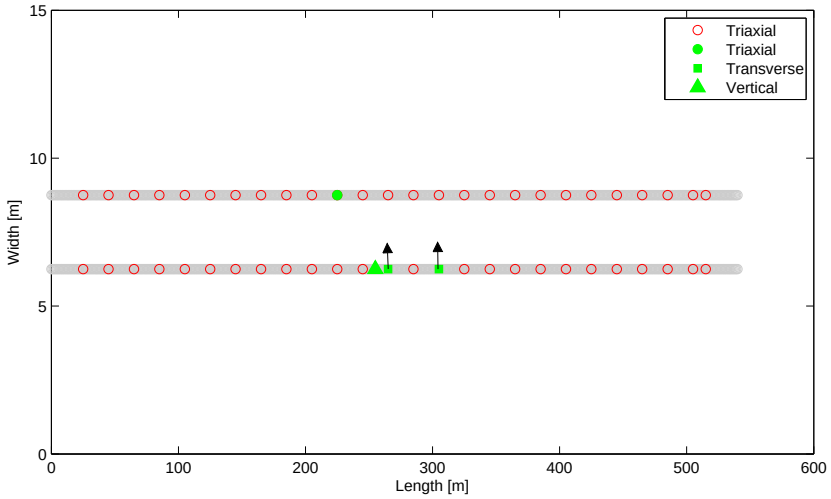


Figure 4.12: Measured locations in the bridge (reference sensors in green)

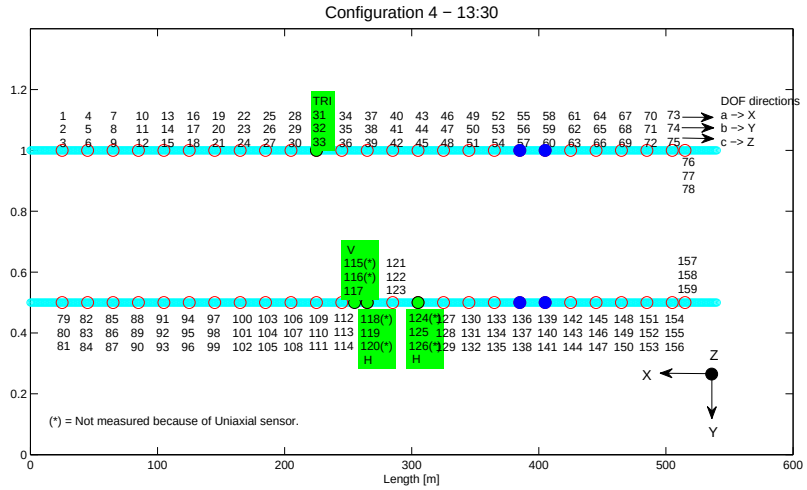


Figure 4.13: Sensor configuration 4

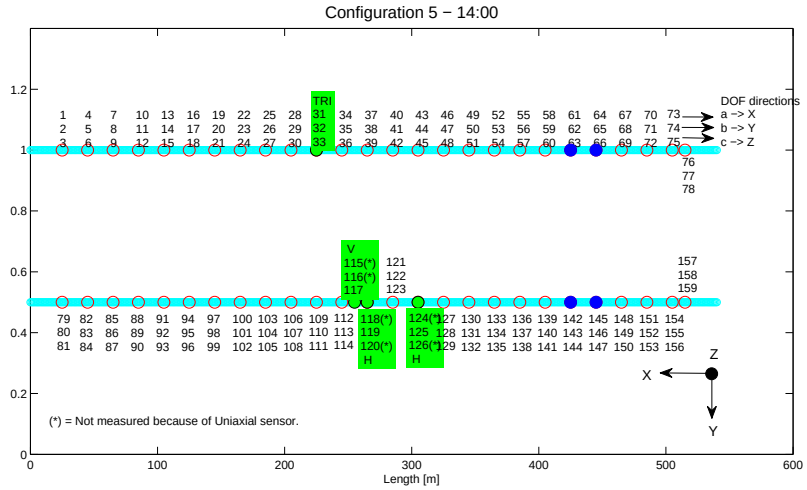


Figure 4.14: Sensor configuration 5

4.4.2 Modal identification

For the estimation of modal properties from the ambient acceleration data, and the assembly of the identified local mode shapes to produce global mode shapes, a software developed in the context of this thesis was used. The software for Ambient (Output-only) Modal Identification uses a recently proposed Bayesian methodology (Au, 2012a,b) in order to estimate the modal frequencies, mode shapes, damping ratios, and their uncertainties for each sensor configuration. The method is based on the Fast

Fourier Transform (FFT) of the acceleration signals in specific frequency bands of interest. The software also includes other modules for data insertion, pre-processing of the signals (for quality-checking of the measurements and for the specification of the frequency bands of interest), and post-processing of the results (mainly for performing mode shape assembly). The mode shape assembly methodology is similar to the one proposed by Au (2011) and is developed in detail in Section 4.4.3. The software also supports insertion of the geometry of the measured structure, for visual representation of the mode shapes in the post-processing module. More details about the software can be found in Appendix A. Illustrative ambient acceleration signals can be found in Appendix B.

Using the above mentioned software, the modal frequencies and modal damping ratios of the bridge were extracted, and the mode shape components of each configuration were combined to produce the full mode shapes at all 159 sensor locations covered by the 13 configurations. Specifically, the first 20 modal frequencies and modal damping ratios of the bridge were successfully identified, along with 11 mode shapes. The mode shapes of all the modes up to the 12th were successfully identified, except the 10th which was very poorly identified and excluded from the set. Due to the fact that the modal properties were identified from each of the 13 sensor configurations separately, their values vary slightly from one configuration to the other which forces us to consider the mean and standard deviation across all configurations. Table 4.4 presents the mean and standard deviation of the experimentally identified modal frequencies and modal damping ratios for all 20 identified modes of the Metsovo bridge. It also compares the identified frequencies and mode shapes with those predicted by the nominal FE model. The identified values for the modal damping ratios can not be compared since the FE model does not take damping effects into account. The experimental and nominal model predicted mode shapes are compared using their MAC value which is a scalar measure of correlation between two mode shapes ranging from 0 to 1, with a value of 1 indicating perfect correlation. The identified mode shapes are shown in Figures 4.15 - 4.18 and compared with the corresponding mode shapes predicted by the nominal FE model of the bridge. From both the MAC values of Table 4.4 and mode shapes of Figures 4.15 - 4.18 it can be clearly seen that the mode shapes predicted by the nominal FE model match almost perfectly the corresponding experimentally identified mode shapes with MAC values higher than 0.99 for the 11 identified mode shapes. This indicates that the nominal FE model is already very accurate in predicting the mode shapes of the bridge and no significant improvement can be achieved by the model updating procedure, since the MAC values are already too high for all practical purposes. However, there appears to be a significant mismatch between the experimental and nominal FE model modal frequencies which indicates that some or all of the parameters of the nominal model should be updated in order to achieve a closer fit with the experimental frequencies in the updated model.

Table 4.4: First 20 experimentally identified modal frequencies and modal damping ratios of the Metsovo bridge (mean and standard deviation across all 13 configurations) compared with the modal frequencies and mode shapes predicted by the nominal FE model of the bridge.

Mode	Experimental Frequency (Hz)		Nominal model Frequency (Hz)	Mode shape MAC value	Experimental Damping ratio	
	mean	std			mean	std
1	0.3063	0.0018	0.293	0.9987	0.0111	0.0025
2	0.6034	0.0070	0.5738	0.9881	0.0178	0.0024
3	0.6227	0.0017	0.6192	0.9984	0.0085	0.0014
4	0.9646	0.0047	0.8489	0.9994	0.0136	0.0195
5	1.0468	0.0079	1.0495	0.9984	0.0194	0.0105
6	1.1389	0.0065	1.0702	0.9984	0.0106	0.0043
7	1.4280	0.0048	1.388	0.9935	0.0143	0.0064
8	1.6967	0.0098	1.5784	0.9973	0.0161	0.0158
9	2.0053	0.0078	1.6895	0.9891	0.0113	0.0041
10	2.3034	0.0135	1.9664		0.0085	0.0016
11	2.3666	0.0100	2.1562	0.9954	0.0083	0.0039
12	2.5901	0.0106	2.3165	0.995	0.0080	0.0034
13	2.7226	0.0067	2.4996		0.0109	0.0043
14	3.0861	0.0126	2.7445		0.0098	0.0043
15	3.1266	0.0154	2.8149		0.0071	0.0015
16	3.4801	0.0159	2.8761		0.0184	0.0021
17	3.8608	0.0056	2.9499		0.0095	0.0023
18	4.0585	0.0116	3.3195		0.0091	0.0085
19	4.2101	0.0105	3.3809		0.0121	0.0032
20	4.4102	0.0121	3.5205		0.0084	0.0024

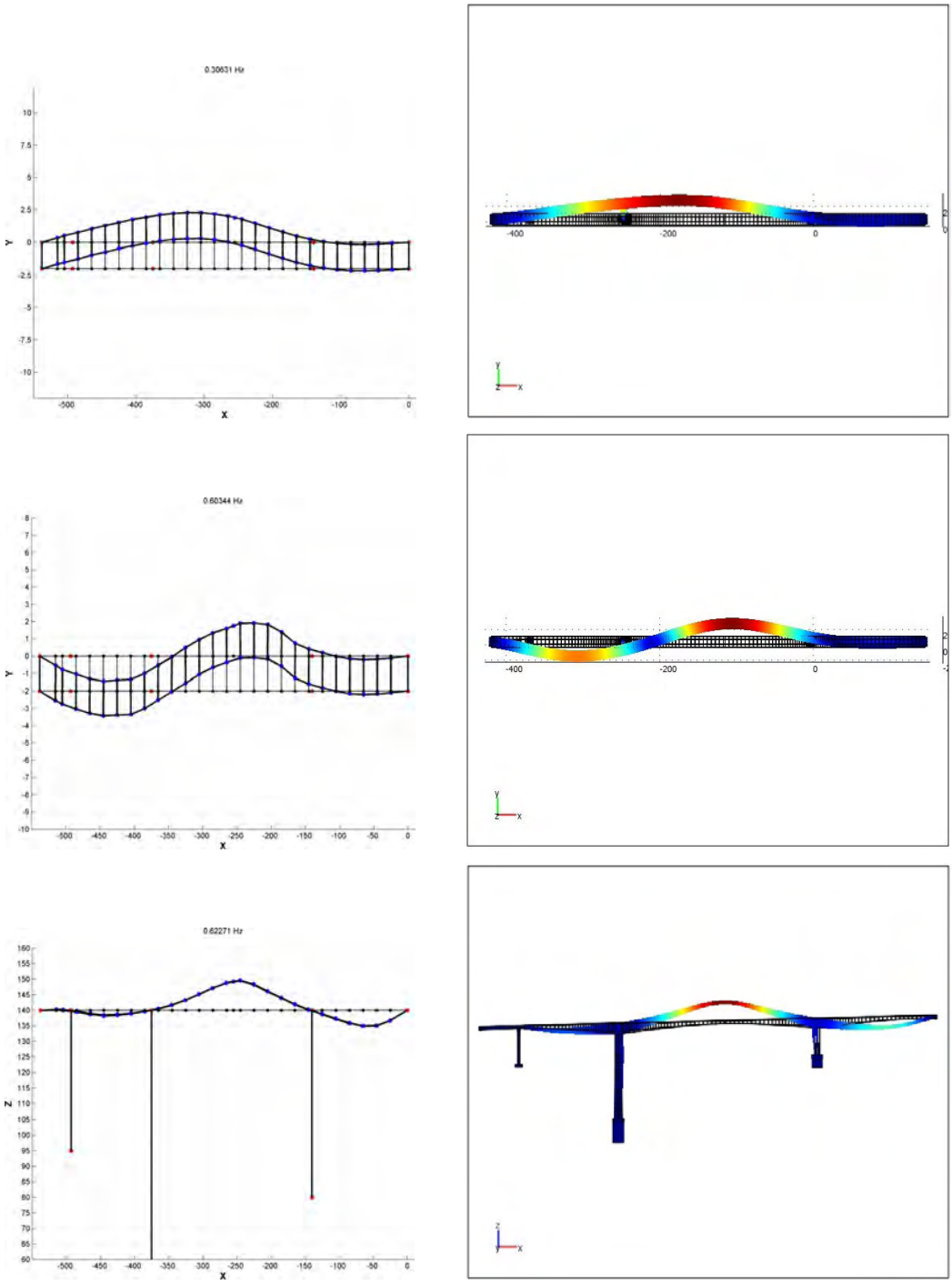


Figure 4.15: Comparison between the experimentally identified (left column) and nominal FE model predicted (right column) mode shapes of the Metsovo bridge. Modes 1-3.

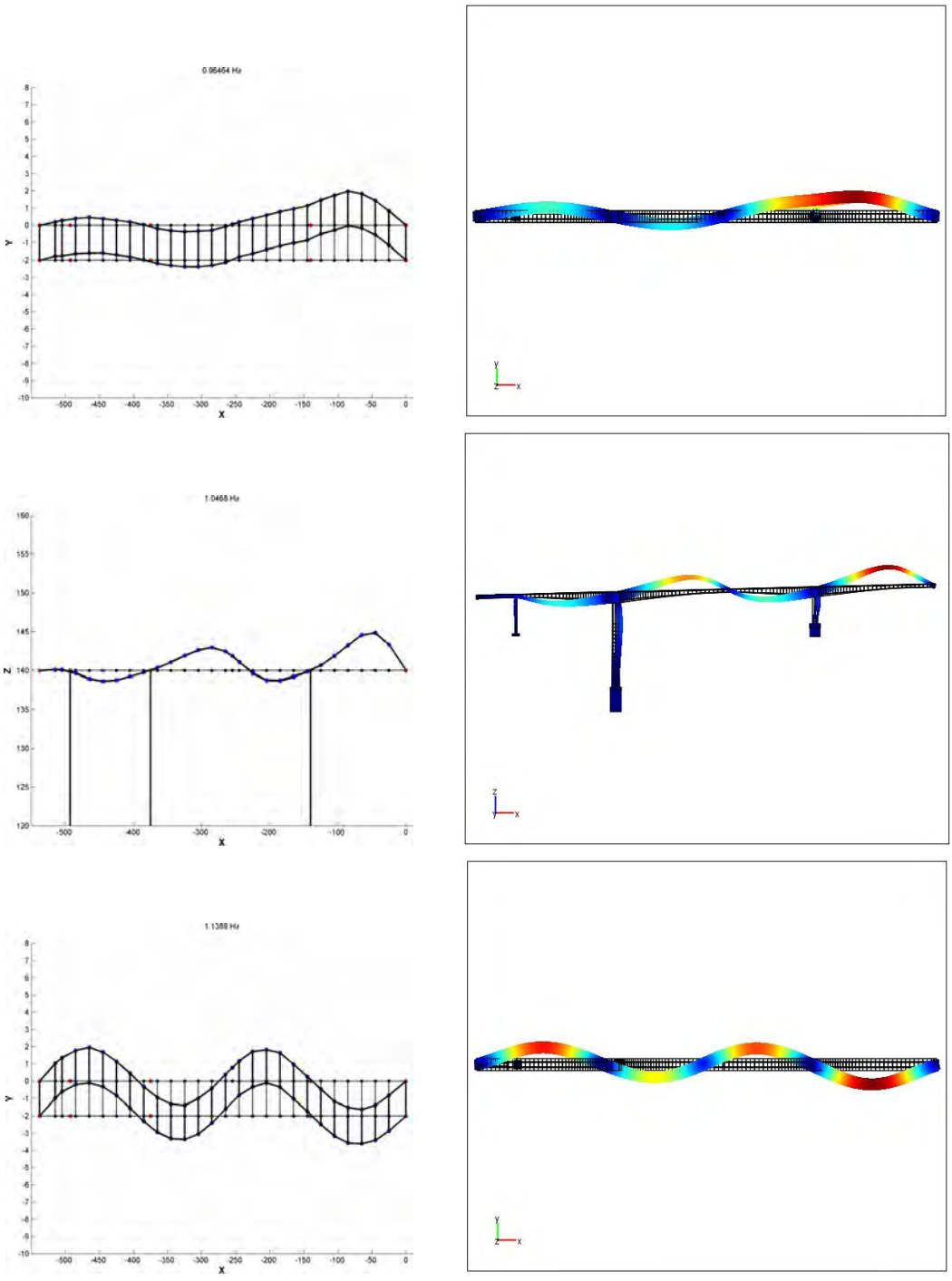


Figure 4.16: Comparison between the experimentally identified (left column) and nominal FE model predicted (right column) mode shapes of the Metsovo bridge. Modes 4-6.

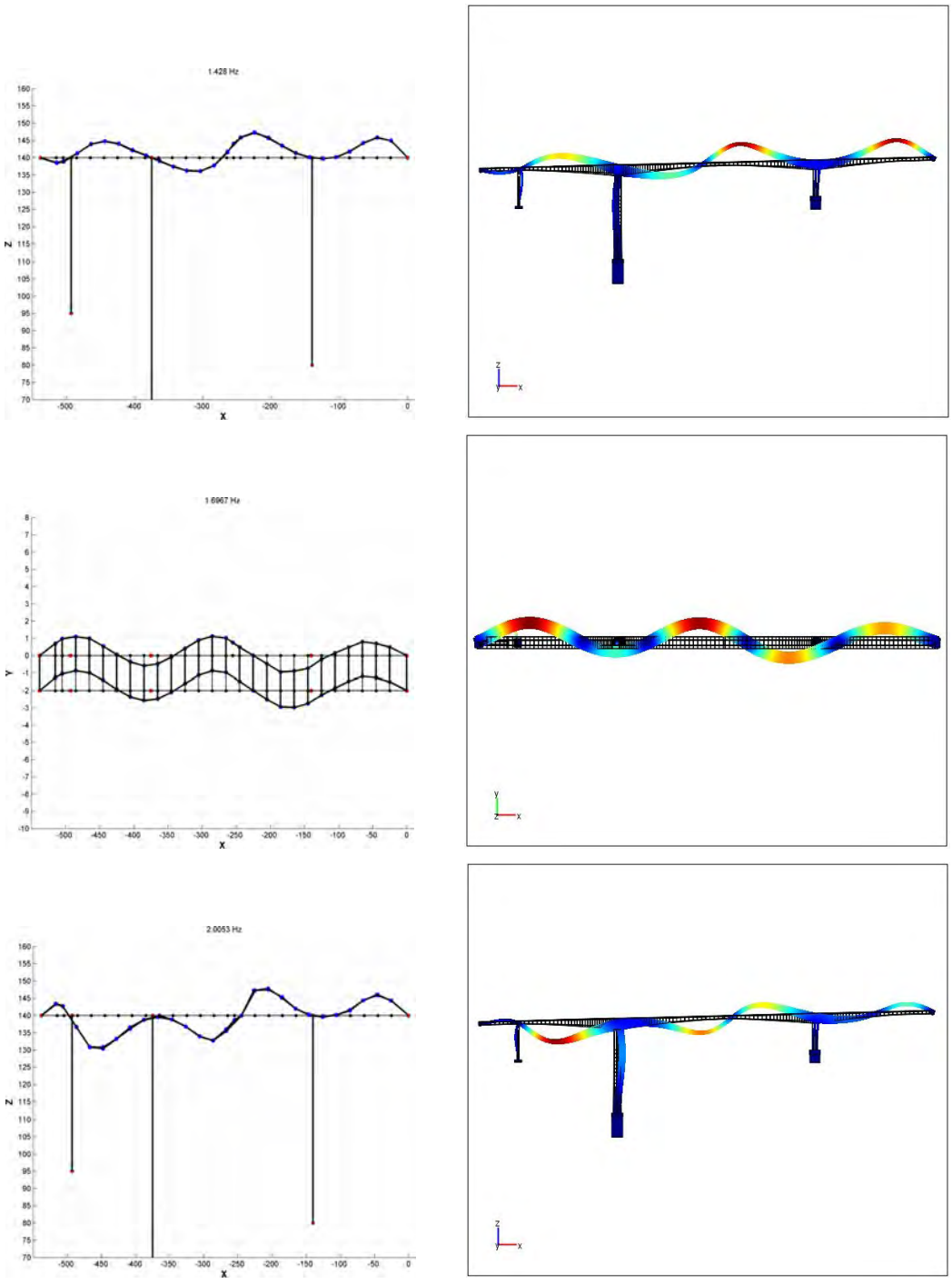


Figure 4.17: Comparison between the experimentally identified (left column) and nominal FE model predicted (right column) mode shapes of the Metsovo bridge. Modes 7-9.

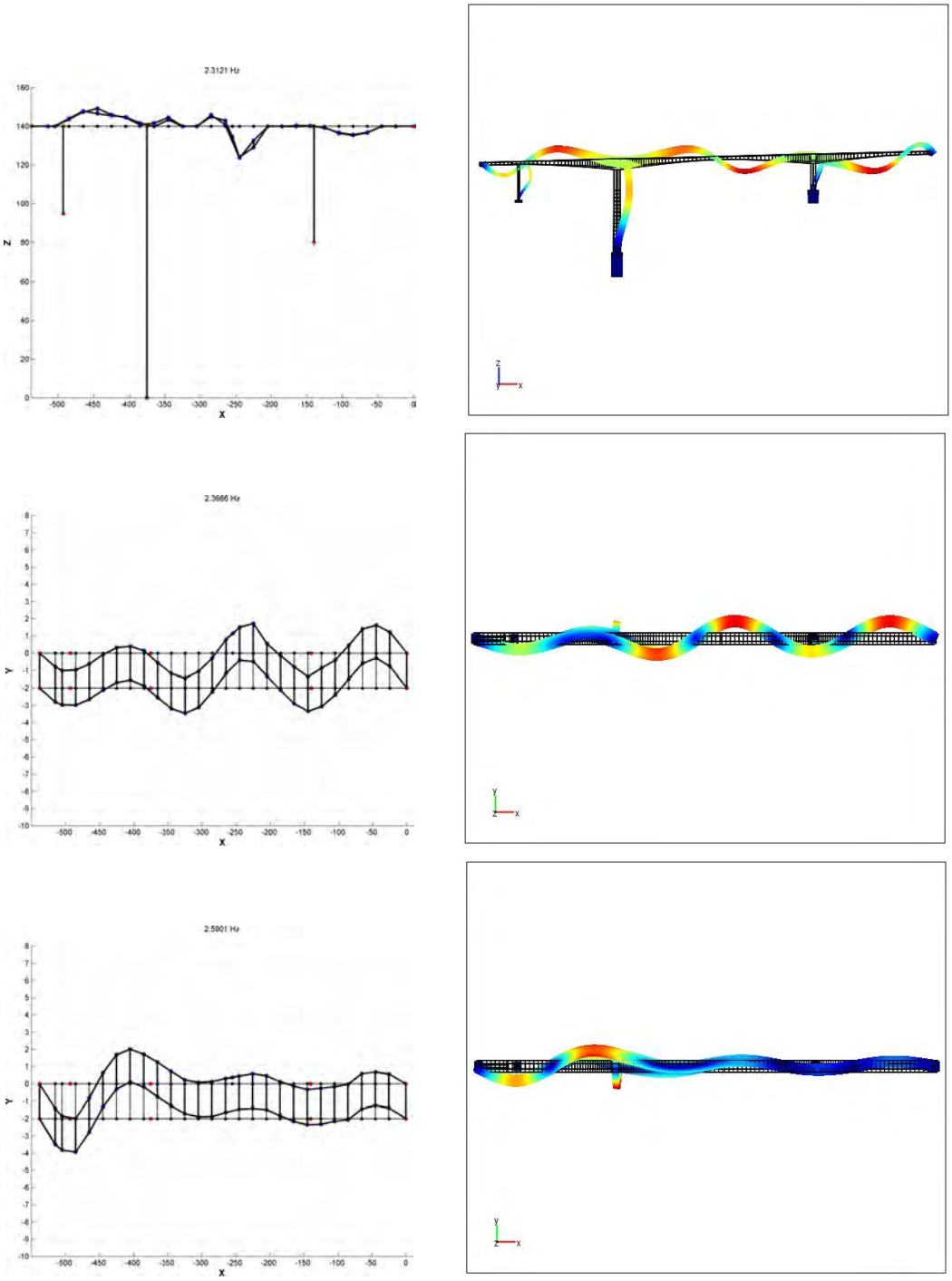


Figure 4.18: Comparison between the experimentally identified (left column) and nominal FE model predicted (right column) mode shapes of the Metsovo bridge. Modes 10-12.

4.4.3 Mode shape assembly

Problem description

When modal identification is performed we can extract certain modal properties for the r^{th} mode:

$$\omega_r,\, \zeta_r,\, \phi_{kr} :\; r = 1 \dots m,\, k = 1 \dots n \tag{4.1}$$

where m is the total number of identified modes and n is the total number of DOFs where we placed our sensors. Notice that we extract the mode shape matrix row elements ϕ_{kr} in the *specific* DOFs where the measurements were performed, namely the k^{th} row of the mode shape matrix for $k = 1 \dots n$. In order to obtain an accurate view of the mode shapes we need to extract the mode shape components in several DOFs, depending on the complexity of each mode. Higher modes tend to have more complex geometric shapes, and therefore require us to extract more mode shape components (increase the spatial resolution of mode shapes) in order to avoid the problem of spatial-aliasing.

However, in practical situations, there is only a limited number of available sensors and many points in the structure that need to be measured. This is the case in ambient vibration testing of large structures such as bridges or buildings. Obviously, it is impossible to measure all the desired DOFs at the same time, so we measure only a few of them at a time. The problem is the following:

Mode shapes identified from individual setups can have different scaling. Depending on the number of reference DOFs (common DOFs among different setups) and the quality of identified mode shapes, implementation issues can arise when determining the optimal mode shape that compromises among different setups. In this section we are going to present a method for determining the optimal mode shape that fits the mode shapes identified from multiple setups in a least square sense. The method presented in this study, is a simplification of the method proposed by Au (2011) and uses the same symbols and notation. The measure of fit function is defined as the squared difference between the theoretical and identified mode shapes suitably oriented and scaled to the same norm. The most important element of the method is the presence of *reference DOFs*. Since the mode shapes identified from individual setups are arbitrary in sense and scaling, common reference DOFs must be present across different setups in order to allow their mode shapes to be assembled. The reference DOFs must have significant frequency response in the modes of interest. There must also be at least one common reference DOF across any two setups.

Proposed methodology

Let n_s be the number of different setups, n_i be the number of measured DOFs in the i^{th} setup, and n be the total number of DOFs measured (in all setups). Let $\boldsymbol{\phi} \in R^n$ be the mode shape (for the r mode) covering the measured DOFs in all setups (unknown). Also, let $\hat{\boldsymbol{\phi}}_i \in R^{n_i}$ be the measured mode shape vector from the i^{th} configuration. Vector $\hat{\boldsymbol{\phi}}_i$ has n_i components because the i^{th} configuration had n_i DOFs measured. Note that $\sum_{i=1}^{n_s} n_i > n$ because some DOFs are measured in more than one configuration.

We need a way to compare the measured mode shape of the i^{th} configuration with the corresponding DOFs of the full unknown mode shape $\boldsymbol{\phi}$. This is achieved by using a selection matrix $\boldsymbol{L}_i \in R^{n_i \times n}$ defined for every configuration as : $\boldsymbol{L}_i(j,k) = 1$, if the j^{th} data channel in the i^{th} setup gives the k^{th} DOF of the full mode shape; and zero otherwise. Therefore, the mode shape $\boldsymbol{\phi}_i$ in the i^{th} setup can be mathematically related to the assembled mode shape $\boldsymbol{\phi}$ by:

$$\boldsymbol{\phi}_i = \boldsymbol{L}_i \boldsymbol{\phi} \quad , \quad \boldsymbol{\phi}_i \in R^{n_i} \tag{4.2}$$

Our objective is to determine the 'best' assembled mode shape $\boldsymbol{\phi}$ so that $\{\boldsymbol{\phi}_i : i = 1, \ldots, n_s\}$ are closest to their measured counterparts $\left\{\hat{\boldsymbol{\phi}}_i : i = 1, \ldots, n_s\right\}$ in a least square sense. The optimal mode shape is formulated as the one that minimizes a measure-of-fit objective function. When formulating the objective function, the issue of scaling should be addressed. Each $\hat{\boldsymbol{\phi}}_i$ might be differently scaled, but the shapes of two vectors can only be compared fairly if they have the same Euclidean norm. This suggests that the discrepancy should be measured based on the difference between $\boldsymbol{\phi}_i$ and $c_i\hat{\boldsymbol{\phi}}_i$ rather than directly between $\boldsymbol{\phi}_i$ and $\hat{\boldsymbol{\phi}}_i$; in the former case both vectors have the same norm.

The objective function that should be minimized is :

$$J(\boldsymbol{\phi}) = \sum_{i=1}^{n_s} \|\boldsymbol{\phi}_i - c_i\hat{\boldsymbol{\phi}}_i\|^2 \tag{4.3}$$

subject to the constraint:

$$\|\boldsymbol{\phi}\|^2 = 1 \Rightarrow \boldsymbol{\phi}^T\boldsymbol{\phi} = 1 \Rightarrow \boldsymbol{\phi}^T\boldsymbol{\phi} - 1 = 0 \tag{4.4}$$

Constrained optimization is converted to unconstrained using Lagrange multipliers as follows :

$$J(\boldsymbol{\phi}, \lambda) = J(\boldsymbol{\phi}) + \lambda(1 - \boldsymbol{\phi}^T\boldsymbol{\phi}) \tag{4.5}$$

In order to proceed we need to find the values of $c_i : i = 1, \ldots, n_s$

Optimal scaling factors This is the part which significantly deviates from the methodology proposed by Au (2011), leading to a different form of the solution. In order to find the optimal c_i we now consider a different minimization problem with respect to c_i this time. Minimize the objective function :

$$G(c_i) = \|\boldsymbol{\phi}_i - c_i\hat{\boldsymbol{\phi}}_i\|^2 = (\boldsymbol{\phi}_i - c_i\hat{\boldsymbol{\phi}}_i)^T(\boldsymbol{\phi}_i - c_i\hat{\boldsymbol{\phi}}_i) \tag{4.6}$$

This minimization can be performed analytically as follows:

$$\begin{aligned}
\frac{\partial G}{\partial c_i} = 0 \Rightarrow -\hat{\boldsymbol{\phi}}_i^{\,T}(\boldsymbol{\phi}_i - c_i\hat{\boldsymbol{\phi}}_i) + (\boldsymbol{\phi}_i - c_i\hat{\boldsymbol{\phi}}_i)^T(-\hat{\boldsymbol{\phi}}_i) = 0 \\
\Rightarrow -\hat{\boldsymbol{\phi}}_i^{\,T}\boldsymbol{\phi}_i + c_i\hat{\boldsymbol{\phi}}_i^{\,T}\hat{\boldsymbol{\phi}}_i - \boldsymbol{\phi}_i^T\hat{\boldsymbol{\phi}}_i + c_i\hat{\boldsymbol{\phi}}_i^{\,T}\hat{\boldsymbol{\phi}}_i = 0 \\
-2\hat{\boldsymbol{\phi}}_i^{\,T}\boldsymbol{\phi}_i + 2c_i\hat{\boldsymbol{\phi}}_i^{\,T}\hat{\boldsymbol{\phi}}_i = 0 \\
\Rightarrow c_i = \frac{\hat{\boldsymbol{\phi}}_i^{\,T}\boldsymbol{\phi}_i}{\hat{\boldsymbol{\phi}}_i^{\,T}\hat{\boldsymbol{\phi}}_i} = \frac{\hat{\boldsymbol{\phi}}_i^{\,T}\boldsymbol{\phi}_i}{\|\hat{\boldsymbol{\phi}}_i\|^2}
\end{aligned}$$

For our analysis, we normalize $\hat{\boldsymbol{\phi}}_i$ such that $\|\hat{\boldsymbol{\phi}}_i\|^2 = 1$ so finally the expression for the c_i is :

$$c_i = \hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{\phi}_i \tag{4.7}$$

Minimization problem Thus, returning to our original minimization problem we have to minimize the objective function :

$$J(\boldsymbol{\phi}, \lambda) = J(\boldsymbol{\phi}) + \lambda(1 - \boldsymbol{\phi}^T\boldsymbol{\phi}) = \sum_{i=1}^{n_s} \|\boldsymbol{\phi}_i - (\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{\phi}_i)\hat{\boldsymbol{\phi}}_i\|^2 + \lambda(1 - \boldsymbol{\phi}^T\boldsymbol{\phi}) \tag{4.8}$$

By substituting $\boldsymbol{\phi}_i = \boldsymbol{L}_i\boldsymbol{\phi}$ in the above equation we have :

$$\begin{aligned} J(\boldsymbol{\phi}, \lambda) &= \sum_{i=1}^{n_s} \|\boldsymbol{L}_i\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{L}_i\boldsymbol{\phi}\|^2 + \lambda(1 - \boldsymbol{\phi}^T\boldsymbol{\phi}) \\ &= \sum_{i=1}^{n_s} \|(\boldsymbol{L}_i - \hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{L}_i)\boldsymbol{\phi}\|^2 + \lambda(1 - \boldsymbol{\phi}^T\boldsymbol{\phi}) \\ = \sum_{i=1}^{n_s} \boldsymbol{\phi}^T(\boldsymbol{L}_i - \hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{L}_i)^T(\boldsymbol{L}_i - \hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{L}_i)\boldsymbol{\phi} &+ \lambda(1 - \boldsymbol{\phi}^T\boldsymbol{\phi}) \end{aligned}$$

And in more compact form :

$$J(\boldsymbol{\phi}, \lambda) = \sum_{i=1}^{n_s} \boldsymbol{\phi}^T\boldsymbol{A}_i\boldsymbol{\phi} + \lambda(1 - \boldsymbol{\phi}^T\boldsymbol{\phi}) \tag{4.9}$$

where we have defined $\boldsymbol{A}_i$ as :

$$\boldsymbol{A}_i = (\boldsymbol{L}_i - \hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{L}_i)^T(\boldsymbol{L}_i - \hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{L}_i) \tag{4.10}$$

A more careful examination of $\boldsymbol{A}_i$ yields :

$$\begin{aligned} \boldsymbol{A}_i &= (\boldsymbol{L}_i - \hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{L}_i)^T(\boldsymbol{L}_i - \hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{L}_i) \\ &= (\boldsymbol{L}_i^T - \boldsymbol{L}_i^T\hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T)(\boldsymbol{L}_i - \hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T \boldsymbol{L}_i) \\ = \boldsymbol{L}_i^T\boldsymbol{L}_i - \boldsymbol{L}_i^T\hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T\boldsymbol{L}_i - \boldsymbol{L}_i^T\hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T\boldsymbol{L}_i &+ \boldsymbol{L}_i^T\hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T\hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T\boldsymbol{L}_i \end{aligned}$$

Taking into account that we normalize our measured mode shapes, $\|\hat{\boldsymbol{\phi}}_i\|^2 = 1 \Rightarrow \hat{\boldsymbol{\phi}}_i{}^T\hat{\boldsymbol{\phi}}_i = 1$ we can further simplify the expression for $\boldsymbol{A}_i$:

$$\boldsymbol{A}_i = \boldsymbol{L}_i^T\boldsymbol{L}_i - \boldsymbol{L}_i^T\hat{\boldsymbol{\phi}}_i\hat{\boldsymbol{\phi}}_i{}^T\boldsymbol{L}_i \tag{4.11}$$

From equation (4.11) we can deduce that $\boldsymbol{A}_i \in R^{n\times n}$, $\boldsymbol{A}_i^T = \boldsymbol{A}_i$ and $\boldsymbol{A}_i$ is independent of $\boldsymbol{\phi}$. Thus, returning to our objective function we note from equation (4.9) that $J(\boldsymbol{\phi}, \lambda)$ is quadratic in $\boldsymbol{\phi}$ components, thus we expect to solve it analytically, using derivatives :

$$J(\boldsymbol{\phi}, \lambda) = \sum_{i=1}^{n_s} \boldsymbol{\phi}^T\boldsymbol{A}_i\boldsymbol{\phi} + \lambda(1 - \boldsymbol{\phi}^T\boldsymbol{\phi})$$

$$\Rightarrow \frac{\partial J}{\partial \phi_j} = \sum_{i=1}^{n_s} (\frac{\partial \boldsymbol{\phi}}{\partial \phi_j})^T \boldsymbol{A}_i \boldsymbol{\phi} + \boldsymbol{\phi}^T \boldsymbol{A}_i \frac{\partial \boldsymbol{\phi}}{\partial \phi_j} - \lambda(\frac{\partial \boldsymbol{\phi}^T}{\partial \phi_j}\boldsymbol{\phi} + \boldsymbol{\phi}^T \frac{\partial \boldsymbol{\phi}}{\partial \phi_j}) \tag{4.12}$$

Further simplification is achieved when noticing that the assembled mode shape vector $\boldsymbol{\phi}$ has constant elements, therefore the derivative of the vector with respect to a specific element is a zero-vector with only one non-zero element equal to one, in the position of the aforementioned derivation element.

$$\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_j, \ldots, \phi_n)^T \Rightarrow \frac{\partial \boldsymbol{\phi}}{\partial \phi_j} = (0, 0, \ldots, 1, \ldots, 0)^T$$

Using Kronecker's delta we can rewrite the derivative in a more compact form :

$$\frac{\partial \boldsymbol{\phi}}{\partial \phi_j} = \boldsymbol{\delta}_j \in R^n \tag{4.13}$$

By substituting into equation (4.12) we obtain :

$$\frac{\partial J}{\partial \phi_j} = \sum_{i=1}^{n_s} \boldsymbol{\delta}_j^T \boldsymbol{A}_i \boldsymbol{\phi} + \boldsymbol{\phi}^T \boldsymbol{A}_i \boldsymbol{\delta}_j - \lambda(\boldsymbol{\delta}_j^T \boldsymbol{\phi} + \boldsymbol{\phi}^T \boldsymbol{\delta}_j)$$

We notice that the terms of the resulting equation are scalar, and therefore they are equal to their transpose. Using the matrix identity : $(\mathbf{ABC})^T = \mathbf{C}^T\mathbf{B}^T\mathbf{A}^T$, and by using the fact that $\boldsymbol{A}_i^T = \boldsymbol{A}_i$ from equation (4.11), we further simplify the derivative to :

$$\frac{\partial J}{\partial \phi_j} = 2\sum_{i=1}^{n_s} (\boldsymbol{\delta}_j^T \boldsymbol{A}_i \boldsymbol{\phi}) - 2\lambda \boldsymbol{\delta}_j^T \boldsymbol{\phi} = 0 \ , \ \ j = 1, \ldots, n \tag{4.14}$$

Equations (4.14) form a system of n equations linear in $\boldsymbol{\phi}$. We can estimate $\boldsymbol{\phi}$ from its solution. Because of its linearity, this system of equations can be solved analytically. First we rewrite the equations in matrix form. To do that we note that the product $\boldsymbol{A}_i\boldsymbol{\phi}$ is a column vector of size n. Its pre-multiplication with $\boldsymbol{\delta}_j^T$ returns a scalar which is just the j^{th} element of $\boldsymbol{A}_i\boldsymbol{\phi}$. The same holds for the second term of equations (4.14). Therefore, the matrix form of the system is :

$$\sum_{i=1}^{n_s} \boldsymbol{A}_i \boldsymbol{\phi} - \lambda \boldsymbol{\phi} = \mathbf{0} \Rightarrow \boldsymbol{A}_0 \boldsymbol{\phi} - \lambda \boldsymbol{\phi} = \mathbf{0} \Rightarrow \boldsymbol{A}_0 \boldsymbol{\phi} = \lambda \boldsymbol{\phi} \ , \ \boldsymbol{A}_0 = \sum_{i=1}^{n_s} \boldsymbol{A}_i \tag{4.15}$$

Equation (4.15) constitutes an *eigenvalue* problem. This means that in order to find our unknown assembled mode shape $\boldsymbol{\phi}$ we have to find the eigenvectors of the matrix $\boldsymbol{A}_0$ given in the above equation. An obvious issue that we have to address is the multitude of eigenvectors. Specifically, since matrix $\boldsymbol{A}_0$ is of dimension $n \times n$ the solution of the eigenproblem will yield n eigenvectors, but we want one solution. We choose the eigenvector $\boldsymbol{\phi}^{(r)}$ that minimizes our objective function $J(\boldsymbol{\phi}, \lambda)$:

$$J(\boldsymbol{\phi}, \lambda) = \sum_{i=1}^{n_s} \boldsymbol{\phi}^T \boldsymbol{A}_i \boldsymbol{\phi} + \lambda(1 - \boldsymbol{\phi}^T \boldsymbol{\phi}) = \boldsymbol{\phi}^T (\sum_{i=1}^{n_s} \boldsymbol{A}_i)\boldsymbol{\phi} + \lambda(1 - \boldsymbol{\phi}^T \boldsymbol{\phi})$$

$$\Rightarrow J(\boldsymbol{\phi}, \lambda) = \boldsymbol{\phi}^T \boldsymbol{A}_0 \boldsymbol{\phi} + \lambda(1 - \boldsymbol{\phi}^T \boldsymbol{\phi}) \tag{4.16}$$

For the r^{th} eigenvector we have that : $\boldsymbol{A}_0\boldsymbol{\phi}^{(r)} = \lambda^{(r)}\boldsymbol{\phi}^{(r)}$ and we also assume that we have normalized the eigenvector such that $\|\boldsymbol{\phi}^{(r)}\|^2 = 1$

By substituting into the objective function we have :

$$J(\boldsymbol{\phi}^{(r)}, \lambda) = \boldsymbol{\phi}^{(r)T}\boldsymbol{A}_0\boldsymbol{\phi}^{(r)} = \boldsymbol{\phi}^{(r)T}\lambda^{(r)}\boldsymbol{\phi}^{(r)} = \lambda^{(r)}\boldsymbol{\phi}^{(r)T}\boldsymbol{\phi}^{(r)} = \lambda^{(r)}\|\boldsymbol{\phi}^{(r)}\|^2$$

$$\Rightarrow J(\boldsymbol{\phi}^{(r)}, \lambda) = \lambda^{(r)} \tag{4.17}$$

It is obvious that the objective function is minimized when $\lambda^{(r)}$ is minimum, i.e. the lowest eigenvalue of the problem. Therefore, the solution to the problem is the eigenvector that corresponds to the lowest eigenvalue, namely $\boldsymbol{\phi}^{(1)}$.

To sum up the important equations, first, we find $\boldsymbol{A}_i$ based on the configuration matrix $\boldsymbol{L}_i$ and the measurements $\hat{\boldsymbol{\phi}}_i$ from (4.11). Then we use equation (4.15) to find $\boldsymbol{A}_0$ from the $\boldsymbol{A}_i$ and find the eigenvalues and eigenvectors of $\boldsymbol{A}_0$. The eigenvector corresponding to the lowest eigenvalue is the solution to our problem.

4.5 Bayesian parameter estimation

4.5.1 Likelihood formulation using modal data

To apply the Bayesian formulation for parameter estimation of linear FE models, we consider that the data D consists of the square of the modal frequencies, $\hat{\lambda}_r = \hat{\omega}_r^2$, and the mode shapes $\underline{\hat{\phi}}_r \in R^{N_{0,r}}$, $r = 1, \ldots, m$, experimentally estimated using vibration measurements, where m is the number of identified modes and $N_{0,r}$ is the number of measured components for mode r.

Consider a parameterized linear FE model class $\tilde{M}$ of a structure and let $\underline{\theta} \in R^{N_\theta}$ be a vector of free structural model parameters to be estimated using the set of modal properties identified from vibration measurements. Let $\omega_r(\underline{\theta})$ and $\underline{\phi}_r(\underline{\theta}) \in R^{N_{0,r}}$ be the r-th modal frequency and mode shape at $N_{0,r}$ DOFs, respectively, predicted by the model for a given value $\underline{\theta}$ of the model parameters. The likelihood $p(D|\underline{\theta}, \tilde{M})$ in (2.2) is built up using the following considerations. The prediction error equation for the r-th modal frequency is taken as:

$$\hat{\omega}_r^2 = \omega_r^2(\underline{\theta}) + \varepsilon_{\lambda_r} \tag{4.18}$$

where ε_{λ_r} is the prediction error for the r-th frequency taken to be Gaussian with zero mean and standard deviation $\sigma_{\omega_r}\hat{\omega}_r$, with the unknown parameter σ_{ω_r} to be included in the parameter set $\underline{\theta}_e$ to be estimated from the data.

The prediction error equation for the r-th mode shape is

$$\underline{\hat{\phi}}_r = \beta_r(\underline{\theta})\underline{\phi}_r(\underline{\theta}) + \underline{\varepsilon}_{\underline{\phi}_r} \tag{4.19}$$

where $\underline{\varepsilon}_{\underline{\phi}_r}$ is the prediction error vector for the r-th mode shape taken to be Gaussian with zero mean and covariance matrix $\sigma^2_{\underline{\phi}_r}\Sigma_{\underline{\phi}_r}$, where the matrix $\Sigma_{\underline{\phi}_r}$ specifies the possible correlation structure between

the components of the prediction error vector of the r-th mode shape, the unknown scalar $\sigma^2_{\underline{\phi}_r}$ is included in the parameter set to be estimated, and

$$\beta_r(\underline{\theta}) = \hat{\underline{\phi}}_r^T \underline{\phi}_r(\underline{\theta}) \,/\, \left\| \underline{\phi}_r(\underline{\theta}) \right\|^2 \tag{4.20}$$

is a normalization constant such that the measured mode shape $\hat{\underline{\phi}}_r$ at the measured DOFs is closest to the model mode shape $\beta_r(\underline{\theta})\underline{\phi}_r(\underline{\theta})$ predicted by the particular value of $\underline{\theta}$, and $||\underline{z}||^2 = \underline{z}^{\mathrm{T}}\underline{z}$ is the usual Euclidean norm. For the simplest case of uncorrelated mode shape prediction error vectors their covariance matrix simplifies to a diagonal matrix

$$\underline{\varepsilon}_{\underline{\phi}_r} \sim N\left(\underline{0},\, diag(\sigma^2_{\underline{\phi}_r} ||\hat{\underline{\phi}}_r||^2)\right) \tag{4.21}$$

The squares of the modal frequencies $\lambda_r(\underline{\theta}) = \omega_r^2(\underline{\theta})$ and the mode shape components $\underline{\phi}_r(\underline{\theta}) = L_r\underline{\varphi}_r(\underline{\theta}) \in \mathbb{R}^{N_{0,r}}$ at the $N_{0,r}$ measured DOFs are computed from the full mode shapes $\underline{\varphi}_r(\underline{\theta}) \in \mathbb{R}^n$ that satisfy the eigenvalue problem:

$$[K(\underline{\theta}) - \lambda_r(\underline{\theta})M(\underline{\theta})]\underline{\varphi}_r(\underline{\theta}) = \underline{0} \tag{4.22}$$

where $K(\underline{\theta}) \in \mathbb{R}^{n\times n}$ and $M(\underline{\theta}) \in \mathbb{R}^{n\times n}$ are the stiffness and mass matrices respectively of the FE model of the structure, n is the number of model DOFs, and $L_r \in R^{N_{0,r}\times n}$ is an observation matrix, usually comprised of zeros and ones, that maps the n model DOFs to the $N_{0,r}$ observed DOFs for mode r. For a model with large number of DOFs, $N_{0,r} \ll n$.

The structural model class $\tilde{M}$ is augmented to include the prediction error model class that postulates zero-mean Gaussian models for the modal frequency and mode shape error terms ε_{λ_r} and $\varepsilon_{\underline{\phi}_r}$ in (4.18) and (4.19), respectively, with equal variances for all modal frequency and mode shape errors. Assuming $\sigma^2_{\omega_r} = \sigma^2_{\underline{\phi}_r} = \sigma^2$, the likelihood function can then be readily obtained in the form

$$p(D|\underline{\theta}, \tilde{M}) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^{m(N_0+1)}} \exp\left[-\frac{1}{2\sigma^2} J(\underline{\theta})\right] \tag{4.23}$$

where

$$J(\underline{\theta}) = J_1(\underline{\theta}) + J_2(\underline{\theta}) \tag{4.24}$$

is the modal residual term. In (4.24) the following modal frequency residuals

$$J_1(\underline{\theta}) = \sum_{r=1}^{m} \varepsilon^2_{\lambda_r}(\underline{\theta}) = \sum_{r=1}^{m} \frac{\left[\lambda_r(\underline{\theta}) - \hat{\lambda}_r\right]^2}{\hat{\lambda}_r^2} \tag{4.25}$$

and mode shape residuals

$$J_2(\underline{\theta}) = \sum_{r=1}^{m} \varepsilon^2_{\underline{\phi}_r}(\underline{\theta}) = \sum_{r=1}^{m} \frac{\left\| \beta_r(\underline{\theta})\underline{\phi}_r(\underline{\theta}) - \hat{\underline{\phi}}_r \right\|^2}{\left\| \hat{\underline{\phi}}_r \right\|^2} \tag{4.26}$$

measure the differences ε_{λ_r} and $\varepsilon_{\underline{\phi}_r}$ for the modal frequencies and mode shape components between the identified modal data and the model predicted modal data, respectively. It is worth noting that it can be shown that the square of the mode shape residuals in (4.26) is related to the modal assurance criterion (MAC) value of the mode r by (Papadimitriou et al., 2011)

$$\varepsilon^2_{\underline{\phi}_r}(\underline{\theta}) = 1 - MAC_r^2(\underline{\theta}) = 0 \leq 1 - \left[\frac{\left[\underline{\phi}_r\right]^T \hat{\underline{\phi}}_r}{\left\|\underline{\phi}_r\right\| \left\|\hat{\underline{\phi}}_r\right\|} \right]^2 \leq 1 \tag{4.27}$$

since $0 \leq MAC_r^2 \leq 1$. Thus $J_2(\underline{\theta})$ in (4.26) is also a measure of the distance of the square MAC value from one, or equivalently, a measure of the correlation between the model predicted mode shape and the measured mode shape.

Based on the above-mentioned note, a second formulation can be derived for the use of the mode shape data based on the MAC value. Specifically, the scalar quantity $0 \leq 1 - MAC_r^2(\underline{\theta}) \leq 1$ for mode r can be reasonably assumed to follow a probability distribution defined on the interval [0, 1], such as the Beta distribution. The 2 positive shape parameters of the beta distribution can also be included in the uncertain model parameter vector $\underline{\theta}$ to be identified from the data. This formulation reduces the number of data for mode shape r from $N_{0,r}$ mode shape components (vector) to just a scalar number. The benefit of such an approach for the mode shape data is that when using the mode shape as a vector, one must either assume the components of the vector to be uncorrelated with each other, or introduce correlation described by a correlation function. In many practical cases both choices might be very difficult to make because some correlation is bound to exist between the mode shape components, but finding the correct correlation function might be very difficult. Then reducing the mode shape data from a vector to a scalar is a convenient simplifying assumption. The beta distribution is not the only distribution that can serve for this purpose, other distributions for real numbers in [0, 1] can also be explored.

4.5.2 Computational tools

The Bayesian tools for identifying FE models as well as performing robust prediction analyses require a moderate to very large number of repeated system analyses to be performed over the space of uncertain parameters. Consequently, the computational demands depend highly on the number of system analyses and the time required for performing a system analysis. For linear FE models with large number of DOFs, this can increase substantially the computational effort to excessive levels. Computational savings are achieved by adopting parallel computing algorithms to efficiently distribute the computations in available multi-core CPUs (Angelikopoulos et al., 2012; Hadjidoukas et al., 2015). Specifically, the Π4U software (Hadjidoukas et al., 2015), based on a parallelized version of the Transitional MCMC (TMCMC) algorithm (Ching and Chen, 2007), was used to efficiently distribute the computations in available multi-core CPUs and draw samples from the posterior probability density function (PDF) of the model parameters. This sampling algorithm requires only point-wise evaluation of the likelihood function for many different values of the model and prediction error parameters.

In structural dynamics, fast and accurate component mode synthesis (CMS) techniques, consistent with the finite element model parameterization, were integrated with Bayesian techniques to reduce efficiently and drastically the model and thus the computational effort (Papadimitriou and Papadioti, 2013; Jensen et al., 2014). The reduction achieved by applying the CMS technique in the finite element model of the bridge was described in Section 4.3.2.

4.6 Model updating and model selection results

The FE model of the bridge-soil system is next updated using a subset of the experimentally identified modal frequencies following the formulation presented in Section 4.5.1 with the modal residual term of Equation (4.24) taking into account only the modal frequencies

$$J(\underline{\theta}) = J_1(\underline{\theta}) \tag{4.28}$$

This approach allows one to use the rest of the experimentally identified modal data to validate the updated model by checking its predictive capabilities with data that was not used to update the model. Specifically, the first 15 identified modal frequencies are used to estimate the model parameters and their uncertainty (model updating), while the rest 5 modal frequencies and 11 mode shapes are used in order to validate the updated model. Remind that the MAC values for the 11 identified mode shapes were already very close to 1 with the nominal model, before the updating.

As already described in Section 4.3.1 the FE model of the bridge-soil system is parameterized using three parameters associated with the modulus of elasticity of the deck (θ_1), piers (θ_2) and soil (θ_3). This parametrization is depicted graphically in Figure 4.19. Since the entire deck is constructed from the same material there is no reason to assume that its modulus of elasticity exhibits any spatial variation which would require more parameters to describe the deck stiffness. The same applies for the three piers and soil blocks. The model parameters multiply the nominal values of the corresponding moduli of elasticity which are $37GPa$ for the deck, $34GPa$ for the piers and $1GPa$ for the soil. The nominal values for the deck and piers are accurate estimates since they are the moduli of elasticity of the concrete used in design for the deck and piers respectively, and therefore their updated values are expected to lie close to 1. It would be sensible to expect a slight lowering in the values of those parameters rather than an increase, since an increase would indicate that the structure has stiffened over time since its construction which would be counter-intuitive. However, as far as the soil is concerned, its nominal value is only a rough estimate since the soil is much more difficult to model compared to the bridge because its exact composition and spatial variation is not known. Therefore, its nominal value should not be trusted and rather it should be dealt with a large uncertainty in the model updating procedure. These modeling considerations regarding the initial parameter uncertainties are taken into account in the Bayesian framework through the prior PDF.

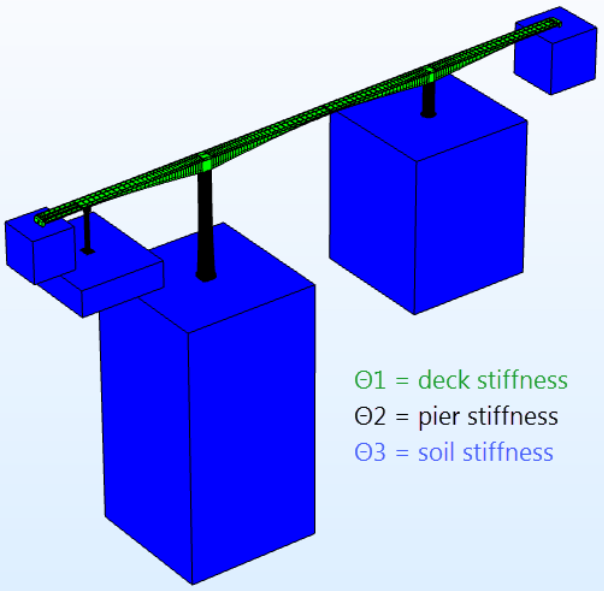


Figure 4.19: Parametrization of the FE model of the bridge-soil system.

The prior distribution for the parameters was assumed to be uniform with bounds in the domain $[0.1, 10] \times [0.1, 10] \times [0.1, 1000]$ for the deck, pier and soil parameters respectively, and in the domain $[0.001, 1]$ for the prediction error parameter σ. The domain for the soil parameter was deliberately chosen much larger in order to account for the large uncertainty in the values of the soil stiffness reported in the design and be able to explore the full effect of the soil stiffness on the model behavior.

Model updating results are obtained using the parallelized TMCMC algorithm for the bridge-soil FE model. The TMCMC is used to generate samples from the posterior PDF of the structural model and prediction error parameters. Those samples represent the posterior PDF and therefore our updated state of knowledge about the parameters given the experimental data. After the posterior samples are drawn which describe the updated uncertainty in the model parameters, this uncertainty is propagated to the predictions of the first 15 modal frequencies of the bridge. This is done in order to check the fit of the updated model with the experimental data that was used to perform the model updating. 1000 samples per TMCMC stage are used, resulting in a total runtime of approximately 10 minutes using the reduced 1.891 DOF model in a 8-core 3.20 GHz computer.

The TMCMC samples which represent the posterior PDF are visualized through their marginal distributions and 2D projections in Figure 4.20. The sample statistics are shown in Table 4.5. The posterior parameter uncertainty is propagated through the model using the samples to yield the robust model predictions of the 15 first frequencies which were used as data to update the model. The fit is shown in Figure 4.21. The improvement of the updated model compared to the nominal model is evident. All experimental frequencies except the 9th fall well within the 90% uncertainty interval of the model predictions and close to the mean. This is a strong indication of the accuracy and predictive capability of

the updated model. It is also interesting to note how the nominal model starts to miss the experimental frequencies significantly only after the 9th frequency. Up to the 8th frequency there is no significant improvement of the updated model over the nominal model. Also it is evident that the uncertainty in the predictions of the updated model starts to grow larger after the 8th frequency. This implies that the higher modes are more sensitive to variations in the values of the model parameters compared to the lower modes.

It can be seen that the values of the deck and pier parameters lie close to 1 as expected, and slightly below it. The updated most probable values for the deck and pier stiffness parameters are estimated to be approximately 0.90 and 0.95 times their nominal values with uncertainties of the order of 5% and 12% respectively. From the $\theta_1 - \theta_2$ 2D projection of Figure 4.20 it is evident that a negative correlation exists between the deck and pier stiffnesses. This is reasonable since an increase in the stiffness of the deck can be counterbalanced by a decrease in the stiffness of the piers such that the modal frequencies are maintained, and vice versa. In fact, part of the posterior uncertainty that exists in the values of the deck and pier stiffnesses is due to this correlation and not only due to the measurement and model error. The updated mean value for the prediction error parameter σ is 0.1 which indicates an average 10% mismatch between the frequency predictions from the 1000 samples and the experimental frequencies with respect to the experimental frequencies. This can be qualitatively seen in the fit of Figure 4.21 by comparing the 90% uncertainty interval of the predictions with the experimental frequencies.

As far as the updated soil stiffness is concerned, the only (but important) new information that is acquired by the model updating is that its value can be arbitrarily large, as long as it exceeds a threshold. The threshold value appears to be approximately 70 which is the minimum value that the updated soil parameter can attain, as seen from its posterior marginal distribution in Figure 4.20. A value of 70 implies a soil modulus of elasticity of $70GPa$ which is more than double of the updated (and nominal) value of the pier modulus of elasticity ($34GPa$). The soil parameter can increase arbitrarily above this value without affecting the fit with the experimental data, that is, without causing any variation in the predicted modal frequencies of the model. Considering that the uniform prior bound for the soil stiffness was set to [0.1, 1000] it is obvious that lower values which attribute to the soil some flexibility similar to that of the piers are avoided. Since the piers are embedded into the soil blocks and the later turn out to be much stiffer, the physical implication of this is that the soil behaves as a rigid body compared to the piers and therefore it completely restrains their movement. This is the reason that any large value of the soil stiffness makes no difference since the soil is already practically rigid compared to the piers and acts like a fixed constraint.

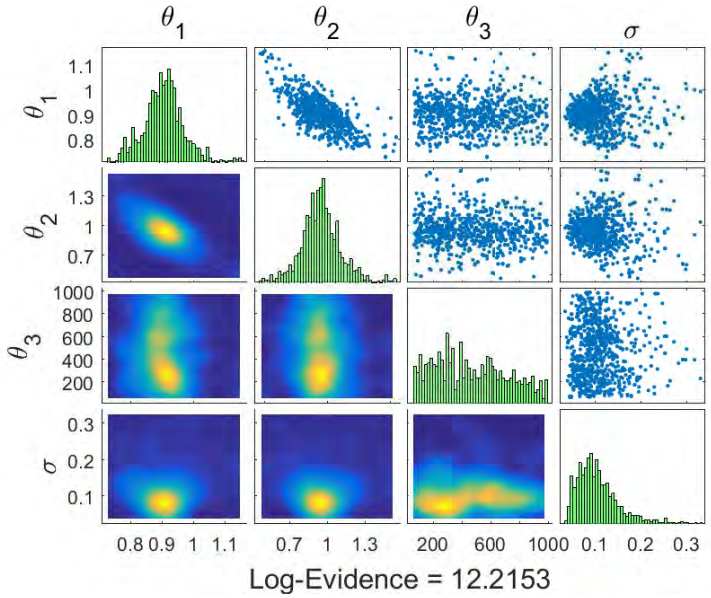


Figure 4.20: Posterior marginal distributions and 2D sample projections of model parameters. θ_1: Deck, θ_2: Piers, θ_3: Soil, σ: Prediction error

Table 4.5: Mean and quantiles of posterior samples.

	θ_1	θ_2	θ_3	σ
Mean	0.9070	0.9534	468.0994	0.1066
5% Quantile	0.8002	0.6997	113.8553	0.0508
95% Quantile	1.0117	1.2297	899.0796	0.1963

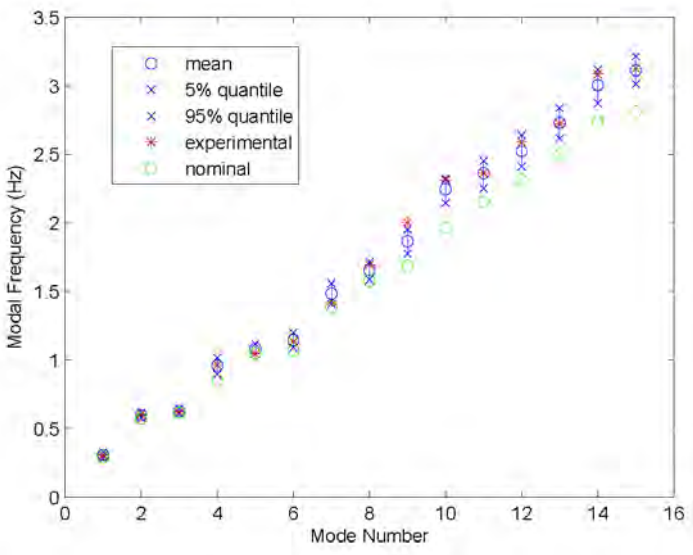


Figure 4.21: Uncertainty propagation to the first 15 modal frequencies compared with the experimental data and nominal model.

These results suggest that the bridge appears to be completely fixed to the ground and the soil exhibits no flexibility at all. This idea naturally leads to a second improved modeling option, which is to eliminate the soil parameter by fixing its value to a large number, simulating the rigid soil conditions which were found from the first model. Therefore, the new rigid soil model has two parameters instead of three, the modulus of elasticity of the deck (θ_1) and piers (θ_2), while the soil parameter is fixed to any value allowed by its posterior marginal distribution of Figure 4.20. A value of 300 is used. This rigid-soil model is expected to be better than the flexible-soil model since it has less parameters and makes the same predictions. Bayesian model selection (Beck and Yuen, 2004) is used as a rational framework to decide about which model is the best using the experimental data. This method compares several models using their evidence term $p(D|M_i)$. The model with the highest evidence is the best. This method takes into account both the complexity of the model in the form of the number of its parameters and the fit it achieves with the data in order to obtain a trade-off between the two. Models with more parameters are penalized unless they have a better fit with the data compared to models with less parameters. One merit of using the TMCMC algorithm for Bayesian model updating is that the evidence of the model is calculated as a byproduct of the algorithm. Therefore, by performing model updating on both models they can be easily compared using their evidence values. The log-evidence for the first flexible-soil model with three parameters was found to be 12.21 as shown in Figure 4.20.

The posterior PDF samples for the rigid-soil model with two parameters are visualized using their marginal distributions and 2D projections in Figure 4.22. The sample statistics are shown in Table 4.6. The posterior parameter uncertainty is propagated through the model using the samples to yield the

robust model predictions of the 15 first frequencies which were used as data to update the model. The fit is shown in Figure 4.23. It can be seen that, as expected, the model updating results both in terms of the updated values of the parameters and in terms of the fit with the data are almost identical to the results obtained from the flexible-soil model. The updated most probable values for the deck and pier stiffness parameters are again estimated to be approximately 0.90 and 0.95 times their nominal values with uncertainties of the order of 5% and 12% respectively, while the negative correlation between them still exists in the results. Note that in Figure 4.23 the predictions of the nominal model are larger due to the increase of the soil parameter to the fixed value of 300 in order to simulate the rigid-soil conditions which led to an increase of the modal frequencies of the nominal model. However, the updated model still provides the same fit with the data as was found with the flexible-soil model in Figure 4.21.

Bayesian model selection rewards the rigid-soil model for having one less parameter than the flexible-soil model while achieving the same fit with the experimental data by assigning a larger log-evidence of 14.99 to the rigid-soil model. The log-evidence for the flexible-soil model was found to be 12.21. Therefore, the Bayesian model selection methodology shows a clear preference of the rigid-soil model with two parameters over the flexible-soil model with three parameters.

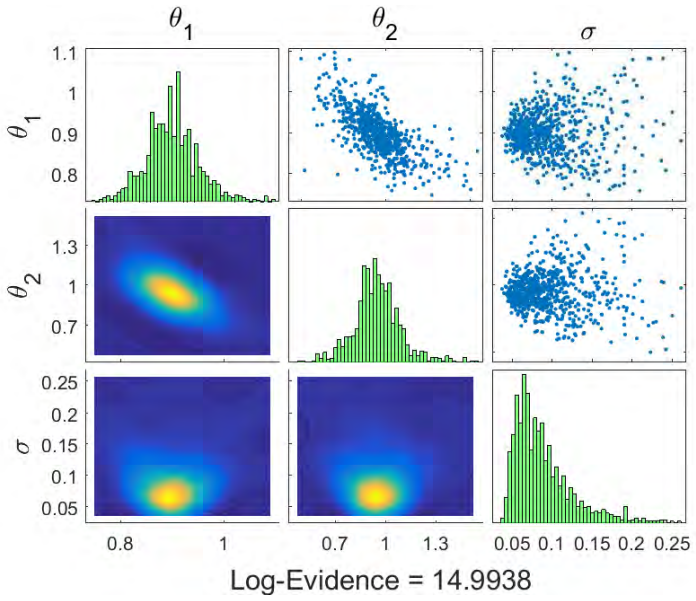


Figure 4.22: Posterior marginal distributions and 2D sample projections of model parameters. θ_1: Deck, θ_2: Piers, σ: Prediction error

Table 4.6: Mean and quantiles of posterior samples.

	θ_1	θ_2	σ
Mean	0.8997	0.9555	0.0904
5% Quantile	0.8180	0.7305	0.0483
95% Quantile	0.9888	1.2094	0.1714

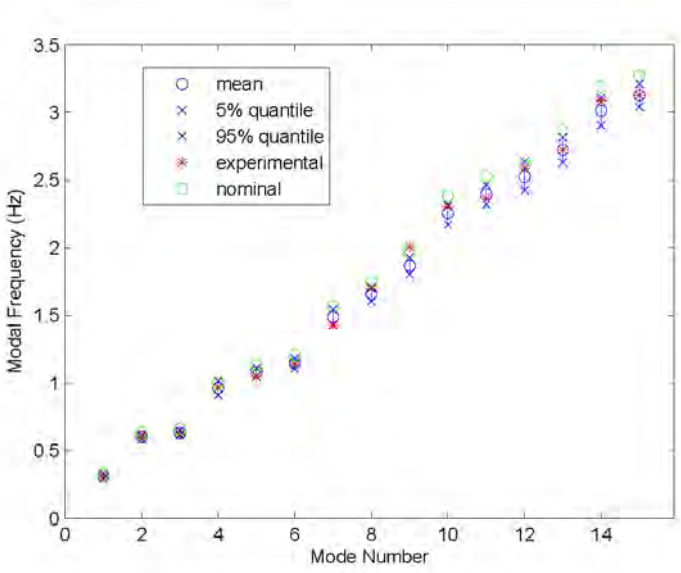


Figure 4.23: Uncertainty propagation to the first 15 modal frequencies compared with the experimental data and nominal model.

Up to this point the updated model was checked against data that were used to update it. To further check the accuracy, reliability and predictive capabilities of the model, it is checked against new data that was not used in the model updating procedure in a process that is known as model validation. The last 5 experimentally identified modal frequencies (16th-20th modes) and the 11 identified mode shapes (1st-12th except the 10th) are used for validating the model. The updated parameters and their uncertainty is propagated using the posterior samples through the rigid-soil model in order to make robust posterior predictions for the modal frequencies of the 16th-20th modes and for the mode shapes of modes 1-12, discarding the mode shape for the 10th mode. The robust posterior predictions are compared against their experimentally identified counterparts. The comparison is shown in Figure 4.24 for the modal frequencies and in Figure 4.25 for the mode shapes using their MAC value.

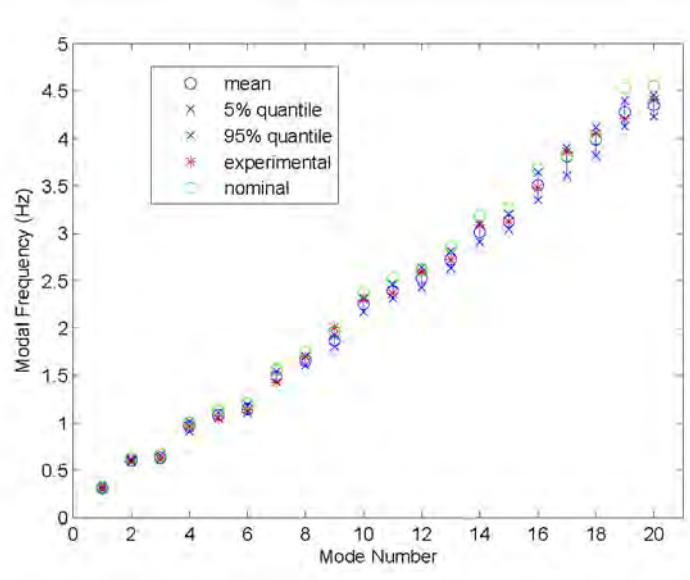


Figure 4.24: Uncertainty propagation to the 20 first modal frequencies compared with the experimentally identified frequencies. Modes 1-15 were used in the model updating while modes 16-20 were used for validation only.

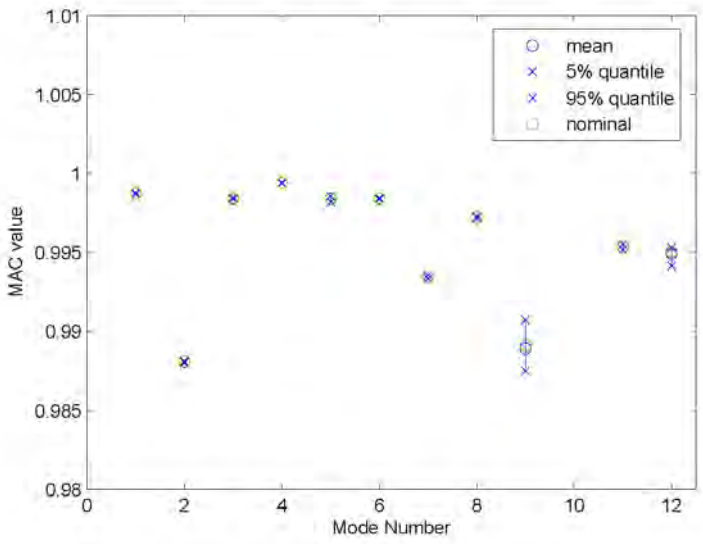


Figure 4.25: Uncertainty propagation to the first 12 mode shapes compared with the experimentally identified mode shapes using their MAC value.

It can be seen that the updated rigid-soil model continues to predict the modal frequencies for modes

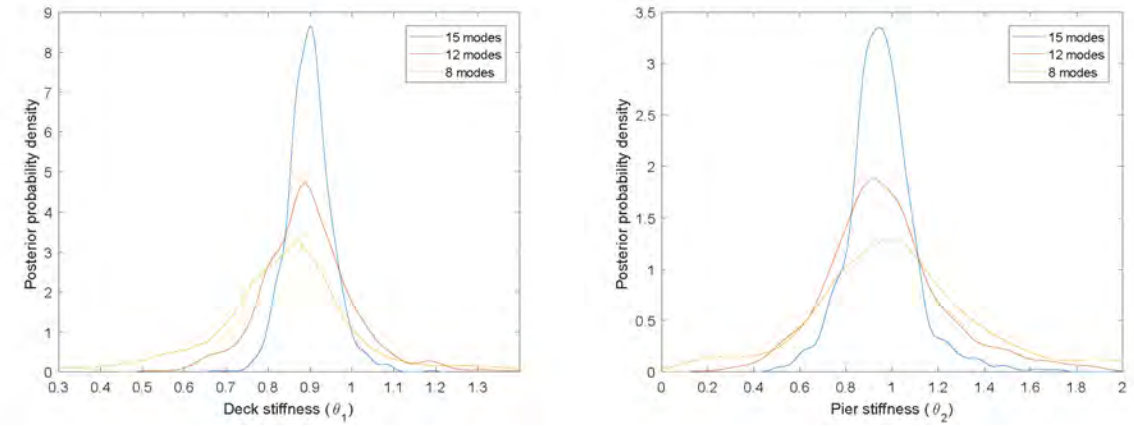
16-20 accurately, even though they were not used in the model updating. The experimental frequencies lie well within the 90% uncertainty intervals of the model predicted frequencies and close to the mean value. As far as the mode shapes are concerned, the MAC values between the experimental and updated rigid-soil model predicted mode shapes are exactly the same as the ones obtained with the nominal rigid-soil model. That is, the mode shapes were not affected at all by the new values of the model parameters and their uncertainty. This shows that the mode shapes are highly insensitive to variations in the model parameters. Also, the nominal MAC values were already very close to 1, indicating a perfect correlation with the experimental mode shapes, which means that they could not be improved in any significant amount. The invariability of the mode shapes with respect to the model parameters shows that they are not useful for model updating of the particular model since they do not reveal any information about the model parameters. The model validation results clearly demonstrate that the updated rigid-soil model is very accurate in describing the behavior of the real structure since it can accurately predict its modal frequencies and mode shapes that were not used in the model updating. The overall fit of the updated rigid-soil model with the experimental modal frequencies is quantitatively shown in Table 4.7.

Table 4.7: Fit of updated rigid-soil model with the experimental modal frequencies.

Mode	Experimental Frequency (Hz)	Updated Model Frequency (Hz)		Nominal Model Frequency (Hz)
	mean	mean	std	
1	0.3063	0.3093	0.0121	0.3209
2	0.6034	0.6010	0.0100	0.6300
3	0.6227	0.6315	0.0107	0.6611
4	0.9646	0.9634	0.0336	1.0007
5	1.0468	1.0833	0.0179	1.1362
6	1.1389	1.1464	0.0249	1.2052
7	1.4280	1.4879	0.0356	1.5667
8	1.6967	1.6583	0.0321	1.7448
9	2.0053	1.8661	0.0409	1.9646
10	2.3034	2.2545	0.0489	2.3796
11	2.3666	2.3946	0.0476	2.5233
12	2.5901	2.5265	0.0680	2.6149
13	2.7226	2.7240	0.0595	2.8630
14	3.0861	3.0145	0.0631	3.1870
15	3.1266	3.1259	0.0529	3.2720
16	3.4801	3.5113	0.0953	3.6789
17	3.8608	3.8107	0.0915	3.8180
18	4.0585	3.9898	0.0938	3.9993
19	4.2101	4.2790	0.0841	4.5334
20	4.4102	4.3538	0.0714	4.5547

As already mentioned, the data set used for model updating consists of the first 15 experimentally identified modal frequencies of the bridge. The model updating results, both in terms of the updated parameter values and in terms of the robust posterior predictions depend on the number of experimental modes used in the data set. In order to illustrate this dependence the model updating procedure is repeated two times, using as data sets the first 12 and 8 experimental modes. The results are compared with the results obtained using the 15 modes. Figure 4.26 shows the posterior marginal distributions for the deck and pier parameters using 15, 12 and 8 modes, while Figure 4.27 shows the robust posterior predictions for the first 20 modal frequencies for the 3 cases, compared with the experimentally identified frequencies. Quantitative results in terms of the mean and standard deviation of the updated parameter values and robust predictions are shown in Tables 4.8 and 4.9 respectively. It can be seen that as the

number of modes included in the data set decreases, the uncertainty in the deck and pier parameters increases. In fact, the posterior standard deviation for the deck and pier parameters almost doubles when reducing the number of modes from 15 to 12, and from 12 to 8. This increased uncertainty in the parameters gets propagated and results in an increased uncertainty in the predictions of the modal frequencies. Again it is seen that the standard deviation of the predicted modal frequencies gets doubled when reducing the number of modes from 15 to 12 and to 8. Table 4.10 shows the results in terms of the coefficient of variation measure, expressed as a percentage. This can be thought of as a dimensionless measure of the error in the modal frequency estimate. It can be seen that using 15 modes in the data set results in errors in the predicted modal frequencies of the order of $1-3\%$, while 12 modes lead to errors of the order of $3-5\%$ and for 8 modes the error in the modal frequency estimates can be as large as $7-9\%$.



(a) Deck parameter

(b) Pier parameter

Figure 4.26: Posterior marginal distributions using 15, 12 and 8 modes in the model updating data set.

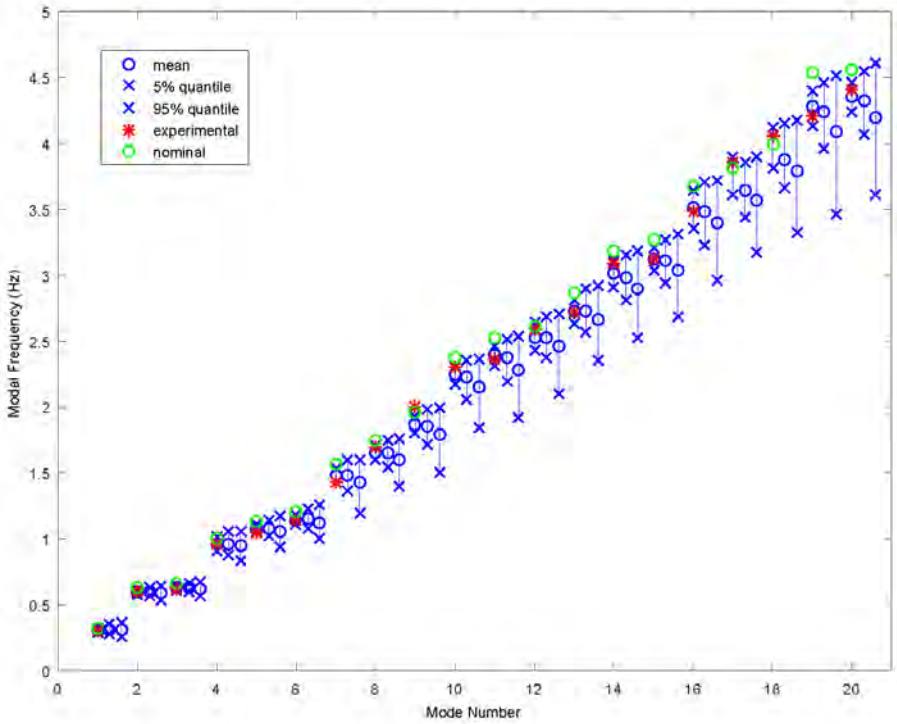


Figure 4.27: Robust posterior predictions for the first 20 modal frequencies using 15, 12 and 8 modes compared with the experimental frequencies.

Table 4.8: Deck and pier posterior uncertainties for 15, 12 and 8 modes used in the data set.

		Number of modes		
		15	12	8
Deck	mean	0.90	0.89	0.87
	std	0.05	0.11	0.26
Pier	mean	0.95	0.94	1.00
	std	0.15	0.33	0.70

Table 4.9: Robust posterior predictions for the first 20 modal frequencies for 15, 12 and 8 modes used in the data set.

Mode	Experimental Frequency	Updated Model					
		15 modes		12 modes		8 modes	
	mean	mean	std	mean	std	mean	std
1	0.3063	0.3093	0.0121	0.3115	0.0235	0.3135	0.0339
2	0.6034	0.6010	0.0100	0.6000	0.0192	0.5879	0.0384
3	0.6227	0.6315	0.0107	0.6312	0.0201	0.6230	0.0372
4	0.9646	0.9634	0.0336	0.9652	0.0536	0.9571	0.0748
5	1.0468	1.0833	0.0179	1.0830	0.0367	1.0606	0.0761
6	1.1389	1.1464	0.0249	1.1491	0.0505	1.1260	0.0836
7	1.4280	1.4879	0.0356	1.4821	0.0730	1.4280	0.1306
8	1.6967	1.6583	0.0321	1.6504	0.0616	1.6016	0.1242
9	2.0053	1.8661	0.0409	1.8578	0.0844	1.7932	0.1558
10	2.3034	2.2545	0.0489	2.2256	0.0881	2.1582	0.1777
11	2.3666	2.3946	0.0476	2.3727	0.0968	2.2869	0.1931
12	2.5901	2.5265	0.0680	2.5272	0.1025	2.4639	0.2034
13	2.7226	2.7240	0.0595	2.7293	0.1007	2.6617	0.1994
14	3.0861	3.0145	0.0631	2.9888	0.1063	2.9013	0.2210
15	3.1266	3.1259	0.0529	3.1121	0.1019	3.0342	0.2200
16	3.4801	3.5113	0.0953	3.4867	0.1492	3.4027	0.2749
17	3.8608	3.8107	0.0915	3.6443	0.1319	3.5639	0.2703
18	4.0585	3.9898	0.0938	3.8757	0.1521	3.7955	0.2982
19	4.2101	4.2790	0.0841	4.2359	0.1516	4.0897	0.3452
20	4.4102	4.3538	0.0714	4.3296	0.1449	4.1985	0.3372

Table 4.10: Coefficient of variation (= $std/mean \times 100\%$) of robust model predictions for 15, 12 and 8 modes.

Mode	Updated Model (cov)		
	15 modes	12 modes	8 modes
1	3.9121	7.5441	10.8134
2	1.6639	3.2000	6.5317
3	1.6944	3.1844	5.9711
4	3.4876	5.5533	7.8153
5	1.6524	3.3887	7.1752
6	2.1720	4.3947	7.4245
7	2.3926	4.9254	9.1457
8	1.9357	3.7324	7.7547
9	2.1917	4.5430	8.6884
10	2.1690	3.9585	8.2337
11	1.9878	4.0797	8.4437
12	2.6915	4.0559	8.2552
13	2.1843	3.6896	7.4915
14	2.0932	3.5566	7.6173
15	1.6923	3.2743	7.2507
16	2.7141	4.2791	8.0789
17	2.4011	3.6194	7.5844
18	2.3510	3.9245	7.8567
19	1.9654	3.5789	8.4407
20	1.6399	3.3467	8.0314

4.7 Conclusions

A Bayesian framework was presented for model updating and model selection in structural dynamics simulations using experimentally identified modal data. The proposed framework was used for estimating the uncertainties in the structural and prediction error model parameters and propagating those uncertainties for making robust predictions. Sampling algorithms were proposed for Bayesian uncertainty quantification and propagation. In Bayesian sampling techniques, the TMCMC is a highly parallelizable algorithm that can be used in a parallel computing environment to efficiently distribute the large number of independent system simulations in available multi-core CPUs.

The implementation of the framework in structural dynamics was outlined for linear structural models using experimentally identified modal frequencies and mode shapes. The model used was a high-fidelity

detailed FE model of the Metsovo bridge. To efficiently handle large-order models of hundreds of thousands or millions degrees of freedom, fast and accurate component mode synthesis techniques, consistent with the finite element model parameterization, were employed that achieve drastic reductions in the model order and the computational effort.

Bayesian model updating was performed in order to estimate the structural and prediction error model parameters and their associated uncertainty, while Bayesian model selection was used to decide which model is the best amongst two classes of competing models, based on the experimental data. The structural model parameter uncertainties were propagated in order to obtain robust posterior model predictions. The updated model was validated against data that was not used in the model updating procedure. Finally, an analysis was performed in order to investigate the effect of the number of modes used in the data set on the model updating and validation results.

It is demonstrated with this application on a full scale bridge that these parallel computing and model reduction techniques, integrated within Bayesian tools, can be effective in calibrating the uncertainty of finite element models with hundred of thousands of degrees of freedom, achieving drastic reductions in computational effort by more than three orders of magnitude. The integration of model reduction and parallel computing techniques within Bayesian uncertainty quantification and propagation tools can result in drastic reduction of computational time to manageable levels for complex models used for simulations of structural dynamics and related engineering systems.

Chapter 5

Bayesian identification of non-linear seismically isolated structures

5.1 Introduction

Bayesian inference is used for quantifying and calibrating uncertainty models in structural dynamics based on vibration measurements, as well as propagating these uncertainties in structural dynamics simulations for updating robust predictions of system performance, reliability and safety (Papadimitriou et al., 2001). Asymptotic and stochastic simulation tools can be used to quantify and propagate uncertainties. Asymptotic formulations (Tierney and Kadane, 1986) are approximate and may miss-represent the posterior distribution of the model parameters for concave supports, multimodal or unidentifiable posterior distributions. They also require that the derivatives and hessians of measures of fit between the response quantities of interest (QoI) and corresponding measurements are available. Adjoint formulations, in particular, can be used to substantially reduce the computations related to estimating the sensitivities of the measures of fit with respect to the number of parameters. For nonlinear models of structures subjected to earthquake-like excitation, the adjoint formulation might not exist or may not be easily implemented in software. Stochastic simulation algorithms are more convenient to use since can better represent complex posterior distributions and do not require the estimation of derivatives of response quantities with respect to the model parameters.

Computationally intensive stochastic simulation algorithms (e.g. Transitional MCMC (Ching and Chen, 2007)) are well suited tools for identifying system and uncertainty models as well as performing robust prediction analyses. The stochastic simulation tools involve generating samples for tracing and then populating the important uncertainty region in the parameter space, as well as evaluating integrals over high-dimensional spaces of the uncertain model parameters by sample estimates. They require a very large number of system analyses to be performed over the space of uncertain parameters which may lead to excessive computational time. Consequently, the computational demands depend highly on the

number of system analyses and the time required for performing a system analysis. This in turn depends on the complexity of the model of the analyzed system as well as the number of uncertain parameters involved. Efficient computing techniques have been integrated with the Bayesian framework to handle large-order models and localized nonlinear actions (Angelikopoulos et al., 2012). Specifically, surrogate models can be adopted to reduce the number of full system runs, and parallel computing algorithms can be used to efficiently distribute the computations in available multi-core CPUs (Angelikopoulos et al., 2012). Depending on the number of available computer workers, drastic reduction in computational effort to manageable levels can be achieved for models for which the time to execute one system simulation is of the order of several seconds, minutes or even hours.

This work is concerned with calibrating models of seismically isolated structures equipped with nonlinear isolator devices. Thus, the equations of motion are nonlinear due to the nonlinear models used for the isolator devices. Parameterized model structures of the isolator devices are introduced and the Bayesian framework is used to estimate the model parameter values and their associated uncertainties. The estimation is based on experimental data obtained by shake table tests on the combined bridge-isolator system. Calibration of the uncertainties in the model parameters is based on full response time histories predicted by the model and measured by a network of sensors. The Transitional Markov Chain Monte Carlo (TMCMC) algorithm (Ching and Chen, 2007) is used for identifying system and uncertainty models as well as for performing robust prediction analyses. This algorithm is used to represent the posterior distribution of the parameters of the non-linear isolation system, as well as propagate this uncertainty to obtain the uncertainty in response quantities of interest. For nonlinear models of the type analyzed in this work, the time to execute one simulation of the system is of the order of a few seconds which poses moderate computational requirements. However, parallel computing algorithms (Angelikopoulos et al., 2012) are used to efficiently distribute the computations in available multi-core CPUs, thus achieving further computational savings.

5.2 Bayesian formulation using response time histories

5.2.1 Parameter estimation

The Bayesian methodology (Beck and Katafygiotis, 1998; Yuen, 2010) is used to calibrate and estimate the uncertainties in the parameters appearing in the non-linear models describing the behavior of the bridge under seismic excitation, using measured response time histories of displacements, accelerations, and forces at different parts of the bridge. According to the methodology, if $\underline{\theta}$ denotes the vector of the parameters of the non-linear models, and $\underline{\sigma}$ denotes an additional set of prediction error parameters to be defined later, the uncertainty in these parameters given the measured data D is quantified by the posterior distribution that is obtained from Bayes' theorem as

$$p\left(\underline{\theta},\underline{\sigma}|D\right)=\frac{p\left(D|\underline{\theta},\underline{\sigma}\right)\pi\left(\underline{\theta},\underline{\sigma}\right)}{p\left(D\right)} \tag{5.1}$$

where$p\left(D|\underline{\theta},\underline{\sigma}\right)$ is the likelihood, $\pi\left(\underline{\theta},\underline{\sigma}\right)$ is the prior distribution of the uncertain parameters and$p\left(D\right)$ is the evidence of the model. The posterior distribution $p\left(\underline{\theta},\underline{\sigma}|D\right)$ quantifies the uncertainty in the model and prediction error parameters by measuring the plausibility of each possible set of parameters given the data D.

To apply the Bayesian formulation for parameter calibration of non-linear models, we consider that the data consists of measured time histories $D = \{\hat{x}_j\left(k\right) \in R,\ j = 1, \cdots, N_0,\ k = 1, \cdots, N_D\}$ at time instances $t = k\Delta t$, of N_0 response quantities (displacements, accelerations and forces) at different points in the bridge, where N_D is the number of the samples data using a sampling period Δt.

Let also $\{x_j\left(k;\underline{\theta}\right) \in R,\ j = 1, \ldots, N_0,\ k = 1, \ldots, N_D\}$ be the predictions of the response time histories for the same quantities and points in the structure, from the non-linear model corresponding to a particular value of the parameter set $\underline{\theta}$. The prediction error equation between the sampled response time history of the quantity j at time $t = k\Delta t$ and the corresponding response time history predicted from the model for a particular value of the parameter set $\underline{\theta}$ can now take the form

$$e_j\left(k\right) = \hat{x}_j\left(k\right) - x_j\left(k;\underline{\theta}\right) \tag{5.2}$$

where $j = 1, \ldots, N_0$ and $k = 1, \ldots, N_D$.

Prediction errors, measuring the fit between the measured and the model predicted response time histories, are modeled by Gaussian distributions. The difference between the measured and model predicted response is attributed to both experimental errors and modeling error. The prediction errors of a response time history at different time instants are assumed to be independent zero-mean Gaussian variables with equal variances for all sampling data of a response time history, but each time history is allowed to have a different prediction error associated with it. This formulation takes into account the fact that each measured time history is generally obtained from a different sensor (displacement, acceleration or force sensor) with a different accuracy and noise level, and this results in a number of prediction errors equal to the number of measured time histories.

Under the zero-mean Gaussian assumption for the prediction error, $e_j\left(k\right) \sim N\left(0,\sigma_j^2\right)$, the measured quantity $\hat{x}_j\left(k\right)$ also follows a Gaussian distribution with mean $x_j\left(k;\underline{\theta}\right)$ and variance σ_j^2, $\hat{x}_j\left(k\right) \sim N\left(x_j\left(k;\underline{\theta}\right),\sigma_j^2\right)$. In the analysis that follows, the prediction error parameters $\sigma_j,\ j = 1, \ldots, N_0$ are contained in the prediction error vector $\underline{\sigma} \in R^{N_0}$. The prediction error e_j provides a measure of the discrepancy between the measured and model predicted quantities. As already stated, this generally breaks down to two terms for the prediction error, one for the experimental error and one for the model error. In this study such a distinction is not made, and the prediction error is thought of as a measure of the total discrepancy between measurements and the model predictions without being able to distinguish how much is due to experimental or modeling error. Depending on the problem, and more specifically on the way the data was collected, $\underline{\sigma}$ might be considered known or unknown. In the most general case it is considered unknown and therefore is included in the parameters for calibration, along with the structural model parameters. Herein, the prediction error parameters are considered unknown and from now on are included in the parameters to be calibrated given the data, along with the structural model parameters

in the set $\underline{\theta}$.

The likelihood function $p\left(D|\underline{\theta},\underline{\sigma}\right)$, which quantifies the probability of obtaining the data given a specific set of structural parameters and prediction error parameters, is derived by noting that the measured time histories $\hat{x}_j\left(k\right)$ are implied from (5.2) to be independent Gaussian variables with mean $x_j\left(k;\underline{\theta}\right)$ and variance σ_j^2. Taking advantage of the independence of the measured quantities both at different time instants of the same time history as well as between different time histories, the likelihood is formulated as follows:

$$p\left(D|\underline{\theta},\underline{\sigma}\right)=\prod_{j=1}^{N_0}\prod_{k=1}^{N_D}p\left(\hat{x}_j\left(k\right)|\underline{\theta},\underline{\sigma}\right) \tag{5.3}$$

Substituting with the formula for the Gaussian probability density function and rearranging terms one obtains that

$$p\left(D|\underline{\theta},\underline{\sigma}\right)=\frac{1}{\left(\sqrt{2\pi}\right)^{N_DN_0}\prod_{j=1}^{N_0}\sigma_j^{N_D}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{N_0}\frac{1}{\sigma_j^2}\sum_{k=1}^{N_D}\left[\hat{x}_j\left(k\right)-x_j\left(k;\underline{\theta}\right)\right]^2\right\} \tag{5.4}$$

Introducing the overall fit function

$$J\left(\underline{\theta};\underline{\sigma}\right)=\frac{1}{N_0}\sum_{j=1}^{N_0}\frac{1}{\sigma_j^2}J_j\left(\underline{\theta}\right) \tag{5.5}$$

where

$$J_j\left(\underline{\theta}\right)=\frac{1}{N_D}\sum_{k=1}^{N_D}\left[\hat{x}_j\left(k\right)-x_j\left(k;\underline{\theta}\right)\right]^2 \tag{5.6}$$

represents the measure of fit between the measured and the model predicted response time history for quantity j, the likelihood function can be compactly written in the form

$$p\left(D|\underline{\theta},\underline{\sigma}\right)=\frac{1}{\left(\sqrt{2\pi}\right)^{N_DN_0}\prod_{j=1}^{N_0}\sigma_j^{N_D}}\exp\left\{-\frac{N_DN_0}{2}J\left(\underline{\theta};\underline{\sigma}\right)\right\} \tag{5.7}$$

Substituting (5.7) in (5.1) one derives the posterior probability distribution of the parameters in the form

$$p\left(\underline{\theta},\underline{\sigma}|D\right)=\frac{\pi\left(\underline{\theta},\underline{\sigma}\right)}{p(D)\left(\sqrt{2\pi}\right)^{N_DN_0}\prod_{j=1}^{N_0}\sigma_j^{N_D}}\exp\left\{-\frac{N_DN_0}{2}J\left(\underline{\theta};\underline{\sigma}\right)\right\} \tag{5.8}$$

Herein the Transitional MCMC (Ching and Chen, 2007) is used to draw samples from the posterior distribution.

5.2.2 Uncertainty propagation

The main interest in updating the uncertainty in the model parameters through measured data lies in propagating this uncertainty through the model to quantify the uncertainty in output quantities of interest (QoI). Specifically, given a scalar output QoI $g\left(\underline{\theta}\right)$ which depends on the uncertain parameters $\underline{\theta}$,

the objective is to evaluate its uncertainty given the uncertainty in $\underline{\theta}$ which is quantified by the posterior PDF in (5.8). Simple measures of uncertainty in $g(\underline{\theta})$ are the mean, given by

$$Mean\left[g\left(\underline{\theta}\right)\right] = E\left[g\left(\underline{\theta}\right)\right] = \int g\left(\underline{\theta}\right) p\left(\underline{\theta}|D\right) d\underline{\theta} \tag{5.9}$$

and the variance, given by

$$Var\left[g\left(\underline{\theta}\right)\right] = E\left[g^2\left(\underline{\theta}\right)\right] - E^2\left[g\left(\underline{\theta}\right)\right] \tag{5.10}$$

where

$$E\left[g^2\left(\underline{\theta}\right)\right] = \int g^2\left(\underline{\theta}\right) p\left(\underline{\theta}|D\right) d\underline{\theta} \tag{5.11}$$

is the second moment of $g(\underline{\theta})$.

Stochastic simulation algorithms are used to provide sample estimates of the integrals (5.9) and (5.11). Specifically, the Transitional Markov Chain Monte Carlo (TMCMC) algorithm (Ching and Chen, 2007) is used. The samples $\underline{\theta}^{(j)},\ j = 1, \ldots, N$ drawn from the posterior PDF $p(\underline{\theta}|D)$ can be used to approximate the integrals (5.9) and (5.11) with the sample estimates

$$E\left[g\left(\underline{\theta}\right)\right] \approx \frac{1}{N}\sum_{j=1}^{N} g\left(\underline{\theta}^{(j)}\right) \tag{5.12}$$

and

$$E\left[g^2\left(\underline{\theta}\right)\right] \approx \frac{1}{N}\sum_{j=1}^{N} g^2\left(\underline{\theta}^{(j)}\right) \tag{5.13}$$

The samples $\underline{\theta}^{(j)},\ j = 1, \ldots, N$ are dependent, but are used for statistical averaging as if they were independent, accepting a reduced accuracy in the sample estimate.

5.3 Description of model and instrumentation

5.3.1 Description of bridge and isolation system

The physical system under examination is a scaled model of a bridge which was used to develop and test advanced sliding isolation systems for bridges (Figure 5.1) (Tsopelas et al., 1994). The isolation system consists of flat sliding bearings, rubber restoring force devices and linear viscous fluid dampers. The bridge model is seismically excited on a shake table, and various response quantities are measured by an array of force, displacement and acceleration sensors located in different parts of the structure. The purpose of the isolation system is to prohibit certain response quantities from reaching critical values under seismic excitation. Figure 5.2 shows a schematic diagram of the bridge model and the isolation system.

The studied isolation system consists of three components:

1. Flat sliding bearings to support the weight of the deck and provide a mechanism for energy dissipation.

2. Rubber devices for providing restoring force, that is, re-centering capability.

3. Linear viscous fluid dampers for enhancing the energy dissipation capability of the isolation system.

The three components of the isolation system provide load carrying capacity, restoring force capability (stiffness) and hysteretic and viscous damping which were not interrelated. The design requirements of the isolation system was to minimize the transmission of force to the substructure, that is piers and foundation, while bearing displacements in the scale of the model (length scale factor equal to 4) did not exceed 50 millimeters.

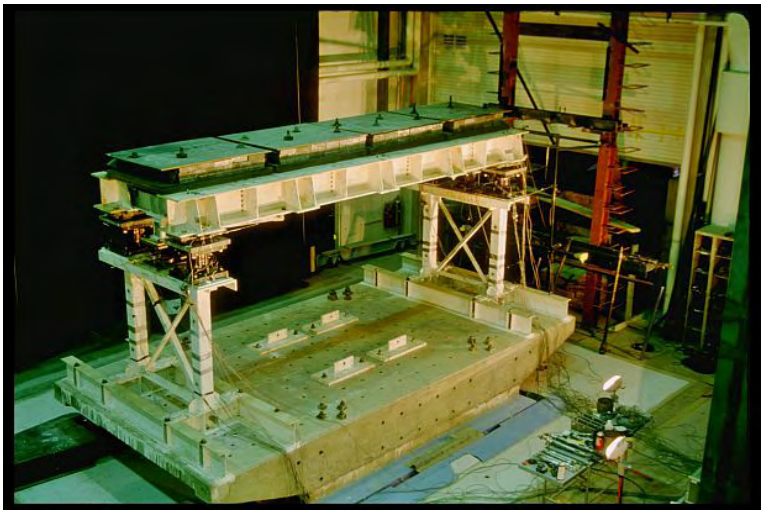


Figure 5.1: View of bridge model with isolation system.

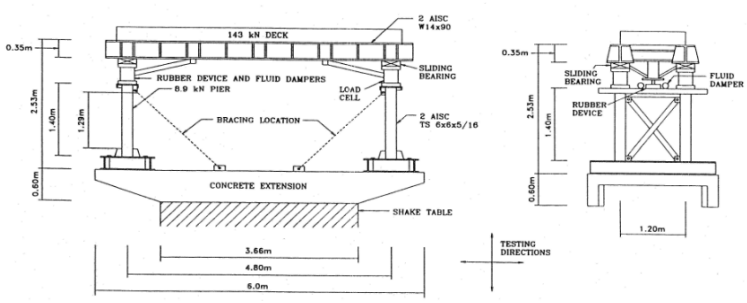


Figure 5.2: Schematic diagram of bridge model and isolation system.

5.3.2 Instrumentation

The instrumentation consisted of load cells, accelerometers, displacement transducers and strain gages. Figure 5.3 shows a schematic diagram of the overall instrumentation, where the abbreviations stand

for: A (acceleration), D (displacement), V-H-T (vertical-horizontal-transverse), D-P (deck-pier), E-W-N-S (east-west-north-south). Not all channels displayed in Figure 5.3 were used for Bayesian parameter calibration. The bearing displacement was monitored by displacement transducers, while the pier shear force was measured by strain gages installed in the pier. The acceleration of the piers measured by accelerometers was also utilized in the Bayesian procedure. Finally, two quantities were not directly measured but rather inferred from other direct measurements. These two quantities were the pier drift and the pier isolator force. The pier drift was measured by subtracting the displacement at the bottom of the pier from the displacement at the top of the pier. The pier isolator force was calculated by adding the frictional force of the bearings (which was measured directly from load cells supporting the bearings) and the combined force from the rubber devices and fluid dampers. The previously mentioned direct and indirect measurements provided the experimental time histories which constitute the data set D, consisted of $N_0 = 5$ response time histories, introduced in the Bayesian formulation. The seismic excitation provided by the shake table was also recorded in order to be used in the analytical model simulations. Details about the experimental setup and the analytical model can be found in Tsopelas et al. (1994).

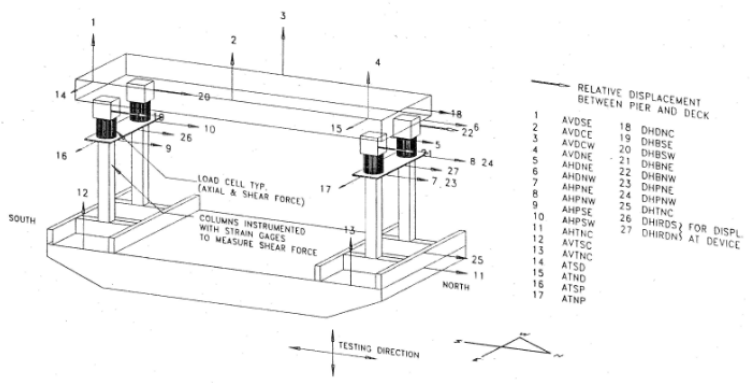
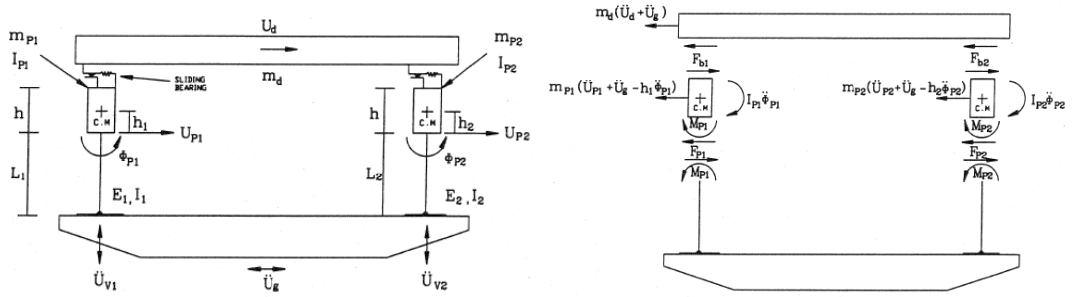


Figure 5.3: Overall instrumentation diagram.

5.3.3 Analytical model

An analytical model was developed in order to predict the behavior of the bridge under different seismic excitations. The model predicts the time histories at locations where measurements are available, given the values of the parameters which appear in the nonlinear model equations. The analytical model accounts for the pier flexibility, pier top rotation, vertical motion effects on the properties of the sliding bearings, and nonlinear hysteretic characteristics of the restoring force devices. The degrees of freedom are selected to be the deck displacement with respect to the table, U_d, the pier displacements with respect to the table, U_{pi}, and the pier rotations, ϕ_{pi} (see Figure 5.4). Each pier is modeled by a beam element of length L_i, moment of inertia I_i, and modulus of elasticity E_i, $i = 1, 2$. The beam element is fixed to

the table and connected at its top to a rigid block of height h, mass m_{pi}, and mass moment of inertia about the center of mass I_{pi}. The center of mass is located at distance h_i from the bottom of the block. This block represents the pier top. The equations of motion are derived by consideration of dynamic equilibrium of the deck and piers in the horizontal direction and of the piers in the rotational direction. Free body diagrams of the deck and pier tops of the bridge model are shown in Figure 5.4.



(a) (b)

Figure 5.4: *Left:* Degrees of freedom of the analytical model. *Right:* Free body diagrams of deck and piers.

The dynamic equilibrium of forces includes the lateral forces in the isolation system which include friction in the sliding bearings, force from the restoring force devices and the forces from the fluid dampers. These forces are summarized in the total isolation system force as follows:

$$F_{bi} = \mu_i \left(\dot{U}_{bi}\right) W_i + F_{ri}\left(U_{bi}\right) + nF_{di}\left(\dot{U}_{bi}\right) \tag{5.14}$$

where μ_i is the coefficient of sliding friction at pier i which is a function of the sliding velocity of the bearing $\dot{U}_{bi}$, W_i is the normal load on two sliding interfaces at pier i, F_{ri} is the restoring force from the rubber device at pier i, n is the number of fluid dampers at pier i, $(n = 2)$ and F_{di} is the damping force of one fluid damper at pier i. The bearing displacement at pier i, U_{bi} is related to the model degrees of freedom with the geometrical relation:

$$U_{bi} = U_d - U_{pi} + h\phi_{pi} \tag{5.15}$$

The coefficient of sliding friction in the sliding interface of the bearings followed the relation (Tsopelas et al., 1994):

$$\mu_i = f_{\max,i} - (f_{\max,i} - f_{\min,i}) \exp\left(-a_i \left|\dot{U}_{bi}\right|\right) \tag{5.16}$$

where $f_{\max,i}$ is the coefficient of friction at high velocity of sliding in pier i, $f_{\min,i}$ is the coefficient of friction at essentially zero velocity of sliding in pier i, a_i is a parameter controlling the variation of the coefficient of friction with velocity of sliding and $\dot{U}_{bi}$ is the velocity of sliding. The parameters $f_{\max,i}$, $\Delta f = f_{\max,i} - f_{\min,i}$ and a_i for $i = 1, 2$ form the first six elements respectively of the uncertain parameter vector $\underline{\theta}$.

The rubber restoring force devices acted as horizontal springs with displacement restraint. They provide stiffness by deforming their inner rubber elements up to a point where the device exhibits increased stiffness and acts as a displacement restrainer. The maximum allowed displacement is 50 millimeters. The devices exhibited nearly linear behavior to displacements of about 35 millimeters. Beyond this limit they displayed increasing stiffness to the limit of about 50 millimeters. After that the devices exhibited nearly rigid behavior. Rubber devices were installed one at each pier location. Due to the non-linear force-displacement relationship of the rubber restoring force devices, higher order polynomials were used to model their behavior.

Specifically, the restoring force F_{ri} in the rubber device at pier i was described by two polynomials as follows:

$$F_{ri} = F_{oi}\left(U_{bi}\right) + F_{Di}\left(U_{bi}\right) \tag{5.17}$$

where $F_{oi}\left(U_{bi}\right)$ is the displacement-dependent skeleton curve and $F_{Di}\left(U_{bi}\right)$ is, also displacement-dependent, the difference between the loading and unloading branches of the hysteresis loop of a rubber restoring force device. $F_{oi}\left(U_{bi}\right)$ and $F_{Di}\left(U_{bi}\right)$ may be expressed as odd and even polynomial functions of displacement, respectively:

$$\begin{aligned} F_{oi} &= \sum_{n=1,3,5,\ldots}^{N} A_n U_{bi}^n \\ F_{Di} &= \sum_{m=0,2,4,\ldots}^{M} B_m U_{bi}^m \end{aligned} \tag{5.18}$$

where the coefficients of the polynomials, A_n and B_m could also be part of the uncertain parameter vector $\underline{\theta}$. In this study the values of these parameters are kept fixed to the values calibrated using component test measurements. Sufficiently good agreement with experimental force-displacement loops (Tsopelas et al., 1994) was achieved by using $N = 11$ and $M = 4$. The values of the parameters obtained from the component tests were used as nominal values in the Bayesian procedure.

The third term in the right hand side of (5.14) is the force delivered by the fluid damper. The fluid viscous damper operates on the principle of fluid flow through orifices. The device consists of a stainless steel piston, with bronze orifice head and an accumulator. It is filled with silicone oil. Unlike typical fluid dampers which utilize cylindrical orifices, this device utilizes a series of specially shaped passages to alter flow characteristics with fluid speed. The damper was modeled using a Maxwell arrangement of a purely elastic spring and a purely viscous dashpot connected in series. The force output F_{di} of the Maxwell model can be represented by the differential equation:

$$\frac{\dot{F}_{di}}{K_{di}} + \frac{F_{di}}{C_{0i}} = \dot{U}_{bi} \tag{5.19}$$

where K_{di} is the elastic modulus of the spring and C_{0i} is the coefficient of viscosity. The parameters K_{di} and C_{0i} are parts of the uncertain parameter vector $\underline{\theta}$.

5.4 Results

The parameters $\underline{\theta}$ are associated with the structural parameters introduced in the model equations (5.16) and (5.19). The correspondence of the parameters in the set $\underline{\theta}$ and the parameters in the model equations (5.16) and (5.19) is given in Table 5.1. Specifically, the first six parameters are associated with the sliding friction isolators located at the left and right edge of the bridge deck, while the next four parameters are associated with the two parameters of the viscous dampers. The nominal values of the model parameters in (5.16) and (5.19) are set to the ones obtained by calibrating the models using measurements from component tests (Tsopelas et al., 1994). The parameters $\underline{\theta}$ scale the nominal values of the properties that they model so that a value of the parameters equal to one corresponds to the nominal values of the isolator parameters in the nonlinear models (5.16) and (5.19). The parameters of the rubber devices in equation (5.18) are kept fixed to their nominal values. The data set D is constituted from experimental measurements of bearing displacement, pier total isolator force, pier drift, pier shear force, and pier acceleration obtained during the shake table tests of the combined bridge-isolators system.

Table 5.1: Definition of the model parameters.

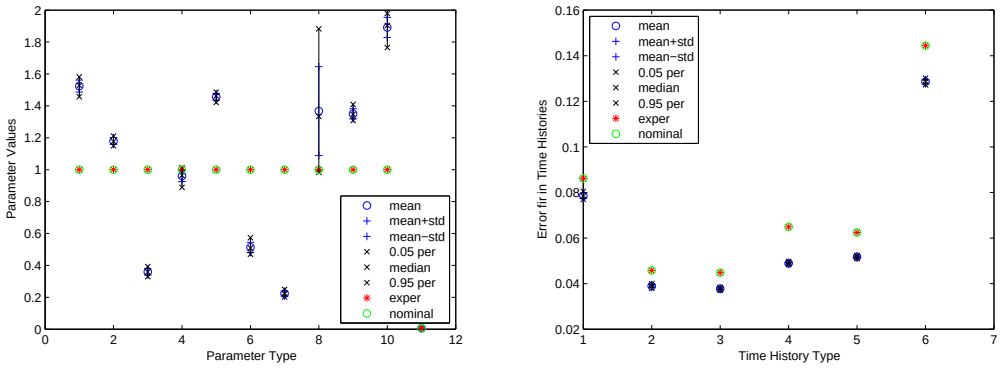
Parameter	***Type***	***Parameter in*** (5.16) ***or*** (5.19)
θ_1	Friction	$f_{\min,1}$
θ_2	Friction	Δf_1
θ_3	Friction	a_1
θ_4	Friction	$f_{\min,2}$
θ_5	Friction	Δf_2
θ_6	Friction	a_2
θ_7	Damper	K_{d1}
θ_8	Damper	C_{01}
θ_9	Damper	K_{d2}
θ_{10}	Damper	C_{02}

The parallelized TMCMC (Angelikopoulos et al., 2012) is used to draw samples from the importance region of the posterior probability density function. Parameter estimation results were obtained with 1000 samples per stage in the TMCMC algorithm and eight computer workers were used to perform in parallel the computations involved in the TMCMC algorithm. The number of stages required to compute the solution is of the order of 20 to 25. The computer programming environment where the model is implemented is also of great importance since it plays a crucial role in the time required to perform a model evaluation for a specific set of parameters, which is required by the stochastic simulation algorithms. In this work, the structural model is implemented in the Fortran programming language which enables it to be very fast and efficient for use in the TMCMC algorithm. The computational time for each simulation

of the system is approximately 10 seconds, resulting in a time-to-solution of the order of several hours.

Figure 5.5(left) shows the uncertainty in the marginal distributions of the 10 model parameters in the friction and damping models (5.16) and (5.19), respectively. Parameter number 11 is the prediction error parameter σ, which was assumed to be the same for all the measured response time histories. The measures of fit $J_j\left(\underline{\theta}\right)$, $j = 1, \ldots, 6$, between the measured response time histories and the model predicted response time histories for the TMCMC samples are given in Figure 5.5(right), along with the statistics (5% quantile, mean - std, mean, mean + std, 5% quantile) computed from propagating the TMCMC samples. The marginal distributions of the parameters are demonstrated in Figure 5.6.

It is clear from these figures that uncertainties in the model parameters vary from 5% level to as high as 50% for the damping coefficient C_{01} for the viscous damper model. The marginal distributions of the parameters Δf_1 and Δf_2 of the sliding friction models are estimated to be bimodal for both friction devices. These uncertainties are expected to affect the predictions of various response quantities of interest that are critical to the performance of such systems. From Figure 5.5(right) it can be seen that compared to the nominal nonlinear models there is a reduction in the error between the experimental time histories and the time histories predicted by the calibrated models. This reduction varies from 5% to 30% depending the measured response quantity considered.



(a) (b)

Figure 5.5: *Left:* Calibrated model parameter values. *Right:* Fit with experimental time histories.

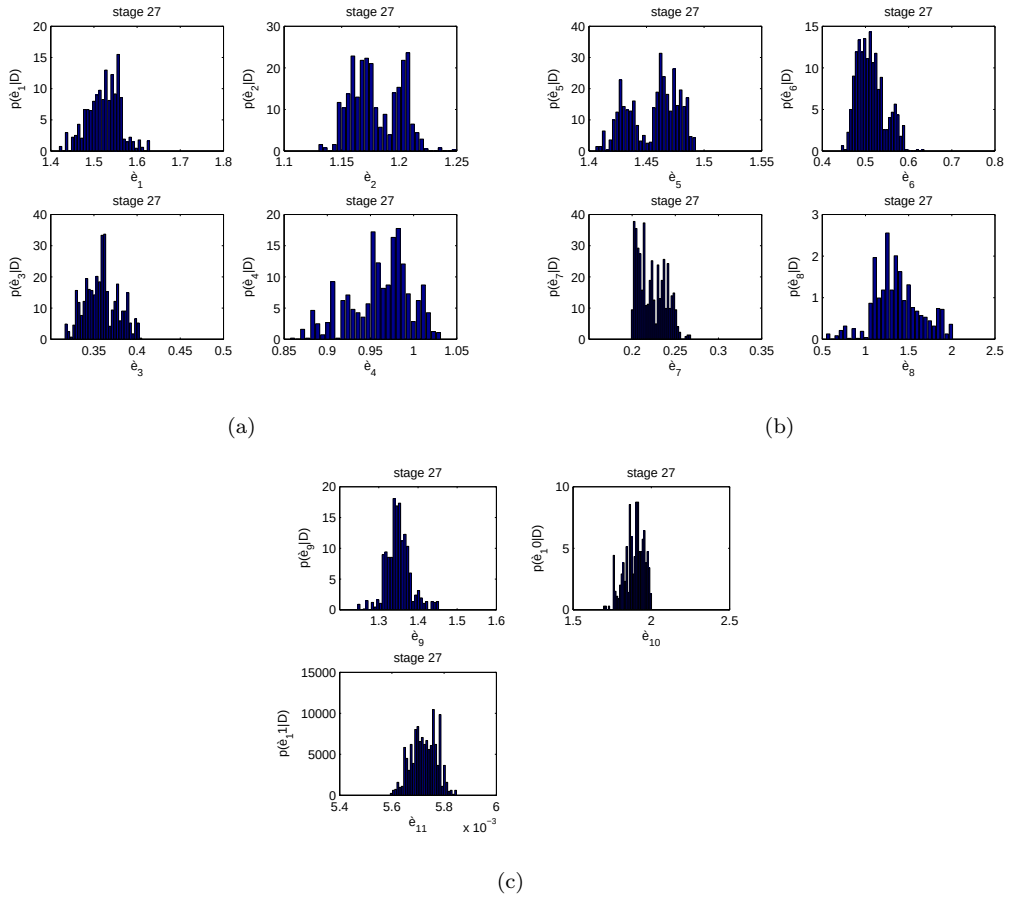


Figure 5.6: Marginal distributions of the uncertain model parameters.

The uncertainties of the model parameters are propagated through the model to estimate the uncertainties in the output response time histories. This was achieved by using the samples drawn from the posterior probability density function to estimate the mean and standard deviation of a function of the uncertain parameters, using (5.12) and (5.13). The robust model predicted time histories are compared with the experimental measurements in Figure 5.7 for four selected response quantities of interest. It is observed that the model predictions adequately fit the measured time histories for all response time histories considered in the calibration. Discrepancies between the experimental measurements and the robust model predictions for the isolator displacement could be attributed to the inadequacy of the isolator models to capture all characteristics of the isolator devices. Also it should be observed that the uncertainty in the predictions of the response time histories is very small compared to the uncertainty in the model parameters shown in Figure 5.5(left) and Figure 5.6.

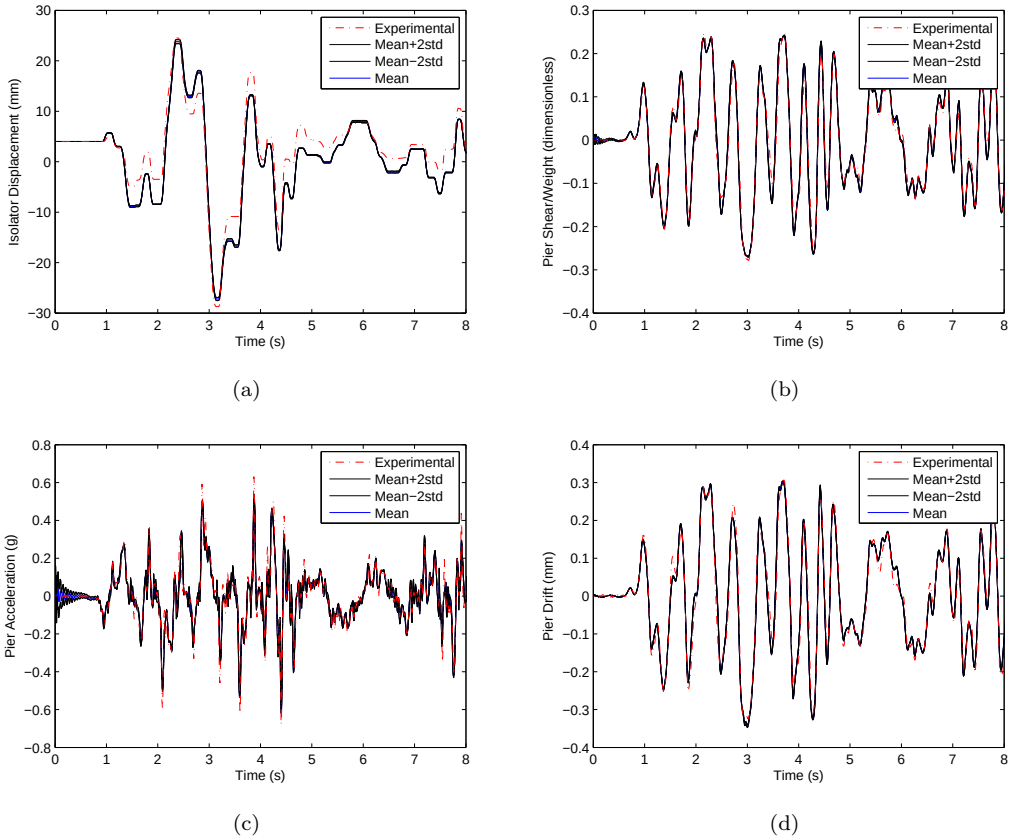


Figure 5.7: Uncertainty propagation: Response uncertainty along with experimental data. Isolator Displacement, Pier Shear force, Pier Acceleration, Pier Drift.

5.5 Conclusions

A Bayesian uncertainty quantification and propagation framework was presented for estimating the parameters of nonlinear models using response time history measurements. The framework was applied to calibrate the parameters of the nonlinear models involved in representing the behavior of a seismically isolated bridge using experimentally measured response time histories. Stochastic simulation algorithms were used to estimate the uncertainty in the model parameters and propagate them through the model to estimate the quality in the fit between the model predicted and measured response time histories. It is found that the identified values of the model parameters are different from the nominal values obtained by calibration procedures at the component level using component tests. This introduces the need for calibrating the model parameters taking into account component and system tests simultaneously. The nonlinear models introduced for the three types of isolator devices are adequate to represent the behavior of the isolated bridge since there is a very good match between the model predicted responses with

the measured responses considered in the Bayesian parameter estimation. Also, the uncertainty in the model parameters is significant for some of the isolator devices. This is expected to affect uncertainties in predictions of critical response quantities of interest. The theoretical and computational developments used in this work can be used to identify and propagate uncertainties in large order nonlinear systems consisting of linear and nonlinear components.

Chapter 6

Bayesian optimal experimental design

6.1 Introduction

Model-based Optimal Experimental Design (OED) is concerned with finding which is the best way to perform an experiment so that a specific purpose is achieved, using a model of the system as a guide. Common purposes include: model parameter inference and making predictions using the model (Chaloner and Verdinelli, 1995). Herein we are interested in model parameter inference as our goal. That is, we seek experimental designs which will result in experimental data that are the most useful for identifying the parameters of the model. The model parameters are uncertain, as are the experimental data, since no experiment has taken place at the time of the design. These uncertainties are treated with the Bayesian framework for uncertainty quantification (Beck and Katafygiotis, 1998), by assigning a prior probability density function (PDF) for the parameters and a probabilistic model for the difference between the model predictions and the data, known as the prediction error. The prior PDF of the model parameters reflects our state of knowledge for the values of the parameters before any data collection experiment takes place. It plays a very important role in the OED process since samples from the prior are used to simulate data samples. The prediction error also plays a very important role since it quantifies how much the experimental data can deviate from the model predictions and is a measure of how good our model is.

Since our objective is to learn about something by observing something else, the most intuitive field to be used as a framework is *Information Theory*. The three basic quantities of Information Theory are: *Entropy*, *Relative Entropy (KL Divergence)* and *Mutual Information* (Cover and Thomas, 2006). These quantities have properties which relate them with each other, e.g. mutual information can be written either as a KL-divergence or as a difference between two entropies in two ways (Cover and Thomas, 2006), and the objective function can be written as a multi-dimensional integral using one of these quantities. These different but equivalent forms can be used to formulate the multi-dimensional integral of the

objective function, and its value can be approximated using either sampling or asymptotic methods.

Herein we use as the objective function the *Conditional Mutual Information*, defined as the mutual information between the parameters and the data, conditioned on the design variables. This leads to the decision-theoretic expected utility approach of Lindley (1956) with the KL-divergence between the posterior and prior taken as the utility function, when the design variables are considered deterministic. However, using the *Conditional Mutual Information* as defined in Information Theory (Cover and Thomas, 2006) as the objective, provides the additional generalization that the design variables are treated as random variables as well instead of considering them as deterministic. This allows for an easy transformation of the design variables to a deterministic quantity whenever we wish, but also allows the possibility of exploring the effect of uncertainty in them in the optimal design by assigning a probability density function to them. That approach would yield optimal designs which are robust to small changes of the design variables from their optimal value. Also the theory for handling nuisance parameters in OED is developed. Nuisance parameters are parameters (either model or prediction error parameters) that enter the formulation, are uncertain, but are not of interest and we do not wish to update them through the data.

In general, the objective function can only be estimated by Monte Carlo sampling. Ryan (2003) has proposed a Monte Carlo estimator which uses samples from the prior and likelihood PDFs to approximate the expected information gain. This estimator of the objective function has been used by Huan and Marzouk (2013, 2012) to perform OED with the help of surrogate models and stochastic optimization techniques. Herein, we simplify the Monte Carlo estimator of (Ryan, 2003) by evaluating one term of the estimator analytically. This improvement leads to an increased quality of the new estimator by reducing its variance.

The new Monte Carlo estimator is shown to be consistent with the theoretical result of Sebastiani and Wynn (2000) concerning the maximum entropy principle of the marginal likelihood, where the joint entropy of the data and parameters was used in order to prove it. Here we demonstrate that it is possible to arrive at the same theoretical result using the *Conditional Mutual Information* between the data and parameters. More importantly, we present a Bayesian interpretation of the maximum entropy theorem which is also apparent in the new proposed estimator of the objective function.

The theory of the asymptotic method is also developed and enriched with a new theorem which explains the effect of a Gaussian prior PDF in the optimal design. The asymptotic method assumes either: 1) a very small prediction error magnitude and a Gaussian posterior PDF or 2) a large number of uncorrelated data, to simplify the estimation of the objective function considerably. It is shown that the Gaussian prior variances of the model parameters act as weighting factors in the identification of the parameters with the highest variance. That is, the optimal design favors the identification of the parameters with the highest prior variance. As far as the new formula for the sampling objective function estimator is concerned, the variance reduction achieved is demonstrated by a numerical example in a simple analytic problem. Also, using the same problem we demonstrate the key difference between the sampling and asymptotic approaches in OED. Summarizing, the work presented herein is:

1. Formulation of OED based on the *Conditional Mutual Information* measure.
2. Probabilistic treatment of the design variables, and management of nuisance parameters.
3. New Bayesian evidence-based interpretation of the OED objective function.
4. New improved sampling estimator of the objective function with one term calculated analytically.
5. Asymptotic formulation of OED based on the small prediction error and Gaussian posterior assumptions.
6. Theorem which explains the effect of the Gaussian prior variances on the optimal design in the asymptotic OED formulation.
7. Numerical example demonstrating the theoretical findings.

6.2 Objective function formulation

Let $\underline{\theta}$ denote the uncertain parameters of a model used to describe the behavior of a system, and $\underline{y}$ the data (observables) resulting from an experiment performed on the system under the design described by $\underline{\delta}$. Let $p(\underline{\theta})$ denote the prior PDF of the model parameters which describes our state of knowledge about the parameters before any experiment takes place. The goal of OED is to find which experimental design (described by $\underline{\delta}$) is the one that will provide data which are most informative for some purpose.

In the context of *parameter inference* as our goal, what we want is to select the design $\underline{\delta}$ such as that the data $\underline{y}$ will be most informative for the model parameters $\underline{\theta}$. The *Mutual Information* between two random variables is exactly this. It is a scalar measure of the amount of information that these two random variables share. Another way to state this is: the reduction in the uncertainty of one of the variables after having observed the other. If one variable tells a lot about the other, then this reduction of uncertainty is large, and thus the Mutual Information is large. The Mutual Information between $\underline{\theta}$ and $\underline{y}$ is defined as (Cover and Thomas, 2006):

$$I(\underline{\theta};\underline{y}) = \int_{\Theta}\int_{Y} p(\underline{\theta},\underline{y}) \log \frac{p(\underline{\theta},\underline{y})}{p(\underline{\theta})\, p(\underline{y})}\, d\underline{\theta}\, d\underline{y} \tag{6.1}$$

which is the *Relative Entropy* between the joint PDF of $\underline{\theta}$ and $\underline{y}$ and the independent approximation of it using the product of their marginals. So it measures how much these two PDFs differ, essentially measuring how much dependence (shared information) exists between them through their joint PDF.

6.2.1 Probabilistic design variables

At this point, we are going to condition everything on the design variables $\underline{\delta}$ in order to obtain the *Conditional Mutual Information* of $\underline{\theta}$ and $\underline{y}$ conditioned on $\underline{\delta}$, because this is what we are really interested in maximizing (with respect to $\underline{\delta}$). This is the mathematically equivalent way of asking: "Which design

$\underline{\delta}$ maximizes the information $\underline{y}$ carries about $\underline{\theta}$?". So the way that the design variables $\underline{\delta}$ enter the formulation is through conditioning. Therefore, the quantity $I(\underline{\theta};\underline{y}|\underline{\delta})$ is our objective function and we want to maximize it as a function of $\underline{\delta}$. In Information Theory, the *Conditional Mutual Information* of $\underline{\theta}$ and $\underline{y}$ conditioned on $\underline{\delta}$ is defined as the expected value of the mutual information of $\underline{\theta}$ and $\underline{y}$ given $\underline{\delta}$ (Cover and Thomas, 2006):

$$I(\underline{\theta};\underline{y}|\underline{\delta}) = E_{\underline{\delta}}[I(\underline{\theta};\underline{y})|\underline{\delta}] = \int_{\Delta} p(\underline{\delta}) \int_{\Theta}\int_{Y} p(\underline{\theta},\underline{y}|\underline{\delta}) \log \frac{p(\underline{\theta},\underline{y}|\underline{\delta})}{p(\underline{\theta}|\underline{\delta})\, p(\underline{y}|\underline{\delta})}\, d\underline{\theta}\, d\underline{y}\, d\underline{\delta} \tag{6.2}$$

where $\underline{\delta}$ is treated as a random variable in general since we are in the Information Theoretic probabilistic framework.

Note how this approach to formulate the objective function is general in the sense that the design variables need not be deterministic. In most OED formulations the design variables are treated deterministically, that is, for each value of $\underline{\delta}$ the objective function uses this exact value of $\underline{\delta}$ without taking into account possible deviations from it. While this is accepted in theory, in practical situations the conditions might not allow for the exact optimal design $\underline{\delta}$ to take place in the experiment, and we have to use a "near-optimal" design $\underline{\delta}^{'} = \underline{\delta} + \underline{\gamma}$ instead, for some small perturbation $\underline{\gamma}$ from the optimal design. This could be the case when placing sensors on a structure, where the actual sensor positions might differ slightly from the optimal positions estimated from the theoretical model. Then we wish to have a robust design capable of withstanding small deviations from the true optimal design without causing significant reduction in the objective function. This can be achieved mathematically in the proposed formulation by keeping the design variables as random variables in Equation (6.2) and assigning to them a PDF $p(\underline{\delta})$ describing their uncertainty. One could for example use a Gaussian PDF for each $\underline{\delta}$ with a mean of $\underline{\delta}$ and a specified covariance matrix $\Sigma_{\underline{\delta}}$, $p(\underline{\delta}) = N(\underline{\delta}, \Sigma_{\underline{\delta}})$, quantifying how much the actual design variables can deviate from the value $\underline{\delta}$ for which we are evaluating the objective function. Then Equation (6.2) yields a robust objective function with respect to changes in the design variables.

6.2.2 Deterministic design variables

We can always simplify to deterministic design variables by using the deterministic Dirac PDF for $p(\underline{\delta})$ which is the limit-case of a Gaussian with zero variance. This work proceeds in that way, using deterministic design variables. Therefore $\underline{\delta}$ is the *optimization variable* and its values are left to be selected by the algorithm which maximizes the objective function. Therefore, using the known property of the Dirac delta function when it appears inside integrals, the objective function in Equation (6.2) becomes:

$$U(\underline{\delta}) = I(\underline{\theta};\underline{y}|\underline{\delta}) = \int_{\Theta}\int_{Y} p(\underline{\theta},\underline{y}|\underline{\delta}) \log \frac{p(\underline{\theta},\underline{y}|\underline{\delta})}{p(\underline{\theta}|\underline{\delta})\, p(\underline{y}|\underline{\delta})}\, d\underline{\theta}\, d\underline{y} \tag{6.3}$$

Next we are going to demonstrate some ways in which this objective function can be expanded to take forms which include other known quantities, such as *KL Divergences* and *Entropies*. Before we examine those ways, we make the reasonable assumption that the prior PDF of the model parameters $p(\underline{\theta})$ does not depend on the design variables $\underline{\delta}$, that is to say, $p(\underline{\theta}|\underline{\delta}) = p(\underline{\theta})$.

By simply expanding the joint PDF as: $p(\underline{\theta}, \underline{y}|\underline{\delta}) = p(\underline{y}|\underline{\delta})\, p(\underline{\theta}|\underline{y}, \underline{\delta})$ and substituting into Equation (6.3) we obtain:

$$\begin{aligned} U(\underline{\delta}) = I(\underline{\theta}; \underline{y}|\underline{\delta}) &= \int_{\Theta}\int_{Y} p(\underline{y}|\underline{\delta})\, p(\underline{\theta}|\underline{y}, \underline{\delta}) \log \frac{p(\underline{y}|\underline{\delta})\, p(\underline{\theta}|\underline{y}, \underline{\delta})}{p(\underline{\theta}|\underline{\delta})\, p(\underline{y}|\underline{\delta})}\, d\underline{\theta}\, d\underline{y} \\ &= \int_{Y} p(\underline{y}|\underline{\delta}) \int_{\Theta} p(\underline{\theta}|\underline{y}, \underline{\delta}) \log \frac{p(\underline{\theta}|\underline{y}, \underline{\delta})}{p(\underline{\theta})}\, d\underline{\theta}\, d\underline{y} \\ &= E_{\underline{y}|\underline{\delta}}[D_{KL}(p(\underline{\theta}|\underline{y}, \underline{\delta}) \,||\, p(\underline{\theta}))] \end{aligned} \tag{6.4}$$

Equation (6.4) shows that the *Conditional Mutual Information* between $\underline{\theta}$ and $\underline{y}$ conditioned on $\underline{\delta}$ can also be expressed as the *Expected KL Divergence (Information Gain)* from the posterior to the prior where the expectation has to be taken over all possible outcomes (data) of the particular experiment with design $\underline{\delta}$. The posterior $p(\underline{\theta}|\underline{y}, \underline{\delta})$ depends on the data $\underline{y}|\underline{\delta}$ so the expectation w.r.t $\underline{y}|\underline{\delta}$ gives an average of the information gain about the parameters $\underline{\theta}$, based on all data that can result from experiment $\underline{\delta}$ as described by the likelihood function. This is in fact the expected utility objective function proposed by Lindley (1956) under a decision-theoretic framework, when the KL Divergence from posterior to prior is used as the utility function. This was also pointed out by Huan and Marzouk (2013, 2012). However, note how this result was obtained here as a special case of the presented formulation, only when the design variables are deterministic.

6.2.3 Handling nuisance parameters

Oftentimes our models have parameters which are uncertain, but at the same time we are not interested in finding out the values of these parameters through the data. That is, in the Bayesian formulation, we do not want to update the uncertainty in those parameters but we do want to take into account their uncertainty. Since we do not want to update the uncertainty in the nuisance parameters using the data, we should not include them into the posterior distribution. So we distinguish between the parameters that are updated ($\underline{\theta}$) and the nuisance parameters which are not updated ($\underline{\phi}$). The nuisance parameters $\underline{\phi}$ can also include other parameters entering the formulation such as prediction error parameters. By also including the nuisance parameters $\underline{\phi}$ in the objective function, Equation (6.4) becomes a function of $\underline{\delta}$ and $\underline{\phi}$

$$U'(\underline{\delta}, \underline{\phi}) = E_{\underline{y}|\underline{\delta},\underline{\phi}}[D_{KL}(p(\underline{\theta}|\underline{y}, \underline{\delta}, \underline{\phi}) \,||\, p(\underline{\theta}))] \tag{6.5}$$

where we assumed that the prior PDF for the $\underline{\theta}$ parameters is not affected by the nuisance parameters $\underline{\phi}$.

Then by taking the expectation over $\underline{\phi}$ the objective function taking into account the uncertainty in the nuisance parameters is:

$$U(\underline{\delta}) = E_{\underline{\phi}}[U'(\underline{\delta}, \underline{\phi})] = \int_{\Phi} E_{\underline{y}|\underline{\delta},\underline{\phi}}[D_{KL}(p(\underline{\theta}|\underline{y}, \underline{\delta}, \underline{\phi}) \,||\, p(\underline{\theta}))]\, p(\underline{\phi})\, d\underline{\phi} \tag{6.6}$$

where $p(\underline{\phi})$ is the probability distribution which describes the uncertainty in the nuisance parameters $\underline{\phi}$. The integral over Φ can be performed numerically using Sparse grid methods (Bungartz and Griebel,

2004; Gerstner and Griebel, 1998) because in most cases $p(\underline{\phi})$ will be a simple PDF such as Gaussian or Uniform and the Sparse Grid points and weights are readily available for these distributions for any number of dimensions of $\underline{\phi}$. Therefore, the inclusion of nuisance parameters into the objective function is done through averaging over them according to the specified PDF which describes their uncertainty. Note that this uncertainty does not get updated since $\underline{\phi}$ is not included in the posterior. Next we assume for simplicity that nuisance parameters do not exist.

6.3 Sampling approximation of the objective function

By applying Bayes' rule and substituting the posterior PDF in Equation (6.4) we can derive a form which includes the likelihood and prior functions, and is suitable for approximation with Monte Carlo sampling:

$$\begin{aligned} U(\underline{\delta}) = I(\underline{\theta}; \underline{y}|\underline{\delta}) &= \int_Y p(\underline{y}|\underline{\delta}) \int_\Theta p(\underline{\theta}|\underline{y}, \underline{\delta}) \log \frac{p(\underline{\theta}|\underline{y}, \underline{\delta})}{p(\underline{\theta})} \, d\underline{\theta} \, d\underline{y} \\ &= \int_Y p(\underline{y}|\underline{\delta}) \int_\Theta \frac{p(\underline{y}|\underline{\theta}, \underline{\delta}) \, p(\underline{\theta})}{p(\underline{y}|\underline{\delta})} \log \frac{p(\underline{y}|\underline{\theta}, \underline{\delta}) \, p(\underline{\theta})}{p(\underline{y}|\underline{\delta}) \, p(\underline{\theta})} \, d\underline{\theta} \, d\underline{y} \\ &= \int_Y \int_\Theta p(\underline{y}|\underline{\theta}, \underline{\delta}) \, p(\underline{\theta}) \log \frac{p(\underline{y}|\underline{\theta}, \underline{\delta})}{p(\underline{y}|\underline{\delta})} \, d\underline{\theta} \, d\underline{y} \\ &= \int_Y \int_\Theta p(\underline{\theta}, \underline{y}|\underline{\delta}) \left[\log p(\underline{y}|\underline{\theta}, \underline{\delta}) - \log p(\underline{y}|\underline{\delta})\right] d\underline{\theta} \, d\underline{y} \end{aligned} \tag{6.7}$$

where samples $(\underline{\theta}^i, \underline{y}^i)|\underline{\delta}$ from the joint PDF $p(\underline{\theta}, \underline{y}|\underline{\delta}) = p(\underline{\theta}) \, p(\underline{y}|\underline{\theta}, \underline{\delta})$ can be used to approximate the integral in Equation (6.7) with a Monte Carlo sum. The evidence $p(\underline{y}|\underline{\delta})$ is itself approximated with another (nested) Monte Carlo sum from its known definition: $p(\underline{y}|\underline{\delta}) = \int_\Theta p(\underline{y}, \underline{\theta}|\underline{\delta}) \, d\underline{\theta} = \int_\Theta p(\underline{y}|\underline{\theta}, \underline{\delta}) \, p(\underline{\theta}) \, d\underline{\theta}$ and therefore the Monte Carlo estimator of the objective function is:

$$U(\underline{\delta}) \approx \frac{1}{N} \sum_{i=1}^{N} \left\{ \log p(\underline{y}^i|\underline{\theta}^i, \underline{\delta}) - \log \left\{ \frac{1}{M} \sum_{j=1}^{M} p(\underline{y}^i|\underline{\theta}^{i,j}, \underline{\delta}) \right\} \right\} \tag{6.8}$$

Equation (6.8) is in fact the estimator proposed by Ryan (2003) which was also used by Huan and Marzouk (2013, 2012). The advantage of the sampling approach is that it solves the exact problem, without making any assumption about the posterior PDF or the form of the integral in Equation (6.7). Also, as a Monte Carlo estimate, it is guaranteed to be more and more accurate as the number of samples increases. The downside is that for (N, M) samples, for the outer and inner sum respectively, there are $N + NM$ likelihood function evaluations involved for different values of the model parameters. Every likelihood function evaluation, whether it is to draw a data sample or to calculate the density at a data value, involves a model run for the specific parameters. So this results in $\mathbf{N + NM}$ **model runs**. This creates an enormous computational burden especially when the model is already computationally expensive such as a finite element model. Also, we must keep a large enough number of inner and outer samples (N, M) in order to reduce the bias and variance of the estimator (Ryan, 2003; Huan and Marzouk, 2013, 2012).

6.3.1 Objective function simplification

In this section, further theoretical steps are taken which reduce the complexity of the objective function of Equation (6.7) and hence improve the quality of the Monte Carlo estimator of Equation (6.8). Also they provide more insight into the nature of the objective function. Instead of using Equation (6.7) as it is to introduce Monte Carlo sampling, we develop it even further as:

$$U(\underline{\delta}) = \int_Y \int_\Theta p(\underline{y}|\underline{\theta},\underline{\delta})\, p(\underline{\theta}) \log p(\underline{y}|\underline{\theta},\underline{\delta})\, d\underline{\theta}\, d\underline{y} - \int_Y \int_\Theta p(\underline{y}|\underline{\theta},\underline{\delta})\, p(\underline{\theta}) \log p(\underline{y}|\underline{\delta})\, d\underline{\theta}\, d\underline{y} \tag{6.9}$$

and note that the first term can be written as:

$$\begin{aligned}
&\int_Y \int_\Theta p(\underline{y}|\underline{\theta},\underline{\delta})\, p(\underline{\theta}) \log p(\underline{y}|\underline{\theta},\underline{\delta})\, d\underline{\theta}\, d\underline{y} \\
&= \int_\Theta p(\underline{\theta}) \left\{ \int_Y p(\underline{y}|\underline{\theta},\underline{\delta}) \log p(\underline{y}|\underline{\theta},\underline{\delta})\, d\underline{y} \right\} d\underline{\theta} \\
&= -\int_\Theta p(\underline{\theta}) \left\{ -\int_Y p(\underline{y}|\underline{\theta},\underline{\delta}) \log p(\underline{y}|\underline{\theta},\underline{\delta})\, d\underline{y} \right\} d\underline{\theta} \\
&= -\int_\Theta p(\underline{\theta}) [H(p(\underline{y}|\underline{\theta},\underline{\delta}))]\, d\underline{\theta} \\
&= -E_{p(\underline{\theta})}[H(p(\underline{y}|\underline{\theta},\underline{\delta}))]
\end{aligned} \tag{6.10}$$

where $H(p(\underline{x})) = -\int_X p(\underline{x}) \log p(\underline{x})\, d\underline{x}$ denotes the entropy of a PDF $p(\underline{x})$, and therefore the first term of the objective function is simply the minus of the expectation of the likelihood entropy over the prior.

At this point we should remind that the likelihood function is built using the prediction error equation for the data:

$$\underline{y} = \underline{g}(\underline{\theta},\underline{\delta}) + \underline{e} \tag{6.11}$$

where $\underline{e}$ is the prediction error term, which is almost always modeled as a Gaussian random vector (due to the Principle of Maximum Entropy (Cover and Thomas, 2006)) with zero mean and a covariance matrix which generally can depend on both the model parameters $\underline{\theta}$ and design variables $\underline{\delta}$:

$$\underline{e} \sim N(\underline{0}, \Sigma_{\underline{e}}(\underline{\theta},\underline{\delta})) \tag{6.12}$$

which means that the likelihood function (when viewed as a function of the data $\underline{y}$ as in Equation (6.10)) $p(\underline{y}|\underline{\theta},\underline{\delta})$ is also Gaussian:

$$p(\underline{y}|\underline{\theta},\underline{\delta}) = N(\underline{g}(\underline{\theta},\underline{\delta}), \Sigma_{\underline{e}}(\underline{\theta},\underline{\delta})) \tag{6.13}$$

which in turn means that its entropy is known and depends only on the covariance matrix of the prediction error vector:

$$H(p(\underline{y}|\underline{\theta},\underline{\delta})) = \frac{N_s}{2}[\log(2\pi) + 1] + \frac{1}{2}\log|\Sigma_{\underline{e}}(\underline{\theta},\underline{\delta})| \tag{6.14}$$

where N_s is the dimension of the prediction error vector $\underline{e}$ which is equal to the dimension of the model prediction vector $\underline{g}(\underline{\theta},\underline{\delta})$, equal to the number of sensors.

Therefore, the term inside the expectation of Equation (6.10) is simply the likelihood entropy which is always known since the likelihood function is defined by the user. So the integral over $\underline{y}$ of the first term of the objective function can always be done analytically no matter what likelihood function we use. In most of the cases the likelihood is Gaussian and the entropy is given by Equation (6.14), but even if not Gaussian, the entropy is still known. So what we achieved with this was effectively to simplify the double integral of the first term down to a single integral over $\underline{\theta}$. This obviously improves the quality of the Monte Carlo estimator since now we do not have to use both data samples $\underline{y}^i$ and parameter samples $\underline{\theta}^i$ for the approximation of the first term, but only $\underline{\theta}^i$ samples from the prior $p(\underline{\theta})$.

Further simplification can be achieved when we use a likelihood function that does not depend on the parameters $\underline{\theta}$. In OED usually the parameters $\underline{\theta}$ only include the model parameters and the analysis is done (or even repeated) for several fixed magnitudes of the prediction error to explore its effect in the optimal design since it is not known in advance and we would like to know its effect for various magnitudes. This makes sense because OED is performed with the goal of finding the optimal design for learning the model parameters and not the prediction error. Assuming this is the case, the covariance matrix of the prediction error vector does not depend on $\underline{\theta}$ and the likelihood entropy in Equation (6.14) becomes independent of $\underline{\theta}$:

$$H(p(\underline{y}|\underline{\theta},\underline{\delta})) = \frac{N_s}{2}[\log(2\pi)+1] + \frac{1}{2}\log|\Sigma_{\underline{e}}(\underline{\delta})| \equiv W(\underline{\delta}) \tag{6.15}$$

where $W(\underline{\delta})$ denotes the likelihood entropy that depends only on the design variables $\underline{\delta}$.

Therefore the expectation of Equation (6.10) w.r.t. the prior parameters $\underline{\theta}$ vanishes and the objective function of Equation (6.9) becomes:

$$U(\underline{\delta}) = -W(\underline{\delta}) - \int_Y \int_\Theta p(\underline{y}|\underline{\theta},\underline{\delta})\, p(\underline{\theta}) \log p(\underline{y}|\underline{\delta})\, d\underline{\theta}\, d\underline{y} \tag{6.16}$$

If the likelihood does depend on $\underline{\theta}$ then the expectation of Equation (6.10) needs to be performed, but this is still a simple integral only over $\underline{\theta}$ that numerically requires just evaluations of the known likelihood entropy for multiple $\underline{\theta}$ values. It is still simpler than the previous form of the double integral over $\underline{y}$ and $\underline{\theta}$.

Now we deal with the second term which can be written as:

$$\begin{aligned}
&-\int_Y \int_\Theta p(\underline{y}|\underline{\theta},\underline{\delta})\, p(\underline{\theta}) \log p(\underline{y}|\underline{\delta})\, d\underline{\theta}\, d\underline{y} \\
&= -\int_Y \log p(\underline{y}|\underline{\delta}) \left\{ \int_\Theta p(\underline{y}|\underline{\theta},\underline{\delta})\, p(\underline{\theta})\, d\underline{\theta} \right\} d\underline{y} \\
&= -\int_Y \log p(\underline{y}|\underline{\delta}) \left\{ \int_\Theta p(\underline{y},\underline{\theta}|\underline{\delta})\, d\underline{\theta} \right\} d\underline{y} \\
&= -\int_Y \log p(\underline{y}|\underline{\delta})\, p(\underline{y}|\underline{\delta})\, d\underline{y} \\
&= H(p(\underline{y}|\underline{\delta}))
\end{aligned} \tag{6.17}$$

which is the entropy of the data for given design variables $\underline{\delta}$. We should remind that unlike the Bayesian parameter estimation methodology where the data are already known (i.e. from experiment or simulated

data) here in OED the data are not known but treated as a random variable. Therefore, the evidence $p(\underline{y}|\underline{\delta})$ which is a constant scalar value in Bayesian parameter estimation, is a probability density function in OED which has its own entropy. So the objective function takes the form:

$$U(\underline{\delta}) = H(p(\underline{y}|\underline{\delta})) - W(\underline{\delta}) \tag{6.18}$$

where $W(\underline{\delta})$ is the likelihood entropy found from Equation (6.15).

This form of the objective function in Equation (6.18) as the difference between two entropies (evidence and likelihood) is similar to the one obtained later in Equation (6.27), which is also a difference between two entropies (prior and posterior). These forms are equivalent but use different quantities. The form of Equation (6.18) is more difficult to understand since it involves the entropy of the evidence $p(\underline{y}|\underline{\delta})$. Assuming that the likelihood entropy does not depend functionally on $\underline{\delta}$, the objective function is maximized w.r.t. $\underline{\delta}$ when the entropy of the evidence is maximized. Sebastiani and Wynn (2000) arrived at the same theoretical result but using the joint entropy of the parameters and data. However, an intuitive Bayesian explanation of this theoretical result is presented, which is demonstrated after the derivation of the final Monte Carlo estimator of the objective function for clarity purposes.

6.3.2 Simplified Monte Carlo estimator

In order to use the new form of the objective function with the simplified first term we still have to introduce Monte Carlo sampling and obtain the estimator:

$$U(\underline{\delta}) \approx -W(\underline{\delta}) - \frac{1}{N}\sum_{i=1}^{N} \log\, p(\underline{y}^i|\underline{\delta}) \tag{6.19}$$

where N random samples $(\underline{\theta}^i,\, \underline{y}^i)|\underline{\delta}$ from the joint PDF $p(\underline{\theta}, \underline{y}|\underline{\delta}) = p(\underline{\theta})\, p(\underline{y}|\underline{\theta}, \underline{\delta})$ were used to approximate the integral with a Monte Carlo sum. The $\underline{\theta}^i$ samples from the prior PDF $p(\underline{\theta})$ are easily drawn since the prior PDF is a well known distribution such as a Uniform or Gaussian PDF, and then the $\underline{y}^i$ data samples are drawn from the corresponding $\underline{\theta}^i$ samples by simulating a draw from the likelihood function which is known. Then the evidence $p(\underline{y}^i|\underline{\delta}) = \int_\Theta p(\underline{y}^i, \underline{\theta}|\underline{\delta})\, d\underline{\theta} = \int_\Theta p(\underline{y}^i|\underline{\theta}, \underline{\delta})\, p(\underline{\theta})\, d\underline{\theta}$ is approximated using another sum with M samples from $p(\underline{\theta})$:

$$U(\underline{\delta}) \approx -W(\underline{\delta}) - \frac{1}{N}\sum_{i=1}^{N} \log \left\{ \frac{1}{M}\sum_{j=1}^{M} p(\underline{y}^i|\underline{\theta}^{i,j}, \underline{\delta}) \right\} \tag{6.20}$$

where the $\underline{\theta}^j$ samples can again be drawn randomly from $p(\underline{\theta})$. However, it is highly recommended for those $\underline{\theta}^j$ samples to not be drawn again but rather be the same with the $\underline{\theta}^i$ samples drawn for the outer sum, because of the reduced number of required model evaluations from $N + NM$ to just N, but also because of the extremely small values of the evidence that may arise if different samples are used, especially for small prediction error and high dimensions of the model predictions. That would require

far too many M samples to help the evidence estimation which would increase the computational effort dramatically. So the estimator of the objective function finally is:

$$U(\underline{\delta}) \approx -W(\underline{\delta}) - \frac{1}{N}\sum_{i=1}^{N} \log \left\{ \frac{1}{N}\sum_{j=1}^{N} p(\underline{y}^i|\underline{\theta}^j, \underline{\delta}) \right\} \tag{6.21}$$

where the $\underline{\theta}^j$ samples are the same $\underline{\theta}^i$ samples drawn for the approximation of the outer integral in Equation (6.19) and j is just a dummy index.

Equation (6.21) constitutes an improved version of Equation (6.8) because one term less is approximated with Monte Carlo, which means that there is less room for error in the estimate. This was due to the analytic steps taken in Equations (6.9),(6.10),(6.15),(6.16) to calculate one of the terms of the objective function exactly, therefore eliminating the need to approximate that term with Monte Carlo. Also thanks to Equation (6.17) we now know that the second term (the one approximated with sampling) is the entropy of the data for the given design $\underline{\delta}$, $H(p(\underline{y}|\underline{\delta}))$. For Gaussian likelihoods that depend on the parameters $\underline{\theta}$, the only change would be that $|\Sigma_{\underline{e}}(\underline{\delta})|$ should be replaced by $\frac{1}{N}\sum_{i=1}^{N} |\Sigma_{\underline{e}}(\underline{\delta}, \underline{\theta}^i)|$.

Equation (6.21) can be developed further as:

$$\begin{aligned}
U(\underline{\delta}) &\approx -W(\underline{\delta}) - \frac{1}{N}\sum_{i=1}^{N} \left\{ -\log N + \log \sum_{j=1}^{N} p(\underline{y}^i|\underline{\theta}^j, \underline{\delta}) \right\} \\
&= -W(\underline{\delta}) - \frac{1}{N} \left\{ -N\log N + \sum_{i=1}^{N} \log \sum_{j=1}^{N} p(\underline{y}^i|\underline{\theta}^j, \underline{\delta}) \right\} \\
&= -W(\underline{\delta}) + \log N - \frac{1}{N}\sum_{i=1}^{N} \log \sum_{j=1}^{N} p(\underline{y}^i|\underline{\theta}^j, \underline{\delta}) \\
&= \log N - W(\underline{\delta}) - \frac{1}{N} \log \prod_{i=1}^{N} \sum_{j=1}^{N} p(\underline{y}^i|\underline{\theta}^j, \underline{\delta})
\end{aligned} \tag{6.22}$$

Note that the quantity in the logarithm of the right term is the product of the evidences $p(\underline{y}^i|\underline{\delta})$ for all $\underline{y}^i$ data samples, for the given design variables $\underline{\delta}$. Let that be denoted by:

$$P(\underline{\delta};\, \underline{y}^s, \underline{\theta}^s) = \prod_{i=1}^{N} \sum_{j=1}^{N} p(\underline{y}^i|\underline{\theta}^j, \underline{\delta}) \tag{6.23}$$

where $\underline{y}^s$ and $\underline{\theta}^s$ denote the full set of N samples from $p(\underline{\theta}, \underline{y}|\underline{\delta}) = p(\underline{\theta})\, p(\underline{y}|\underline{\theta}, \underline{\delta})$.

Then the objective function estimator becomes:

$$U(\underline{\delta}) \approx \log N - W(\underline{\delta}) - \frac{1}{N} \log\, P(\underline{\delta};\, \underline{y}^s, \underline{\theta}^s) \tag{6.24}$$

Without loss of generality we can exponentiate the objective function to remove the many logarithms of the right hand side. This poses no issue with the optimization because our goal is to maximize $U(\underline{\delta})$ and since the exponential is a monotonic function the maximum of $\exp[U(\underline{\delta})]$ occurs in the same $\underline{\delta}$ with

the maximum of $U(\underline{\delta})$. If we are also interested in the actual value of the objective function except the location of the maximum we can always find it by taking the logarithm.

$$\begin{aligned} \exp[U(\underline{\delta})] &\approx N \left\{(2\pi e_b)^{N_s} |\Sigma_{\underline{e}}(\underline{\delta})|\right\}^{-1/2} P^{-1/N} \Rightarrow \\ \exp[U(\underline{\delta})] &\approx \frac{N}{(2\pi e_b)^{N_s/2} \ |\Sigma_{\underline{e}}(\underline{\delta})|^{1/2} \ P^{1/N}} \end{aligned} \tag{6.25}$$

where $P = P(\underline{\delta}; \underline{y}^s, \underline{\theta}^s)$ was defined in Equation (6.23) and e_b is the base of the natural logarithm.

Further simplification could be achieved by assuming that the prediction error covariance matrix is diagonal of the form $\Sigma_{\underline{e}} = \sigma^2 I$ where I is the identity matrix, which would calculate its determinant analytically. However, this is not done in order to allow for the more interesting general case of the correlated prediction error matrix which means a non-diagonal $\Sigma_{\underline{e}}$ matrix.

Equation (6.25) is an improvement of Equation (6.8) as an estimator of the objective function for the following reasons:

1. It uses Monte Carlo sampling for one term less so it is expected to be more accurate in general.
2. It is useful for gaining physical insight into the nature of the objective function and Optimal Experimental Design.

6.3.3 Intuitive Bayesian interpretation of the objective function

Next we examine Equation (6.25) to gain some physical insight into the problem of model-based Optimal Experimental Design. We seek to maximize the objective function $U(\underline{\delta})$. Under the assumption that the likelihood entropy does not depend on the design variables $\underline{\delta}$, this is equivalent to minimizing $P = P(\underline{\delta}; \underline{y}^s, \underline{\theta}^s)$. We now proceed to show that this result is consistent with the theoretical result of Sebastiani and Wynn (2000) stated earlier.

From the definition of P in Equation (6.23) we can see that since P is the product of evidences of all data samples $\underline{y}^i$, it is numerically minimized w.r.t. $\underline{\delta}$ when all data samples $\underline{y}^i$ attain their lowest evidence value. This might seem counter-intuitive at first, since we know from Bayesian parameter estimation (model updating) that a high evidence is a sign of a good model and something we would welcome. But here we should note an important difference with OED. In contrast to model updating where the data are known and fixed, in OED they are unknown and treated as a random variable. This means that many different data values should be accounted for. The optimal design is the one which gives good parameter estimates for *all* data values that may arise, and doesn't favor some data over others. This is achieved by keeping the distribution of the evidence as spread as possible over the data space, which effectively lowers the values of the distribution at the individual data points. We need to keep in mind that since $p(\underline{y}|\underline{\delta})$ is a probability distribution, low values of $p(\underline{y}^i, \underline{\delta})$ simply imply that the distribution is as spread as possible over the data space. This is not to say that we prefer low evidences to high evidences. To make this clear, lets assume for simplicity that the data $\underline{y}^i$ for 3 different designs follow normal distributions (1d data) with different standard deviations:

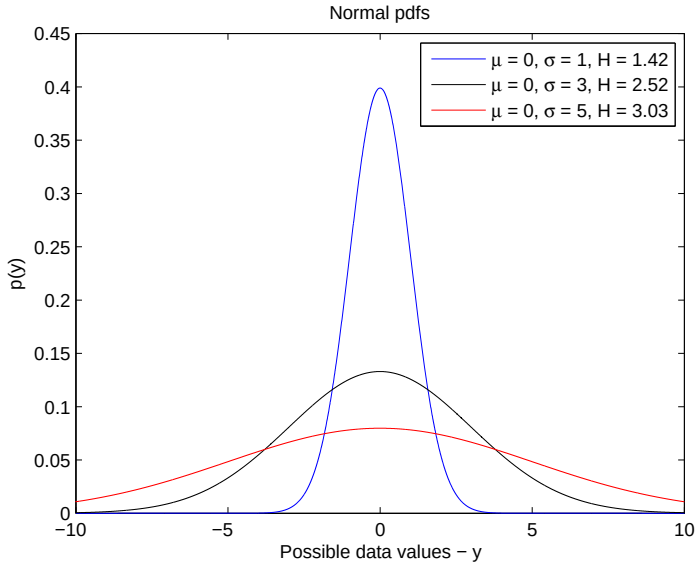


Figure 6.1: Evidence densities for different designs.

We can see that even though the first design (blue) gives a higher evidence at the central data values, it fails for the others. The situation gets better in the second design (black) and is best in the third design (red) where an adequate fit is preserved for all possible data values. Also note that the entropies of the distributions are increasing. This is in agreement with Equation (6.18) and (Sebastiani and Wynn, 2000), which suggested that the optimal design is the one which maximizes the entropy of the evidence distribution.

So even though a small evidence value might sound bad at first for a specific data point, from the OED perspective it is more important to be able to keep a constant fit with all the data despite their uncertainty, than to have a good fit with some data that might arise and a bad fit with all the rest. By minimizing the individual evidences of all data samples we are increasing the entropy of the evidence distribution, making it more spread out in the data space, and therefore ensuring that it is more like the red one than the blue one in Figure 6.1. Therefore, we can expect to have a good fit with the data *on average*, despite their uncertainty. This constitutes a novel "flat-evidence" interpretation of OED.

6.4 Asymptotic approximation of the objective function

From the properties of the *Conditional Mutual Information* we know that it can also be expressed as (Cover and Thomas, 2006):

$$U(\underline{\delta}) = I(\underline{\theta}; \underline{y}|\underline{\delta}) = H(p(\underline{\theta}|\underline{\delta})) - H(p(\underline{\theta}|\underline{y}, \underline{\delta})) \tag{6.26}$$

where the entropy of the prior PDF conditioned on $\underline{\delta}$, $H(p(\underline{\theta}|\underline{\delta}))$, is independent of $\underline{\delta}$ because $p(\underline{\theta}|\underline{\delta}) = p(\underline{\theta})$ and thus $H(p(\underline{\theta}|\underline{\delta})) = H(p(\underline{\theta})) = constant = c$ as far as the optimization w.r.t $\underline{\delta}$ is concerned. So we have:

$$U(\underline{\delta}) = I(\underline{\theta}; \underline{y}|\underline{\delta}) = H(p(\underline{\theta})) - H(p(\underline{\theta}|\underline{y}, \underline{\delta})) = c - H(p(\underline{\theta}|\underline{y}, \underline{\delta})) \tag{6.27}$$

Equation (6.27) can be interpreted as the reduction in uncertainty of $\underline{\theta}$ due to the knowledge of $\underline{y}|\underline{\delta}$. This added knowledge that came from $\underline{y}|\underline{\delta}$ about $\underline{\theta}$ is their mutual information. Equation (6.27) is easily proved by continuing from Equation (6.4) - the expected KL Divergence form - and breaking down the logarithm of the posterior/prior fraction into a difference, instead of applying Bayes' rule as was done earlier. Specifically:

$$\begin{aligned}
U(\underline{\delta}) = I(\underline{\theta}; \underline{y}|\underline{\delta}) &= \int_Y p(\underline{y}|\underline{\delta}) \int_\Theta p(\underline{\theta}|\underline{y}, \underline{\delta}) \log \frac{p(\underline{\theta}|\underline{y}, \underline{\delta})}{p(\underline{\theta})} \, d\underline{\theta} \, d\underline{y} \\
&= \int_Y \int_\Theta p(\underline{\theta}, \underline{y}|\underline{\delta}) \log p(\underline{\theta}|\underline{y}, \underline{\delta}) \, d\underline{\theta} \, d\underline{y} - \int_Y \int_\Theta p(\underline{\theta}, \underline{y}|\underline{\delta}) \log p(\underline{\theta}) \, d\underline{\theta} \, d\underline{y} \\
&= -H(p(\underline{\theta}|\underline{y}, \underline{\delta})) - \int_\Theta \int_Y p(\underline{\theta}, \underline{y}|\underline{\delta}) \, d\underline{y} \, \log p(\underline{\theta}) \, d\underline{\theta} \\
&= -H(p(\underline{\theta}|\underline{y}, \underline{\delta})) - \int_\Theta p(\underline{\theta}|\underline{\delta}) \log p(\underline{\theta}) \, d\underline{\theta} \\
&= H(p(\underline{\theta})) - H(p(\underline{\theta}|\underline{y}, \underline{\delta})) \\
&= c - H(p(\underline{\theta}|\underline{y}, \underline{\delta}))
\end{aligned} \tag{6.28}$$

which shows that what is actually being optimized (minimized) w.r.t $\underline{\delta}$ is the *Conditional Entropy of* $\underline{\theta}$ *conditioned on* $\underline{y}$ *and* $\underline{\delta}$, which is by definition the expectation w.r.t $\underline{y}|\underline{\delta}$ of the entropy of the posterior for some specific value of the data:

$$\begin{aligned}
H(p(\underline{\theta}|\underline{y}, \underline{\delta})) &= -\int_Y \int_\Theta p(\underline{\theta}, \underline{y}|\underline{\delta}) \log p(\underline{\theta}|\underline{y}, \underline{\delta}) \, d\underline{\theta} \, d\underline{y} \\
&= -\int_Y p(\underline{y}|\underline{\delta}) \int_\Theta p(\underline{\theta}|\underline{y}, \underline{\delta}) \log p(\underline{\theta}|\underline{y}, \underline{\delta}) \, d\underline{\theta} \, d\underline{y} \\
&= E_{\underline{y}|\underline{\delta}} \left\{ H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta})) \right\}
\end{aligned} \tag{6.29}$$

This form of the objective function is the simplest to understand since it breaks down to two terms one of which does not affect the optimization w.r.t $\underline{\delta}$ because of the independence of the prior on $\underline{\delta}$. This, however, is not to say that the prior does not play a role in the objective. It still affects the results through the posterior from Bayes' theorem. This form is suitable for asymptotic approximations because it directly involves the posterior PDF, in contrast with the previous form which involved the likelihood and prior and hence were convenient for sampling approaches. The direct involvement of the posterior offers the possibility of approximating it with a Gaussian PDF whose entropy has a known form that depends only on the covariance matrix of the Gaussian.

In the asymptotic method first the posterior is approximated with a Gaussian PDF (which requires first and second derivatives of the model prediction w.r.t. parameters $\underline{\theta}$) so that its entropy has a known

form, and then the integral over the data $\underline{y}$ is simplified under the assumption of either: 1) very small prediction error magnitude or 2) large number of uncorrelated data. Note that when the assumption 2) is made, this automatically implies the validity of the Gaussian posterior assumption due to the Bayesian Central Limit Theorem. By introducing the Gaussian posterior assumption we have:

$$
\begin{aligned}
H(p(\underline{\theta}|\underline{y},\underline{\delta})) &= -\int_Y p(\underline{y}|\underline{\delta}) \int_\Theta p(\underline{\theta}|\underline{y},\underline{\delta}) \log p(\underline{\theta}|\underline{y},\underline{\delta})\, d\underline{\theta}\, d\underline{y} \\
&= \int_Y p(\underline{y}|\underline{\delta})\, H(p(\underline{\theta}|\underline{Y}=\underline{y},\underline{\delta}))\, d\underline{y} \\
&= \int_Y \left\{ \int_\Theta p(\underline{y}|\underline{\theta},\underline{\delta})\, p(\underline{\theta})\, d\underline{\theta} \right\} H(p(\underline{\theta}|\underline{Y}=\underline{y},\underline{\delta}))\, d\underline{y} \\
&= \int_Y \int_\Theta H(p(\underline{\theta}|\underline{Y}=\underline{y},\underline{\delta}))\, p(\underline{y}|\underline{\theta},\underline{\delta})\, p(\underline{\theta})\, d\underline{\theta}\; d\underline{y} \\
&= \int_\Theta \int_Y H(p(\underline{\theta}|\underline{Y}=\underline{y},\underline{\delta}))\, p(\underline{y}|\underline{\theta},\underline{\delta})\, p(\underline{\theta})\, d\underline{y}\; d\underline{\theta} \\
&\approx \frac{1}{N} \sum_{i=1}^{N} H(p(\underline{\theta}|\underline{Y}=\underline{y}^i,\underline{\delta})) \\
&\approx \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \log((2\pi e_b)^{N_{\underline{\theta}}} |\Sigma_{p(\underline{\theta}|\underline{y}^i,\underline{\delta})}|) \\
&\approx \frac{1}{2N} \sum_{i=1}^{N} \left[N_{\underline{\theta}} (\log(2\pi)+1) + \log|\Sigma_{p(\underline{\theta}|\underline{y}^i,\underline{\delta})}| \right] \\
&\approx \frac{1}{2N} \left[N N_{\underline{\theta}} (\log(2\pi)+1) + \sum_{i=1}^{N} \log|\Sigma_{p(\underline{\theta}|\underline{y}^i,\underline{\delta})}| \right] \\
&\approx \frac{N_{\underline{\theta}}}{2} (\log(2\pi)+1) + \frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1}^{N_{\underline{\theta}}} \log \lambda_i^j
\end{aligned}
\tag{6.30}
$$

where N samples $(\underline{\theta}^i, \underline{y}^i)|\underline{\delta}$ from the joint PDF $p(\underline{\theta},\underline{y}|\underline{\delta}) = p(\underline{\theta})\, p(\underline{y}|\underline{\theta},\underline{\delta})$ were used to approximate the double integral with a Monte Carlo sum, and λ_i^j is the j-th of the $N_{\underline{\theta}}$ eigenvalues of the posterior PDF covariance matrix for the i-th of the N data samples $\underline{y}^i$. The determinant was substituted with the product of the eigenvalues of the covariance matrix, and the logarithm of the product property was used. Note that in the fourth line the posterior entropy for data point $\underline{Y}=\underline{y}$ was moved inside the integral over Θ. This is possible because it does not depend on $\underline{\theta}$, but only on $\underline{y}$ (and $\underline{\delta}$).

Up to this point we have not introduced the small prediction error magnitude or the large number of uncorrelated data assumption, but only the Gaussian posterior assumption. So the objective function is the average (expected) posterior entropy over many data samples. Since the posterior is Gaussian its entropy depends only on the covariance matrix. From the theory of Gaussian approximation of the posterior, for data point $\underline{y}^i$ we know that the covariance matrix of the posterior is the minus inverse of the Hessian w.r.t. model parameters $\underline{\theta}$ (denoted as $\nabla^2_{\underline{\theta}}$) of the log-posterior, evaluated at the most probable value of the parameters $\underline{\theta}_*$ (mean of the Gaussian posterior) which depends on $\underline{y}^i$. The Gaussian

approximation of the posterior is based on the Taylor series expansion of the log-posterior (function of $\underline{\theta}$) around its maximum, and keeping up to the quadratic term with the Hessian (second derivatives) of the log-posterior evaluated at the maximum. The linear term with the first derivatives is zero because the gradient of the log-posterior is zero at the maximum. Then by exponentiating the Taylor series expansion the posterior takes the known Gaussian form. Therefore its precision depends on whether the truncation of the higher-order-derivative terms is important or not. The Taylor series approximation of the log-posterior is done at the point $\underline{\theta}_*$ of maximum posterior probability. By taking the logarithm of Bayes' rule we have the log-posterior:

$$\begin{aligned} \log\, p(\underline{\theta}|\underline{y}^i, \underline{\delta}) &= \log\, p(\underline{y}^i|\underline{\theta}, \underline{\delta}) + \log\, p(\underline{\theta}) - \log\, p(\underline{y}^i|\underline{\delta}) \\ &= \log\, p(\underline{y}^i|\underline{\theta}, \underline{\delta}) + \log\, p(\underline{\theta}) - C_1 \end{aligned} \tag{6.31}$$

where the evidence term $p(\underline{y}^i|\underline{\delta})$ does not depend on $\underline{\theta}$. Then the Gaussian posterior covariance matrix is

$$\begin{aligned} \Sigma_{p(\underline{\theta}|\underline{y}^i, \underline{\delta})} &= [-\nabla^2_{\underline{\theta}}[\log\, p(\underline{\theta}|\underline{y}^i, \underline{\delta})]]^{-1}\,|\,\underline{\theta} = \underline{\theta}_* \\ &= [-\nabla^2_{\underline{\theta}}[\log\, p(\underline{y}^i|\underline{\theta}, \underline{\delta})] - \nabla^2_{\underline{\theta}}(\log\, p(\underline{\theta}))]^{-1}\,|\,\underline{\theta} = \underline{\theta}_* \end{aligned} \tag{6.32}$$

where $\underline{\theta}_*$ maximizes the log-posterior (6.31). Using the prediction error equation to model the discrepancy between the model prediction and measurement

$$\underline{y} = \underline{g}(\underline{\theta}, \underline{\delta}) + \underline{e} \tag{6.33}$$

and assuming that the discrepancy $\underline{e}$ is zero-mean Gaussian with covariance matrix $\Sigma_{\underline{e}}(\underline{\delta})$ the likelihood is also Gaussian with PDF

$$p(\underline{y}^i|\underline{\theta}, \underline{\delta}) = \frac{\exp\left(-\frac{1}{2}\, J(\underline{y}^i, \underline{\theta}, \underline{\delta})\right)}{\sqrt{|2\pi\Sigma_{\underline{e}}(\underline{\delta})|}} \tag{6.34}$$

where $J(\underline{y}^i, \underline{\theta}, \underline{\delta})$ is the quadratic form defined as

$$J(\underline{y}^i, \underline{\theta}, \underline{\delta}) = (\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta}))^T \Sigma_e(\underline{\delta})^{-1} (\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta})) \tag{6.35}$$

then the log-likelihood is

$$\begin{aligned} \log\, p(\underline{y}^i|\underline{\theta}, \underline{\delta}) &= -\frac{1}{2}\, J(\underline{y}^i, \underline{\theta}, \underline{\delta}) - \log\, \left(\sqrt{|2\pi\Sigma_{\underline{e}}(\underline{\delta})|}\right) \\ &= -\frac{1}{2}\, J(\underline{y}^i, \underline{\theta}, \underline{\delta}) - C_2 \end{aligned} \tag{6.36}$$

where C_2 does not depend on $\underline{\theta}$. Then the Hessian of the log-likelihood w.r.t. the model parameters $\underline{\theta}$ (first term of the Gaussian posterior covariance matrix of Equation (6.32)) is

$$\nabla^2_{\underline{\theta}}[\log\, p(\underline{y}^i|\underline{\theta}, \underline{\delta})] = -\frac{1}{2}\nabla^2_{\underline{\theta}}[J(\underline{y}^i, \underline{\theta}, \underline{\delta})]$$

$$= -\frac{1}{2}\nabla^2_{\underline{\theta}}[(\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta}))^T \Sigma_e(\underline{\delta})^{-1} (\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta}))]$$

$$= -D_{\underline{\theta}}[\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta})]^T \Sigma_e(\underline{\delta})^{-1} D_{\underline{\theta}}[\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta})] - \nabla^2_{\underline{\theta}}[\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta})] \Sigma_e(\underline{\delta})^{-1} [\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta})]$$

$$= -D_{\underline{\theta}}[\underline{g}(\underline{\theta},\underline{\delta})]^T \Sigma_e(\underline{\delta})^{-1} D_{\underline{\theta}}[\underline{g}(\underline{\theta},\underline{\delta})] + \nabla^2_{\underline{\theta}}[\underline{g}(\underline{\theta},\underline{\delta})]\Sigma_e(\underline{\delta})^{-1}[\underline{y}^i - \underline{g}(\underline{\theta},\underline{\delta})] \tag{6.37}$$

where $D_{\underline{\theta}}\left[\underline{g}(\underline{\theta},\underline{\delta})\right]$ denotes the $N_s \times N_{\underline{\theta}}$ Jacobian matrix of first partial derivatives of the components of $\underline{g}(\underline{\theta},\underline{\delta})$ w.r.t. $\underline{\theta}$ and $\nabla^2_{\underline{\theta}}\left[\underline{g}(\underline{\theta},\underline{\delta})\right]$ is a row block matrix of Hessian matrices of the components of $\underline{g}(\underline{\theta},\underline{\delta})$, $\nabla^2_{\underline{\theta}}\left[\underline{g}(\underline{\theta},\underline{\delta})\right] = \left[\nabla^2_{\underline{\theta}}\, g_1(\underline{\theta},\underline{\delta}), \ldots, \nabla^2_{\underline{\theta}}\, g_{N_s}(\underline{\theta},\underline{\delta})\right]$.

Note for linear models Note that for linear models w.r.t. the parameters, that is, $\underline{g}(\underline{\theta},\underline{\delta}) = A(\underline{\delta})\underline{\theta} + b(\underline{\delta})$, the Jacobian matrix is independent of $\underline{\theta}$, $D_{\underline{\theta}}\left[\underline{g}(\underline{\theta},\underline{\delta})\right] = A(\underline{\delta})$ and the Hessian is zero $\nabla^2_{\underline{\theta}}\left[\underline{g}(\underline{\theta},\underline{\delta})\right] = 0$. Therefore (6.37) becomes

$$\nabla^2_{\underline{\theta}}\left[\log\, p(\underline{y}^i|\underline{\theta},\underline{\delta})\right] = -A(\underline{\delta})^T \Sigma_e(\underline{\delta})^{-1} A(\underline{\delta}) \tag{6.38}$$

which is independent of $\underline{\theta}$. Also, the Hessian of the log-prior is zero for uniform prior PDF or equal to the inverse of the prior covariance matrix for Gaussian prior PDF (independent of $\underline{\theta}$). Then it is obvious from (6.31) that the Hessian of the log-posterior is independent of $\underline{\theta}$. This means that higher-order derivatives of the log-posterior w.r.t. $\underline{\theta}$ are zero, and therefore the Gaussian approximation is exact for linear models and not an approximation.

The optimal $\underline{\theta}_*$ where the posterior covariance matrix (6.32) is evaluated is the one that maximizes the log-posterior of (6.31). By substituting the log-likelihood from (6.36) to (6.31) and taking the minus, the optimal $\underline{\theta}_*$ equivalently minimizes:

$$\frac{1}{2} J(\underline{y}^i, \underline{\theta}, \underline{\delta}) - \log\, p(\underline{\theta}) + C_3 \tag{6.39}$$

where all the constants that do not depend on $\underline{\theta}$ were merged into C_3. So in order to do the Gaussian approximation we need to solve the minimization problem (6.39) w.r.t. $\underline{\theta}$ to find the optimal $\underline{\theta}_*$ for the given $\underline{y}^i$. We also need to have both the first and second derivatives of the model prediction vector w.r.t. the parameters $\underline{\theta}$. Then the posterior covariance matrix is given by Equations (6.32) and (6.37).

For small prediction error magnitude or a large number of uncorrelated data points, the following simplification occurs in the objective function. Continuing from Equation (6.30):

$$\begin{aligned} U(\underline{\delta}) &= H(p(\underline{\theta}|\underline{y},\underline{\delta})) \\ &= \int_{\Theta} p(\underline{\theta}) \left\{ \int_{Y} H(p(\underline{\theta}|\underline{Y} = \underline{y},\underline{\delta}))\, p(\underline{y}|\underline{\theta},\underline{\delta})\, d\underline{y} \right\}\, d\underline{\theta} \\ &= \int_{\Theta} p(\underline{\theta}) I(\underline{\theta})\, d\underline{\theta} \end{aligned} \tag{6.40}$$

where the integral of the posterior entropy times the likelihood over the data is a function of $\underline{\theta}$ and was substituted with $I(\underline{\theta})$. Next, we show that $I(\underline{\theta})$ can be written in a way such that the Laplace asymptotic approximation of integrals can be applied when either the prediction error magnitude is very small, or when a large number of uncorrelated data is available (Tierney and Kadane, 1986; Bleistein and

Handelsman, 1975).

$$\begin{aligned} I(\underline{\theta}) &= \int_Y H(p(\underline{\theta}|\underline{Y}=\underline{y},\underline{\delta}))\, p(\underline{y}|\underline{\theta},\underline{\delta})\, d\underline{y} \\ &= \int_Y H(p(\underline{\theta}|\underline{Y}=\underline{y},\underline{\delta}))\, \frac{\exp\left(-\frac{1}{2}\,(\underline{y}-\underline{g}(\underline{\theta},\underline{\delta}))^T \Sigma_{\underline{e}}(\underline{\delta})^{-1}(\underline{y}-\underline{g}(\underline{\theta},\underline{\delta}))\right)}{\sqrt{|2\pi\Sigma_{\underline{e}}(\underline{\delta})|}}\, d\underline{y} \end{aligned} \tag{6.41}$$

Without loss of generality, the prediction error covariance matrix can be written as $\Sigma_{\underline{e}}(\underline{\delta}) = \sigma_{\underline{e}}^2 R_{\underline{e}}(\underline{\delta})$, where $\sigma_{\underline{e}}^2$ is the magnitude of the prediction error covariance matrix and $R_{\underline{e}}(\underline{\delta})$ describes its possible correlation structure. Then we set $M = 1/\sigma_{\underline{e}}^2$ and obtain

$$\begin{aligned} I(\underline{\theta}) &= \int_Y H(p(\underline{\theta}|\underline{Y}=\underline{y},\underline{\delta}))\, \frac{\exp\left(M(-\frac{1}{2})\,(\underline{y}-\underline{g}(\underline{\theta},\underline{\delta}))^T R_{\underline{e}}(\underline{\delta})^{-1}(\underline{y}-\underline{g}(\underline{\theta},\underline{\delta}))\right)}{\sqrt{|2\pi\sigma_{\underline{e}}^2 R_{\underline{e}}(\underline{\delta})|}}\, d\underline{y} \\ &= \int_Y H(p(\underline{\theta}|\underline{Y}=\underline{y},\underline{\delta}))\, (2\pi)^{-N_s/2}\, \sigma_{\underline{e}}^{-N_s}\, |R_{\underline{e}}(\underline{\delta})|^{-1/2}\, \exp\left(M\, f(\underline{y})\right)\, d\underline{y} \\ &= \int_Y h(\underline{y})\, \exp\left(M\, f(\underline{y})\right)\, d\underline{y} \end{aligned} \tag{6.42}$$

from which it is immediately obvious that the integral is of Laplace type when M is large, or equivalently when $\sigma_{\underline{e}}$ is very small, with:

$$h(\underline{y}) = H(p(\underline{\theta}|\underline{Y}=\underline{y},\underline{\delta}))\, (2\pi)^{-N_s/2}\, \sigma_{\underline{e}}^{-N_s}\, |R_{\underline{e}}(\underline{\delta})|^{-1/2} \tag{6.43}$$

and

$$f(\underline{y}) = -\frac{1}{2}\,(\underline{y}-\underline{g}(\underline{\theta},\underline{\delta}))^T R_{\underline{e}}(\underline{\delta})^{-1}(\underline{y}-\underline{g}(\underline{\theta},\underline{\delta})) \tag{6.44}$$

In order to apply the Laplace approximation we need the maximum $\underline{y}_0$ of $f(\underline{y})$ and its Hessian matrix evaluated at that point. From the structure of $f(\underline{y})$ as a quadratic form it is obvious that since the correlation matrix $R_{\underline{e}}(\underline{\delta})$ is positive definite, its inverse $R_{\underline{e}}(\underline{\delta})^{-1}$ is also positive definite and $f(\underline{y})$ can only be negative or zero. Therefore $f(\underline{y})$ has a maximum of zero at $\underline{y}_0 = \underline{g}(\underline{\theta},\underline{\delta})$, $f(\underline{g}(\underline{\theta},\underline{\delta})) = 0$. Also, the Hessian matrix of $f(\underline{y})$ w.r.t. $\underline{y}$ is easily found to be $\nabla_{\underline{y}}^2 f(\underline{y}) = -R_{\underline{e}}(\underline{\delta})^{-1}$ which is independent of $\underline{y}$.

Then the Laplace approximation (Tierney and Kadane, 1986; Bleistein and Handelsman, 1975) states that for large M:

$$\int_Y h(\underline{y})\, \exp\left(M\, f(\underline{y})\right)\, d\underline{y} \approx h(\underline{g}(\underline{\theta},\underline{\delta}))\, (2\pi)^{N_s/2}\, M^{-N_s/2}\, |-\nabla_{\underline{y}}^2 f(\underline{g}(\underline{\theta},\underline{\delta}))|^{-1/2}\, \exp\left(M\, f(\underline{g}(\underline{\theta},\underline{\delta}))\right) \tag{6.45}$$

which in our case of $h(\underline{y})$ and $f(\underline{y})$ yields:

$$I(\underline{\theta}) \approx H(p(\underline{\theta}|\underline{Y}=\underline{g}(\underline{\theta},\underline{\delta})))\, (2\pi)^{-N_s/2}\, \sigma_{\underline{e}}^{-N_s}\, |R_{\underline{e}}(\underline{\delta})|^{-1/2}\, (2\pi)^{N_s/2}\, \sigma_{\underline{e}}^{N_s}\, |R_{\underline{e}}(\underline{\delta})|^{1/2}\, \exp(0) \tag{6.46}$$

and after the cancellation of most of the terms the integral approximation simplifies to:

$$I(\underline{\theta}) \approx H(p(\underline{\theta}|\underline{Y}=\underline{g}(\underline{\theta},\underline{\delta}))) \tag{6.47}$$

which is the posterior entropy evaluated for the data point equal to the model prediction for the specific model parameters $\underline{\theta}$. Before substituting the value of the integral $I(\underline{\theta})$ into the objective function of

(6.40), we first demonstrate that $I(\underline{\theta})$ can admit Laplace approximation independently of the magnitude of the prediction error, provided that we have a large number of uncorrelated data. For uncorrelated data the prediction error covariance matrix becomes diagonal. For simplicity and without affecting the proof we assume that the variances of each component of the prediction error vector are the same, that is: $\Sigma_{\underline{e}}(\underline{\delta}) = \sigma_{\underline{e}}^2 \, I_{N_s}$, where I_{N_s} is the $N_s \times N_s$ identity matrix. Then $I(\underline{\theta})$ from (6.41) takes the form:

$$\begin{aligned}
I(\underline{\theta}) &= \int_Y H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta})) \, \frac{\exp\left(-\frac{1}{2} \, (\underline{y} - \underline{g}(\underline{\theta}, \underline{\delta}))^T \Sigma_{\underline{e}}(\underline{\delta})^{-1} (\underline{y} - \underline{g}(\underline{\theta}, \underline{\delta}))\right)}{\sqrt{|2\pi \Sigma_{\underline{e}}(\underline{\delta})|}} \, d\underline{y} \\
&= \int_Y H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta})) \, (2\pi)^{-N_s/2} \, \sigma_{\underline{e}}^{-N_s} \, |I_{N_s}|^{-1/2} \, \exp\left(-\frac{1}{2\sigma_{\underline{e}}^2} \, (\underline{y} - \underline{g}(\underline{\theta}, \underline{\delta}))^T I_{N_s} (\underline{y} - \underline{g}(\underline{\theta}, \underline{\delta}))\right) \, d\underline{y} \\
&= \int_Y H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta})) \, (2\pi)^{-N_s/2} \, \sigma_{\underline{e}}^{-N_s} \, \exp\left(-\frac{1}{2\sigma_{\underline{e}}^2} \, (\underline{y} - \underline{g}(\underline{\theta}, \underline{\delta}))^T \, (\underline{y} - \underline{g}(\underline{\theta}, \underline{\delta}))\right) \, d\underline{y} \\
&= \int_Y H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta})) \, (2\pi)^{-N_s/2} \, \sigma_{\underline{e}}^{-N_s} \, \exp\left(-\frac{1}{2\sigma_{\underline{e}}^2} \, \sum_{i=1}^{N_s} (y_i - g_i(\underline{\theta}, \underline{\delta}))^2\right) \, d\underline{y} \\
&= \int_Y H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta})) \, (2\pi)^{-N_s/2} \, \sigma_{\underline{e}}^{-N_s} \, \exp\left(-\frac{N_s}{2\sigma_{\underline{e}}^2} \, \sum_{i=1}^{N_s} (y_i - g_i(\underline{\theta}, \underline{\delta}))^2 / N_s\right) \, d\underline{y} \\
&= \int_Y h(\underline{y}) \, \exp\left(M \, f(\underline{y})\right) \, d\underline{y}
\end{aligned} \tag{6.48}$$

where

$$h(\underline{y}) = H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta})) \, (2\pi)^{-N_s/2} \, \sigma_{\underline{e}}^{-N_s} \tag{6.49}$$

and

$$f(\underline{y}) = -\frac{1}{2\sigma_{\underline{e}}^2} \, \sum_{i=1}^{N_s} (y_i - g_i(\underline{\theta}, \underline{\delta}))^2 / N_s \tag{6.50}$$

where the sum over the data points is the mean squared difference between the model prediction and the data, which is always a finite and non-negative quantity. In the role of the large number M is now the number of data points, $M = N_s$. Again it is easily seen that the maximum of $f(\underline{y})$ is zero attained at $\underline{y}_0 = \underline{g}(\underline{\theta}, \underline{\delta})$ and the Hessian of $f(\underline{y})$ is $-\sigma_{\underline{e}}^{-2} \, N_s^{-1} \, I_{N_s}$.

Therefore, for large number of data points the Laplace approximation of $I(\underline{\theta})$ is:

$$\begin{aligned}
I(\underline{\theta}) &\approx H(p(\underline{\theta}|\underline{Y} = \underline{g}(\underline{\theta}, \underline{\delta}))) \, (2\pi)^{-N_s/2} \, \sigma_{\underline{e}}^{-N_s} \, (2\pi)^{N_s/2} \, N_s^{-N_s/2} \, |\sigma_{\underline{e}}^{-2} \, N_s^{-1} \, I_{N_s}|^{-1/2} \, \exp\left(N_s \, f(\underline{g}(\underline{\theta}, \underline{\delta}))\right) \\
&\approx H(p(\underline{\theta}|\underline{Y} = \underline{g}(\underline{\theta}, \underline{\delta}))) \, \sigma_{\underline{e}}^{-N_s} \, N_s^{-N_s/2} \, |\sigma_{\underline{e}}^{-2} \, N_s^{-1} \, I_{N_s}|^{-1/2} \, \exp\left(N_s \, f(\underline{g}(\underline{\theta}, \underline{\delta}))\right) \\
&\approx H(p(\underline{\theta}|\underline{Y} = \underline{g}(\underline{\theta}, \underline{\delta}))) \, \sigma_{\underline{e}}^{-N_s} \, N_s^{-N_s/2} \, \sigma_{\underline{e}}^{N_s} \, N_s^{N_s/2} \, \exp(0) \\
&\approx H(p(\underline{\theta}|\underline{Y} = \underline{g}(\underline{\theta}, \underline{\delta})))
\end{aligned} \tag{6.51}$$

which is the same result as the one obtained under the assumption of very small prediction error magnitude in (6.47).

We have proved that the $I(\underline{\theta})$ integral defined in (6.41) admits Laplace asymptotic approximation under two different assumptions:

1. The prediction error magnitude is very small.

2. The number of data points is large and the data are uncorrelated.

The result of the approximation is the same in both cases, and it is equal to the posterior entropy for data equal to the model prediction for the specific model parameters. Then we continue by substituting the approximated value of the integral $I(\underline{\theta})$ into the objective function of (6.40):

$$\begin{aligned} U(\underline{\delta}) &\approx \int_{\Theta} p(\underline{\theta})\, H(p(\underline{\theta}|\underline{Y} = \underline{g}(\underline{\theta}, \underline{\delta}))\, d\underline{\theta} \\ &\approx \sum_{i=1}^{N} w_i\, H(p(\underline{\theta}|\underline{Y} = \underline{g}(\underline{\theta}^i, \underline{\delta})) \\ &\approx \sum_{i=1}^{N} w_i\, \frac{1}{2} \log((2\pi e_b)^{N_{\underline{\theta}}} |\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})}|) \\ &\approx \sum_{i=1}^{N} w_i \frac{1}{2} \left[N_{\underline{\theta}} (\log(2\pi) + 1) + \log |\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})}| \right] \\ &\approx \frac{N_{\underline{\theta}}}{2} (\log(2\pi) + 1) \sum_{i=1}^{N} w_i + \frac{1}{2} \sum_{i=1}^{N} w_i \log |\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})}| \\ &\approx \frac{N_{\underline{\theta}}}{2} (\log(2\pi) + 1) \; + \frac{1}{2} \sum_{i=1}^{N} w_i \left[\sum_{j=1}^{N_{\underline{\theta}}} \log \lambda_i^j \right] \end{aligned} \tag{6.52}$$

where λ_i^j is the j-th of the $N_{\underline{\theta}}$ eigenvalues of the posterior PDF covariance matrix for the i-th of the N data samples $\underline{y}^i$, equal to the i-th model prediction vector $\underline{g}(\underline{\theta}^i, \underline{\delta})$.

The integral over $\underline{\theta}$ was approximated with a weighted sum where the points $\underline{\theta}^i$ and weights w_i can be based on either Monte Carlo or some Sparse Grid method (Bungartz and Griebel, 2004; Gerstner and Griebel, 1998). Note that due to the Laplace approximation of the $I(\underline{\theta})$ integral over the data $\underline{y}$, the integral of Equation (6.52) is only over the model parameters Θ, compared with the double integral of Equation (6.30) which is over Θ and Y. This allows for the application of Sparse Grid methods in order to estimate it, since the prior PDF $p(\underline{\theta})$ which serves as the weighting function is a common known PDF such as uniform or Gaussian. Equation (6.52) is the expected posterior entropy over the prior PDF, also proposed by Papadimitriou et al. (2000) using the robust information entropy concept.

The next simplification that occurs is that the minimization problem required to find $\underline{\theta}_*$ from Equation (6.39) needs not to be performed, since $J(\underline{y}^i, \underline{\theta}, \underline{\delta})$ from Equation (6.35) has a clear minimum of zero at $\underline{\theta} = \underline{\theta}^i$ which was used to simulate $\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})$. Zero is assured to be the minimum because $J(\underline{y}^i, \underline{\theta}, \underline{\delta})$ is a quadratic form with a positive definite matrix $\Sigma_{\underline{e}}(\underline{\delta})^{-1}$. $\Sigma_{\underline{e}}(\underline{\delta})^{-1}$ is positive definite because it is the inverse of the $\Sigma_{\underline{e}}(\underline{\delta})$ prediction error covariance matrix which is positive definite by construction. This is

true if we assume that the minimum of (6.39) w.r.t. $\underline{\theta}$ occurs at the minimum of $J(\underline{y}^i, \underline{\theta}, \underline{\delta})$. This holds when the prior PDF plays little or no role in the minimization w.r.t. $\underline{\theta}$, e.g. when the prior is uniform (no dependence on $\underline{\theta}$) or Gaussian with large variances (the variation of $p(\underline{\theta})$ w.r.t $\underline{\theta}$ is very small compared to the variation of $J(\underline{y}^i, \underline{\theta}, \underline{\delta})$ w.r.t. $\underline{\theta}$). The later case of a uncorrelated Gaussian prior with large variances $\Sigma_{p(\underline{\theta})} = diag(s_1^2, \ldots, s_{N_{\underline{\theta}}}^2)$ can be explained as follows:

$$p(\underline{\theta}) = N(\mu_{p(\underline{\theta})}, \Sigma_{p(\underline{\theta})}) \Rightarrow \log p(\underline{\theta}) = -\frac{1}{2}(\underline{\theta} - \mu_{p(\underline{\theta})})^T \Sigma_{p(\underline{\theta})}^{-1} (\underline{\theta} - \mu_{p(\underline{\theta})}) + C_4 \tag{6.53}$$

then (6.39) becomes:

$$\begin{aligned} &\frac{1}{2} J(\underline{y}^i, \underline{\theta}, \underline{\delta}) + \frac{1}{2}(\underline{\theta} - \mu_{p(\underline{\theta})})^T \Sigma_{p(\underline{\theta})}^{-1} (\underline{\theta} - \mu_{p(\underline{\theta})}) + C_5 \\ &\frac{1}{2}(\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta}))^T \Sigma_e(\underline{\delta})^{-1} (\underline{y}^i - \underline{g}(\underline{\theta}, \underline{\delta})) + \frac{1}{2} \sum_{k=1}^{N_{\underline{\theta}}} \frac{1}{s_k^2} (\underline{\theta}_k - \mu_{p(\underline{\theta})}^k)^2 + C_5 \end{aligned} \tag{6.54}$$

where it is clear that for large prior variances s_k^2 compared to the components of the prediction error covariance matrix $\Sigma_{\underline{e}}$ the second term is much smaller than the first and the main contribution in the minimization comes from minimizing $J(\underline{y}^i, \underline{\theta}, \underline{\delta})$ which is zero for $\underline{\theta} = \underline{\theta}^i$ which was used to simulate $\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})$, as stated above.

If the prior PDF does affect the minimization of (6.39) w.r.t. $\underline{\theta}$ then the minimization can not be avoided and must be performed with some minimization routine. However, that would render the asymptotic method very computationally expensive since this minimization would be required for every data sample $\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})$. We proceed by assuming that the prior is either uniform or Gaussian with large variances so that the minimum of (6.39) for a given $\underline{y}^i$ occurs at the $\underline{\theta} = \underline{\theta}^i$ which was used to simulate $\underline{y}^i$, hence avoiding the minimization.

Also the second term of (6.37) is eliminated since $\underline{g}(\underline{\theta}^i, \underline{\delta}) - \underline{g}(\underline{\theta}, \underline{\delta}) = \underline{0}$ for $\underline{\theta} = \underline{\theta}^i$ and therefore there is no need for second derivatives of the model. So the posterior covariance matrix takes the simpler form:

$$\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta}))} = [D_{\underline{\theta}}[\underline{g}(\underline{\theta}, \underline{\delta})]^T \Sigma_{\underline{e}}(\underline{\delta})^{-1} D_{\underline{\theta}}[\underline{g}(\underline{\theta}, \underline{\delta})] - \nabla_{\underline{\theta}}^2 (\log p(\underline{\theta}))]^{-1} | \underline{\theta} = \underline{\theta}^i \tag{6.55}$$

For uniform prior the second term is obviously zero because the uniform PDF is independent of $\underline{\theta}$

$$\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta}))} = [D_{\underline{\theta}}[\underline{g}(\underline{\theta}, \underline{\delta})]^T \Sigma_{\underline{e}}(\underline{\delta})^{-1} D_{\underline{\theta}}[\underline{g}(\underline{\theta}, \underline{\delta})]]^{-1} | \underline{\theta} = \underline{\theta}^i \tag{6.56}$$

For Gaussian prior with covariance $\Sigma_{p(\underline{\theta})}$ the posterior covariance matrix for data point $\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})$ becomes:

$$\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta}))} = [D_{\underline{\theta}}[\underline{g}(\underline{\theta}, \underline{\delta})]^T \Sigma_{\underline{e}}(\underline{\delta})^{-1} D_{\underline{\theta}}[\underline{g}(\underline{\theta}, \underline{\delta})] + \Sigma_{p(\underline{\theta})}^{-1}]^{-1} | \underline{\theta} = \underline{\theta}^i \tag{6.57}$$

We see that due to the assumption of either small prediction error magnitude or large number of uncorrelated data, combined with the Gaussian approximation of the posterior PDF (which becomes true for large number of data), we only need the first derivatives of the model w.r.t. parameters $\underline{\theta}$ to evaluate the posterior covariance matrix and hence the posterior entropy for data point $\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})$. The first

derivatives can either be provided analytically or estimated numerically by some numerical differentiation scheme. However, numerical estimation of the derivatives is prone to error and is also computationally expensive. Therefore, the asymptotic method is truly useful only when the derivatives of the model w.r.t. parameters are known analytically or can be estimated accurately. Otherwise it is best to use the sampling method instead.

Once the Jacobian matrix is known, then the objective function is evaluated as a weighted sum of the posterior entropies for various $\underline{\theta}^i$ from Equation (6.52), with the posterior covariance matrices for the $\underline{\theta}^i$ given by Equations (6.55), (6.56) or (6.57).

Note that when the number of sensors is smaller than the number of parameters and we use a uniform prior for the parameters, the posterior covariance matrix of Equation (6.56) is singular, leading to some zero eigenvalues and zero determinant. In that case the posterior entropy can not be evaluated. Papadimitriou and Lombaert (2012) used the product of the non-zero eigenvalues (pseudo-determinant) to proceed with a uniform prior. However, if a Gaussian prior is used instead, its covariance matrix is added to the first term and turns the posterior covariance of (6.57) into a non-singular matrix with a non-zero determinant. Therefore, the Gaussian prior enables to perform OED even for cases where the number of sensors is smaller than the number of model parameters we are trying to identify, without the need to use the pseudo-determinant.

6.4.1 Linear models

When the model is linear w.r.t. the parameters, that is, $\underline{g}(\underline{\theta}, \underline{\delta}) = A(\underline{\delta})\underline{\theta} + b(\underline{\delta})$, then the Jacobian matrix is independent of $\underline{\theta}$, $D_{\underline{\theta}}[\underline{g}(\underline{\theta}, \underline{\delta})] = A(\underline{\delta})$ and the posterior covariance matrix in Equation (6.57) does not depend on $\underline{\theta}$. Its determinant also does not depend on $\underline{\theta}$ and the integral of Equation (6.52) reduces to a constant which is the posterior entropy that depends only on $\underline{\delta}$.

$$U(\underline{\delta}) = H(p(\underline{\theta}|\underline{y}, \underline{\delta})) = \frac{1}{2}\log((2\pi e_b)^{N_{\underline{\theta}}}|\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})}|) \tag{6.58}$$

where the posterior covariance matrix is independent of $\underline{\theta}$ and $\underline{y}$:

$$\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})} = [A(\underline{\delta})^T \Sigma_{\underline{e}}(\underline{\delta})^{-1} A(\underline{\delta}) + \Sigma^{-1}_{p(\underline{\theta})}]^{-1} \tag{6.59}$$

In this case the Gaussian approximation of the posterior PDF becomes exact as already proved by (6.31) and (6.38) where we showed that the log-posterior has zero third and higher order derivatives w.r.t. $\underline{\theta}$.

Also, the Laplace approximation of the $I(\underline{\theta})$ integral over the data Y in (6.41) that was valid only under the assumption of either small prediction error magnitude or large number of uncorrelated data, becomes exact since the posterior entropy does not depend on the data any more and gets outside of the integral, yielding the same result:

$$\int_Y H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta}))\, p(\underline{y}|\underline{\theta}, \underline{\delta})\, d\underline{y} = H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta})) \int_Y p(\underline{y}|\underline{\theta}, \underline{\delta})\, d\underline{y} = H(p(\underline{\theta}|\underline{Y} = \underline{y}, \underline{\delta})) \tag{6.60}$$

Note that in the above Equations (6.58), (6.59) and (6.60) the dependence of the posterior on the data is written out in order to indicate that it is the posterior PDF, but there is no actual dependence on the data.

Finally, the drop of the second term of Equation (6.37) which was true only under the assumption of either small prediction error magnitude or large number of uncorrelated data, is now fully justified since for linear models the Hessian of the model prediction vector is zero ($\nabla^2_{\underline{\theta}}[A(\underline{\delta})\underline{\theta} + b(\underline{\delta})] = 0$), reducing the entire term to zero.

The above effectively mean that for linear models the asymptotic method is exact and not an approximation since all of its assumptions become true.

6.4.2 Effect of Gaussian prior variances

It turns out that Equation (6.57) has a very direct implication to the resulting optimal design. Next, a theorem is presented which explains the effect of a Gaussian prior PDF on the optimal design. Specifically, it is shown that the prior Gaussian variances of the parameters $\underline{\theta}$ act as weighting factors which give more preference to the identification of the parameters with the greatest prior variance. Hence, different optimal designs arise when using different prior variances, which favor the identification of the corresponding parameters.

First a diagonal covariance $\Sigma_{p(\underline{\theta})} = diag(s_1^2, \ldots, s_{N_{\underline{\theta}}}^2)$ for the Gaussian prior is assumed and the effect of the values of the variances is examined. These variances control the prior uncertainty in the values of the $N_{\underline{\theta}}$ model parameters. For clarity purposes, the dependence of the Jacobian matrix of the model prediction vector on $\underline{\theta}$ and $\underline{\delta}$ is dropped and we set $D = D_{\underline{\theta}}[\underline{g}(\underline{\theta}, \underline{\delta})]$ to denote the $N_s \times N_{\underline{\theta}}$ Jacobian matrix. Let also $\underline{D}_i$ denote the $N_{\underline{\theta}} \times 1$ column vector which is the transpose of the i-th row of D and corresponds to the first derivatives of the i-th component of $\underline{g}(\underline{\theta}, \underline{\delta})$ w.r.t. the model parameters $\underline{\theta}$.

For simplicity, the case of optimally placing one sensor is considered first and then generalized for the multiple sensor case. In the case of a single sensor at location i, the first term of the posterior covariance matrix of Equation (6.57) reduces to $\sigma^{-2}\underline{D}_i\underline{D}_i^T$. Note that for one sensor the prediction error covariance matrix is scalar $\Sigma_{\underline{e}}(\underline{\delta}) = \sigma^2$. Using the following known result for a square matrix B and two column vectors $\underline{u}$ and $\underline{v}$

$$\det(\underline{u}\underline{v}^T + B) = (1 + \underline{v}^T B^{-1}\underline{u}) \det(B) \tag{6.61}$$

the determinant of the posterior covariance matrix in Equations (6.52) and (6.57) takes the form

$$\det\left[\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})}\right] = \frac{1}{\det[\sigma^{-2}\underline{D}_i\underline{D}_i^T + \Sigma^{-1}_{p(\underline{\theta})}]} \tag{6.62}$$

$$\begin{aligned} \det\left[\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})}\right]^{-1} &= \det[\sigma^{-2}\underline{D}_i\underline{D}_i^T + \Sigma^{-1}_{p(\underline{\theta})}] \\ &= [1 + \sigma^{-2}\underline{D}_i^T\Sigma_{p(\underline{\theta})}\underline{D}_i]\, det(\Sigma^{-1}_{p(\underline{\theta})}) \\ &= [1 + \sigma^{-2}\sum_{k=1}^{N_{\underline{\theta}}} s_k^2 D_{ik}^2]\ \det(\Sigma^{-1}_{p(\underline{\theta})}) \end{aligned} \tag{6.63}$$

Note that the Gaussian prior variance s_k^2 of the k-th model parameter weights the contribution in the sum of the model derivative w.r.t. the k-th parameter, at the sensor location described by $\underline{\delta}$; remind that the Jacobian depends on both $\underline{\theta}$ and $\underline{\delta}$, $D = D_{\underline{\theta}}[\underline{g}(\underline{\theta}, \underline{\delta})]$. The higher the value of the variance s_k^2 is, the higher the contribution of D_{ik} is on the determinant. Since our goal is to minimize the posterior entropy (uncertainty), it is evident that the optimal design $\underline{\delta}$ will give preference to the sensitivities of the parameters that have higher prior uncertainty, i.e. higher s_k^2 values, and place the sensor at a location $\underline{\delta}$ where the derivative w.r.t. the k-th parameter D_{ik} is large. In that way the determinant of (6.63) is maximized, and the determinant of the posterior covariance matrix of (6.62) is minimized, which in turn minimizes the posterior entropy in (6.52).

Next we complete the proof in a similar way for the general case of an arbitrary N_s number of sensors. For this, we use the following known result of Linear Algebra

$$\det(UWV^T + B) = \det(W^{-1} + V^T B^{-1} U) \det(W) \det(B) \tag{6.64}$$

and the determinant of the posterior covariance matrix in Equations (6.52) and (6.57) takes the form

$$\det\left[\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})}\right] = \frac{1}{\det[D^T \Sigma_{\underline{e}}(\underline{\delta})^{-1} D + \Sigma_{p(\underline{\theta})}^{-1}]} \tag{6.65}$$

$$\begin{aligned}
\det\left[\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})}\right]^{-1} &= \det\left[D^T \Sigma_{\underline{e}}(\underline{\delta})^{-1} D + \Sigma_{p(\underline{\theta})}^{-1}\right] \\
&= \det\left[D^{'} \Sigma_{\underline{e}}(\underline{\delta})^{-1} D^{'T} + \Sigma_{p(\underline{\theta})}^{-1}\right] \\
&= \det\left[\Sigma_{\underline{e}}(\underline{\delta}) + D^{'T} \Sigma_{p(\underline{\theta})} D^{'}\right] \det\left[\Sigma_{\underline{e}}(\underline{\delta})^{-1}\right] \det\left[\Sigma_{p(\underline{\theta})}^{-1}\right] \\
&= \det\left[\Sigma_{\underline{e}}(\underline{\delta}) + D\, \Sigma_{p(\underline{\theta})}\, D^T\right] \det\left[\Sigma_{\underline{e}}(\underline{\delta})^{-1}\right] \det\left[\Sigma_{p(\underline{\theta})}^{-1}\right]
\end{aligned} \tag{6.66}$$

where $D^{'}$ was temporarily set as the transpose of the Jacobian matrix D.

The next step of the proof is to expand the matrix product $D\, \Sigma_{p(\underline{\theta})}\, D^T$ as a sum of matrices created from the outer products of the columns of D. For diagonal $\Sigma_{p(\underline{\theta})}$ we know from Linear Algebra that

$$D\, \Sigma_{p(\underline{\theta})}\, D^T = \sum_{k=1}^{N_{\underline{\theta}}} s_k^2\, D^k\, D^{kT} \tag{6.67}$$

where D^k is the k-th column of the Jacobian matrix which has the first derivatives of all N_s model predictions $\underline{g}(\underline{\theta}, \underline{\delta})$ w.r.t. the k-th model parameter. Then the determinant of the posterior covariance matrix of (6.66) becomes

$$\det\left[\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta})}\right]^{-1} = \det\left[\Sigma_{\underline{e}}(\underline{\delta}) + \sum_{k=1}^{N_{\underline{\theta}}} s_k^2\, D^k\, D^{kT}\right] \det\left[\Sigma_{\underline{e}}(\underline{\delta})^{-1}\right] \det\left[\Sigma_{p(\underline{\theta})}^{-1}\right] \tag{6.68}$$

Equation (6.63) is a special case of equation (6.68) when only one sensor is used ($N_s = 1$) and the Jacobian matrix D becomes a row vector and its columns become scalars. In the multiple sensor case we are dealing with a sum of rank-one matrices over the model parameter first derivatives, where each

rank-one matrix is formed from the outer product of the first derivatives of the N_s model prediction components w.r.t. the k-th parameter, D^k, weighted by the corresponding prior variance s_k^2 of that parameter. So we see again that, similarly to the one sensor case, the Gaussian prior variance for the k-th parameter s_k^2 acts as weighting factor, this time by giving greater weight to the matrix $D^k\, D^{kT}$ related to the model sensitivities w.r.t. the k-th parameter. The higher the value of the prior variance s_k^2 is, the higher the contribution of the model sensitivities w.r.t. the k-th parameter is. So the optimal design of sensor locations $\underline{\delta}_*$ is such that the model sensitivities w.r.t. the high-variance parameters are the largest possible. This in turn maximizes the right-hand side of Equation (6.68) and minimizes the posterior covariance matrix determinant (entropy). Therefore, the optimal design gives preference to the identification of parameters which have the greatest prior uncertainty.

Note that the actual objective function is an average of posterior entropies for several points over the parameter space $\underline{\theta}^i$ described by the prior PDF $p(\underline{\theta})$, as shown in Equation (6.52). However, this does not affect the validity of the theoretical result since it applies for every $\underline{\theta}^i$ point individually, and on average over all points.

An intuitive explanation of this theoretical result follows: Our objective is to gain the most information about the model parameters when updating the prior to the posterior PDF. The prior uncertainty in the parameters is quantified by the prior PDF. In order to gain insight about this result, we consider a limiting case. Assume that most of the model parameters have very small prior variances and the rest have large. The parameters with the very small variances are essentially already known, and learning more about them will not increase the information gain substantially. It does not make sense to try to identify parameters that are already known with very small uncertainty. In other words, there is not enough room for information gain regarding those parameters. However, the parameters with the large prior variances offer large information gain potential, since they are not known at all (almost), and learning more about them would contribute largely to the overall information gain. Therefore, their identification is preferred by the optimal design. Different weights (prior variances) can be attributed to all the parameters, indicating their significance in the identification procedure.

This theoretical result enables us to perform more sophisticated OED with the asymptotic method, which gives more weight to the identification of some parameters over others instead of all the parameters having the same weight. This may be proved useful in cases were some of the model parameters are of more interest than others because for example they play a more important role in the behavior of the system. Small variances can be assigned to parameters that are not important so that the optimal design will focus on the more important parameters with larger variances. Equal variances can always be used to indicate that all the parameters are equally important. This theoretical result has been used for the purpose of designing the optimal acceleration sensor locations for modal identification of a bridge, while giving preference to specific modes of interest (Argyris et al., 2016). Whether or not the same theoretical result can be proved under the sampling OED framework is an open question that is subject for future research.

6.5 Numerical application

In order to demonstrate the key difference between the two OED methods, but also to verify the increased quality of the new proposed sampling estimator introduced in Section 6.3.2, a numerical example is carried out in a simple analytical beam model with uncertain load.

6.5.1 Problem description

The illustrative problem that was selected in order to demonstrate the findings is the optimal sensor placement in a simply supported beam that is under an uncertain point load. The beam is shown schematically in Figure 6.2. Both the location a and magnitude P of the load are uncertain, and these are the model parameters $\underline{\theta}$ that we would like to identify from the experiment. We would like to perform the experiment by placing strain sensors at the top surface at any point in the beam. So effectively we are trying to find which is the optimal location of a number of strain sensors so that we identify both the load magnitude and location. The strain of a simply supported beam under a known point load can be found in any classical textbook of Strength of Materials. Therefore, the strain prediction from the model at any location x is given by:

$$\begin{aligned} \epsilon(x) &= AP(L-a)x \quad \text{for } 0 \leq x \leq a \\ &= APa(L-x) \quad \text{for } a < x \leq L \end{aligned} \tag{6.69}$$

where A is a constant that depends on the geometric and material properties of the beam. Without loss of generality we assume that the constants are such that $A = 1$. Since our parameter vector $\underline{\theta}$ is $\underline{\theta} = [P \;\; a]^T$ we can immediately note that even though the model is linear in the load magnitude P, when the load location a is also included as a parameter the model becomes non-linear due to the product terms Pa that appear. Another factor that contributes to the non-linearity of the model predictions w.r.t. the parameters $\underline{\theta}$ is that it is a piece-wise model which changes its function depending on the value x where we evaluate the strain, and the distance where the two pieces separate from each other is a model parameter, namely a. The beam is of length 1 and the priors of the model parameters were chosen to be uniform in the range $[0.55 \;\; 0.85]$ and $[0.85 \;\; 1.15]$ for the load location and magnitude, respectively. A second symmetric prior for the load location was also examined in the range $[0.35 \;\; 0.65]$. Therefore, the uncertain load could be applied anywhere in the intervals $[0.55 \;\; 0.85]$ and $[0.35 \;\; 0.65]$ with equal probability. The prediction error vector of the model predicted strains was chosen to be Gaussian with zero mean and covariance matrix $\Sigma_e = \sigma^2 I$ where σ was chosen to be 1% and 10% of the average model prediction for the given priors.

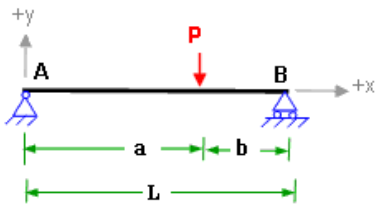


Figure 6.2: Simply supported beam with applied load

Our goal is to find the optimal locations of strain sensors on the beam such that the most information will be contained about the load magnitude and location. The design variables $\underline{\delta}$ are the locations of the strain sensors on the beam. In order to avoid having too many parameters by assigning each design variable to a sensor coordinate, we simplify the design variables to just 2 parameters, the first being the location of the first sensor and the second being the uniform distance between the sensors. In this way we limit the number of optimization variables to just 2, despite the actual number of sensors. This decision was backed up by preliminary results using the coordinates of each sensor as design variables, which clearly showed that indeed they are uniformly spaced. The optimization of the objective function w.r.t. the 2 design variables in order to find the optimal design is done using the CMA algorithm (Hansen et al., 2003). The choice of this algorithm has to do with some of its convenient properties:

1. It is a gradient-free algorithm which does not require derivatives of the model prediction w.r.t. the design variables (which are in many problems impossible to obtain), but only forward model evaluations for given design variables.

2. It is a stochastic optimization algorithm with global-search capabilities and has higher chances to locate the global optimum compared to gradient-based methods which can easily get trapped in local optimal.

3. It is well-suited for noisy objective functions which are estimated through Monte Carlo methods.

4. It can be highly parallelized since at every iteration multiple model runs are required for different design variables which can be executed in parallel and therefore reduce computational time significantly.

6.5.2 Comparison of sampling estimators

In section 6.3.2 a simplified Monte Carlo estimator was derived which exploits the fact that the likelihood entropy is known analytically and needs not to be approximated with samples. It was derived under the assumption of the Gaussian likelihood (Gaussian prediction error vector) which is by far the most common case because of the Maximum Entropy principle (Cover and Thomas, 2006). We claimed that this simplification would result in a more accurate estimate of the objective function. In this section we

demonstrate this with a numerical example. Specifically, we compare the two estimators by evaluating the objective function 100 times with each estimator and compare the statistical descriptors of each case. Remind that since we only have a Monte Carlo estimate of the objective function, there is also unavoidably some noise associated with that estimate due to the random but specific $\underline{\theta}^i$ and $\underline{y}^i$ samples used in the estimate. Therefore, each objective function evaluation gives a slightly different estimate when those samples are randomly drawn again. An attempt is made to measure that variability by evaluating the objective 100 times and taking the mean and standard deviation. This is done for both estimators of Equations (6.8) and (6.25). Next, in order to eliminate any possible random artifacts of the estimators, this is repeated for an increasing (but equal) number of samples for both estimators so that we have more statistic data to compare.

In Figure 6.3 we see the mean ± 2 standard deviations of 100 objective function evaluations for each estimator, for a increasing number of samples. The objective function $U(\underline{\delta})$ was evaluated for 2 sensors placed at the bounds of the uniform priors of the load location at [0.55 0.85] and [0.35 0.65]. In this context, finding the optimal sensor location $\underline{\delta}_*$ is not of interest since our goal is just to compare the two estimators. So any value $\underline{\delta}$ would be sufficient to study the behavior of $U(\underline{\delta})$ under the two estimators.

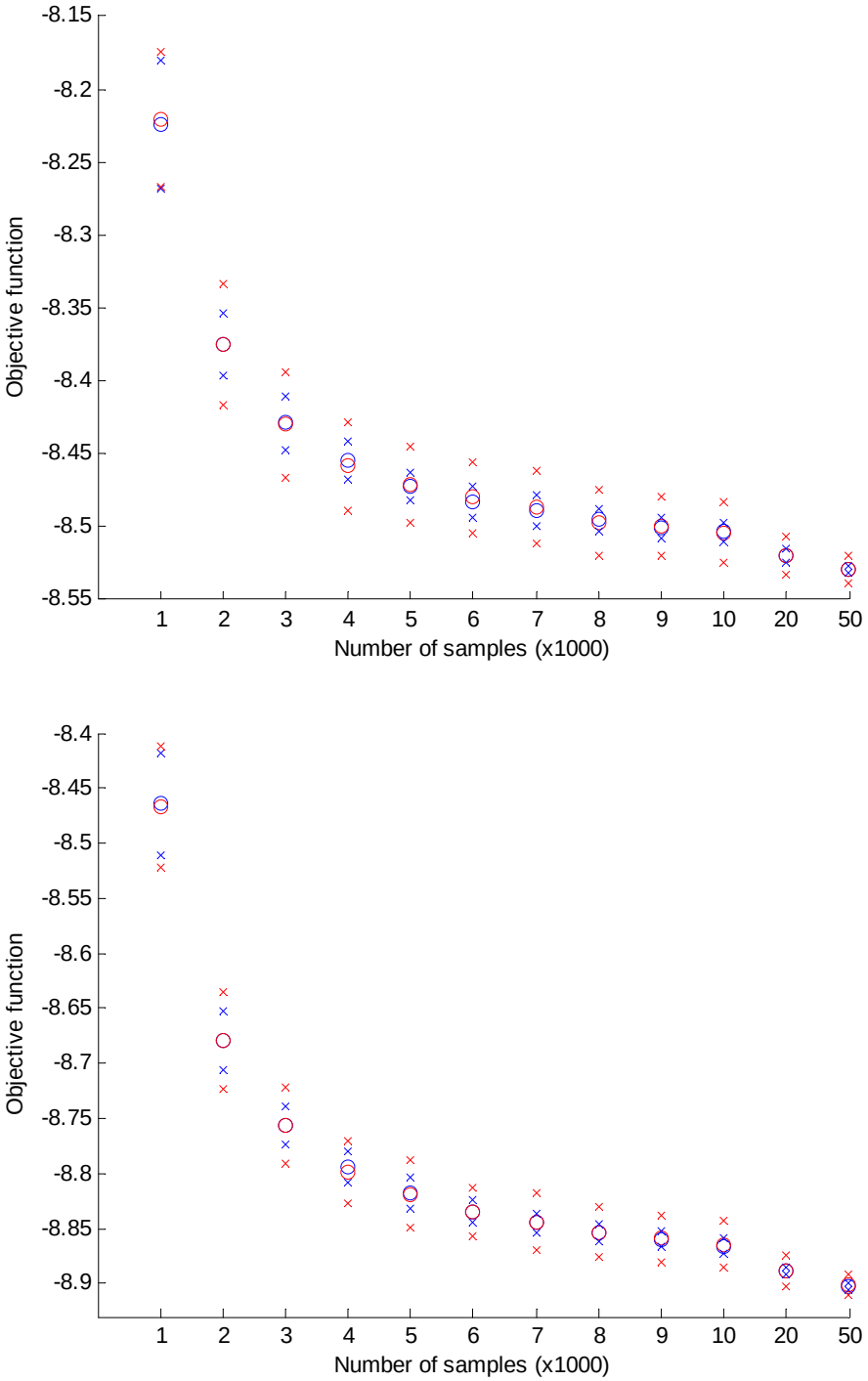


Figure 6.3: Comparison of variance of the two estimators. Blue: Equation (6.25), Red: Equation (6.8). Top: Load location prior [0.55 0.85], Bottom: Load location prior [0.35 0.65].

Notice how the proposed estimator indicated by the blue color has a smaller standard deviation than the estimator proposed by Ryan (2003) indicated by the red color. This is consistent for all numbers of samples and therefore not an artifact caused by the stochastic nature of the estimator. We see the expected result that as the number of samples increases the objective function estimate stabilizes towards the true value, and the variance decreases. This clearly shows that the quality of the estimator has improved not by reducing its bias but by reducing its variance.

6.5.3 Results based on both sampling and asymptotic methods

Both sampling and asymptotic methods are used to perform optimal sensor placement on the beam for model parameter inference. The objective function for the sampling method is Equation (6.25) which is the expected KL-divergence between the posterior and prior PDFs or the Mutual Information between the data and parameters equivalently. The objective function for the asymptotic method is Equation (6.52) which is the expected posterior entropy over the prior. In order to compare results from both methods Equation (6.27) is used which connects the two quantities through the entropy of the prior PDF. In the following results, the expected posterior entropy is used as the objective function, therefore the optimal design is the one that minimizes the objective.

The sampling method requires only evaluation of the model for various parameter points $\underline{\theta}^i$, which is done from Equation (6.69), while the asymptotic method requires the first derivatives of the model w.r.t. $\underline{\theta} = [P \;\; a]^T$ which are also readily available by taking the derivatives of Equation (6.69). Note how the evaluation of the strains or the derivatives of the strains for various parameter points $\underline{\theta}^i$ and given sensor locations $\underline{\delta}$ always requires examining the location of the sensors with respect to the location of the load a which is a parameter. In order to illustrate the effect of the prediction error magnitude in the results using both methods, two cases were examined with the prediction error standard deviation being 1% and 10% of the average model prediction for the given priors.

From the numerical perspective, both the sampling and asymptotic objective functions are integrals which can always be approximated with Monte Carlo sampling. However, the asymptotic objective function offers itself for more accurate numerical integration methods due to its simpler form of Equation (6.52) compared to the sampling objective function of Equation (6.25). The simpler form is due to 2 reasons:

1. The asymptotic objective function is a single integral over Θ whereas the sampling is a double integral over both Θ and Y.

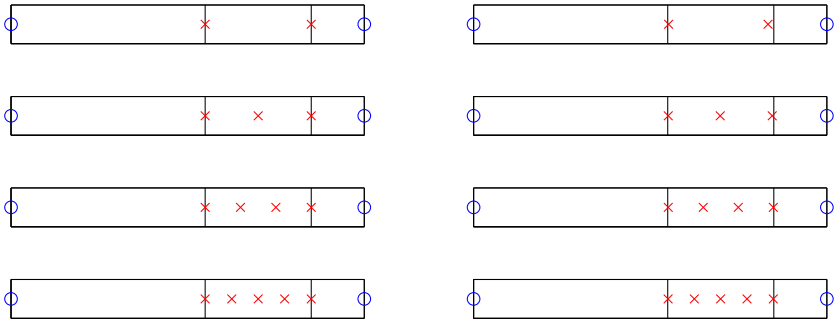
2. The integral over Θ can be conveniently approximated with Sparse Grids since the prior PDF which acts as the weighting function is a well-known PDF, whereas the integral over both Θ and Y can only be approximated with Monte Carlo sampling because the product of the likelihood and prior PDFs which acts as the weighting function is not necessarily a well-known PDF. Also, for small number of model parameters, even numerical quadrature can be used to integrate the asymptotic objective function.

3. The sampling objective function, due to the fact that it also uses data samples $\underline{y}^i$ except from parameter samples $\underline{\theta}^i$ it introduces another source of uncertainty due to $\underline{y}^i$. This is because the data samples $\underline{y}^i$ are sampled randomly by adding a disturbance (prediction error) in the model prediction for the corresponding $\underline{\theta}^i$, $\underline{g}(\underline{\theta}^i, \underline{\delta})$. This added source of randomness does not exist in the asymptotic objective function because it does not require data samples but only parameter samples. The data samples are considered equal to the model prediction for the corresponding $\underline{\theta}^i$ under the assumption of small prediction error.

Finally, we should note that during the optimization of the sampling objective function w.r.t. the design variables, the $\underline{\theta}^i$ samples used to estimate the objective were kept the same for all objective function evaluations. This is common practice in the optimization of noisy objective functions given by Monte Carlo estimates in order to avoid extra noise added to the procedure due to the sample variability, making the optimization harder. In Section 6.5.2 the samples were deliberately re-drawn for every objective function evaluation in order to show the reduced variability of the new estimator.

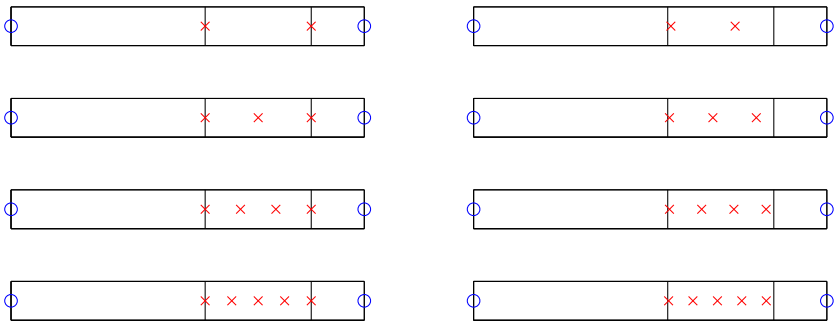
Next, the optimal sensor placement results are presented for 2, 3, 4, 5 and 20 sensors. For the sampling objective function, 8.000 Monte Carlo samples were used to approximate it. In order to assess the effect of that number of samples in the results, indicative runs with 2 sensors were also conducted with 20.000 and 50.000 samples which confirmed the results produced using 8.000 samples, judging that 8.000 samples were sufficient. For the asymptotic objective function, numerical quadrature was used since the integral over the parameters is only 2-dimensional. The number of iterations required to perform the quadrature was of the order of 2.000, which should be compared with the 8.000 Monte Carlo samples where the sampling objective function begins to stabilize, as shown in Section 6.5.2. The minimization of the objective function (expected posterior entropy) in both the asymptotic and sampling methods was done using the CMA algorithm (Hansen et al., 2003) for the reasons described in Section 6.5.1.

First, the case with the asymmetric load location prior of [0.55 0.85] is presented. Figure 6.4 shows the optimal locations of sensors in the beam for 1% prediction error using both methods, and Figure 6.5 shows the same for the 10% case (the 20 sensor case is omitted in these figures for clarity purposes). The results are summarized in Tables 6.1 and 6.2, while Figure 6.6 is the plot of the objective function (expected posterior entropy) versus the number of sensors.



(a) Asymptotic (b) Sampling

Figure 6.4: Case of 1% prediction error: Optimal sensor locations



(a) Asymptotic (b) Sampling

Figure 6.5: Case of 10% prediction error: Optimal sensor locations

Table 6.1: OSL results for 1% prediction error magnitude

	Sampling			Asymptotic		
# of sensors	Location of first sensor	Distance between sensors	Objective function	Location of first sensor	Distance between sensors	Objective function
2	0.5524	0.2815	-8.5128	0.55	0.3	-8.3365
3	0.5514	0.1472	-9.0091	0.55	0.15	-8.9095
4	0.5514	0.0991	-9.2086	0.55	0.1	-9.1480
5	0.5507	0.0745	-9.3504	0.55	0.075	-9.2941
20	0.5507	0.0156	-9.9165	0.55	0.0158	-9.8203

Table 6.2: OSL results for 10% prediction error magnitude

	Sampling			Asymptotic		
# of sensors	Location of first sensor	Distance between sensors	Objective function	Location of first sensor	Distance between sensors	Objective function
2	0.5587	0.1819	-4.6275	0.55	0.3	-3.8750
3	0.5548	0.1229	-4.8973	0.55	0.15	-4.4577
4	0.5546	0.0913	-5.0353	0.55	0.1	-4.7003
5	0.5522	0.0691	-5.2025	0.55	0.075	-4.8490
20	0.5528	0.0153	-5.6719	0.55	0.0158	-5.5842

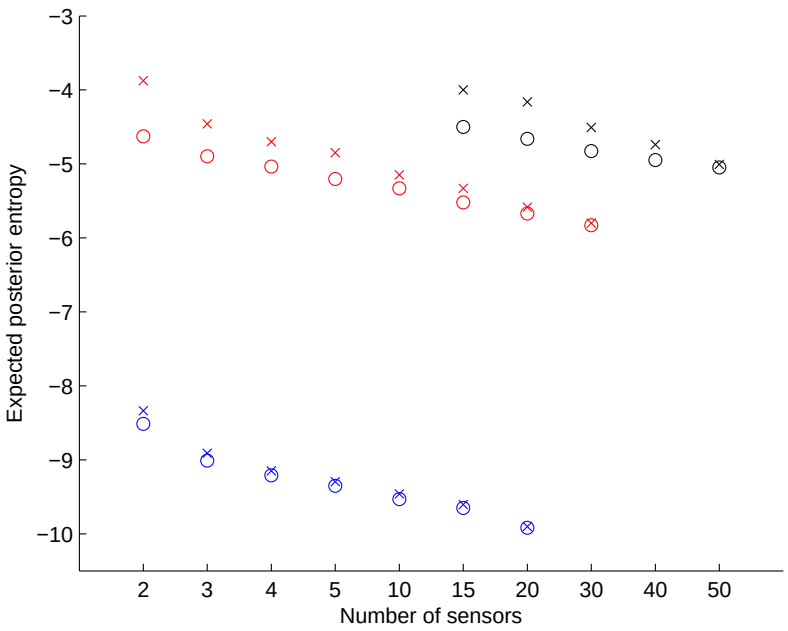


Figure 6.6: Expected posterior entropy versus number of sensors for both methods and prediction error cases. Blue: 1% error, Red: 10% error, Black: 25% error. Circle: Sampling method, X: Asymptotic method.

First we note that for only 2 sensors and 1% error, they are optimally placed in the bounds of the uniform prior of the load location at [0.55 0.85] and not on the interior of the interval. This happens due to the piece-wise nature of the model, which changes its function type depending on the value of the load location parameter. In order to be able to identify both the load magnitude P and location a we need to have strain sensors in both sides of the load location. If sensors are in the same side the model does not have any way to know both P and a, no matter how many sensors we use. One can see that by noting that the equation for the strains (6.69) is the same for any number of sensors, if the x^i coordinates of those sensors are all $\leq a$ or $\geq a$. Then it is impossible to know both P and a using only one equation, no matter how many sensors there are. One can only know their ratio, but not both. But since a is uncertain and uniformly distributed in [0.55 0.85], the only way to ensure that the 2 sensors are going to be in both sides of the load is to place them in the bounds at 0.55 and 0.85. As the number of sensors is increasing, we see the tendency to cover the entire prior interval for the load location. This is done for the same reason, in order to ensure having sensors in both sides of the load.

The next thing we note from the results, and especially from Figure 6.6 is that for 1% prediction error the expected posterior entropy is lower than for the 10% prediction error with both methods (less uncertainty in the identified parameters). This is in agreement with the Bayesian theory because less uncertainty in the data (small prediction error) is expected to give less uncertainty in the identified

parameters (small posterior entropy), and this is expressed in Figure 6.6.

From the same Figure 6.6 we can also draw other useful conclusions about the sampling and asymptotic methods. The most notable difference is how the two methods are very close to each other for the 1% prediction error, and they start to deviate significantly as the prediction error increases to 10% and 25%. This is due to the assumption of the asymptotic method that the prediction error magnitude is very small. Here we see that a 1% prediction error was sufficiently small to cause almost the same results as with the sampling method, but with 4 times less computational effort used in the objective function; 2.000 numerical quadrature iterations for the asymptotic method compared to the 8.000 samples of the sampling method. This applies to both the optimal objective values and optimal designs shown in Figure 6.4.

However, we also proved that the asymptotic method gets more accurate not only by lowering the prediction error magnitude, but by increasing the number of uncorrelated data. This can be seen in the numerical results, where it is evident that even for the 10% and 25% prediction error cases the asymptotic method does approach the exact sampling method as the number of sensors is increasing. The higher the prediction error magnitude is, the more data points are required in order to converge to the sampling method. This shows that the asymptotic method can be considered even for non-negligible prediction errors, provided that the number of sensors (data) is large enough. Plots like Figure 6.6 can be used to decide how many sensors are sufficient.

It is evident from Figure 6.5 that the asymptotic method was not capable of adjusting its optimal designs for up to 5 sensors in response to the increase of the prediction error to 10%. Only the objective values changed (Table 6.2), but the optimal designs remained the same for both 1% and 10% cases of the prediction error magnitude. The sampling method, on the other hand, modified its optimal designs according to the increase in the prediction error by shortening the distance between the sensors. Only when the number of sensors reached 20 did the asymptotic method start to agree with the sampling method. It seems from Tables 6.1 and 6.2 that as we increase the prediction error magnitude, the distance between the sensors decreases, resulting in the far-right sensor moving away from the right prior bound at 0.85 and towards the interior of the prior interval, as can be seen in Figure 6.5.

It is not obvious why this tendency exists, but it is speculated that it has to do with the asymmetric prior of the load location at [0.55 0.85] with respect to the beam, which causes an uneven strain distribution. Figure 6.7 shows the strain distributions in the beam for various random samples from the load location prior PDF, together with the associated prior bounds. For simplicity, the load magnitude P in these plots was set equal to 1. What should be noted in this Figure is how much sharper the strain variations are close to the right bound at 0.85 compared to the left bound at 0.55. These sharp variations cause the strains close to 0.85 to be much lower (on average) than the strains close to 0.55. This becomes a problem when the prediction error increases because it can lead to very small strain readings from sensors close to 0.85. Remind that the data samples $\underline{y}^i$ which are used in the sampling objective function are simulated by adding a disturbance (random draw) from the prediction error PDF $N(\underline{0}, \Sigma_{\underline{e}}(\underline{\delta}))$ to the model predicted strains using the i-th parameter sample $\underline{\theta}^i$. When the prediction error increases,

this added disturbance can lead to very small (also very large, but the small case is the one causing the problem here) strain readings from a sensor close to 0.85 and this is not desired because ideally we would want large strain readings for better identification. This is why the second sensor is optimally located further inside the interval at 0.74 so that small strain readings from a sensor close to 0.85 are avoided. In the left bound at 0.55 this is not a problem because as we can see from Figure 6.7 the strains close to it are quite large on average.

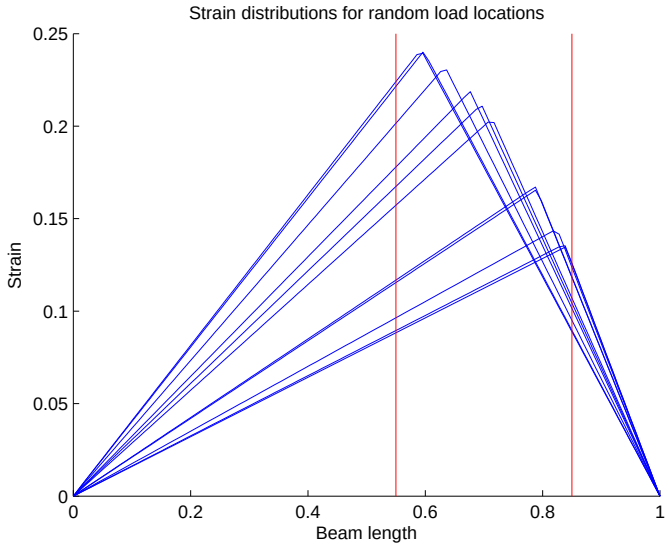


Figure 6.7: Strain distributions for multiple random samples of the load location for the asymmetric prior PDF [0.55 0.85]

In order to check this intuitive explanation for the results, a second case with a symmetric load location prior is used at [0.35 0.65]. In that case, as can be seen in Figure 6.8, the strain distributions are also symmetric on average and there are no areas in the beam with sharper variations than others on average as was the case with the asymmetric prior before. Therefore, according to our explanation, there should not be any reason for the 2 sensors to avoid any bound more than the other, and the 2 sensors should be placed symmetrically as well at the prior bounds. According to the proposed explanation, it is also expected that as the prediction error magnitude increases, both of the sensors would move towards the inside of the symmetric prior interval [0.35 0.65], while remaining symmetric. In the asymmetric prior case the 2 sensors did not remain symmetric with the increase of the prediction error, and this behavior was attributed to the sharper strain variations close to the right bound at 0.85, but with the symmetric prior this reason does not exist anymore and therefore the sensor locations would still remain symmetric even after the increase of the prediction error magnitude.

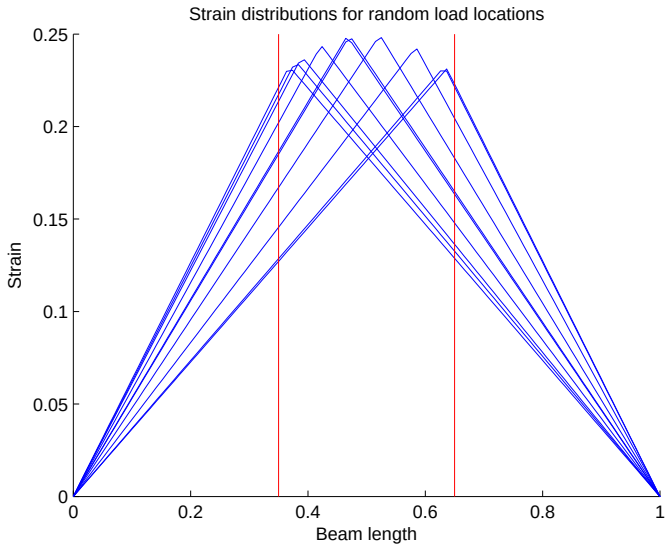


Figure 6.8: Strain distributions for multiple random samples of the load location for the symmetric prior PDF [0.35 0.65]

The CMA algorithm was run again for the case of 2 sensors with the following change in the objective function: the asymmetric load location prior PDF [0.55 0.85] was replaced by the symmetric prior PDF [0.35 0.65]. Again, both cases of 1% and 10% prediction error magnitude were examined. For the 1% error, the 2 sensors are optimally placed at 0.3546 and 0.6467 very close to the prior bounds, and for the 10% error they are optimally placed at 0.3749 and 0.6245. The optimal sensor locations are in fact symmetric, as expected. Also, the idea about the prediction error causing the sensors to move further away from the prior bounds and towards the inside of the prior interval seems to be reinforced as well. The optimal sensor locations are symmetric even for the 10% prediction error case, and we do not see one sensor being more displaced than the other as was the case with the asymmetric prior. For the 1% prediction error they remain symmetric, and approach the prior bounds at [0.35 0.65].

The above analysis allowed us to understand two basic ideas about this problem:

1. The sensors must be placed as close as possible to the uniform prior bounds in order to make sure that we have sensors on both sides of the load application point which is uncertain.

2. We must avoid small strain readings from the sensors.

When the prediction error magnitude is small, both of the above points can be satisfied by placing sensors very close to the uniform prior bounds for the load location. However, when the prediction error increases, the strain readings from the sensor close to 0.85 can become dangerously small, so that the sensor has to move towards the inside of the interval. Then there is a natural trade-off between the two above mentioned ideas and an optimal location must be found where both are satisfied to a lesser degree. The

sampling method is capable of finding that optimal trade-off point while the asymptotic method is not.

As more sensors are added in the beam, the far-right sensor can approach 0.85 because the rest of the sensors make up for its potential small strain readings, while covering the entire prior interval which is desired for having sensors in both sides of the load on average. The above clearly demonstrate that there are notable disagreements between the two methods when the prediction error magnitude increases, and that its effect cannot be studied with the asymptotic method but rather the computationally more expensive sampling method must be used instead if we would like to perform such analysis.

Next, we should note from Figure 6.6 that especially for the 10% prediction error case, the objective values obtained with the sampling method are lower (better) than the asymptotic method. This suggests that the modified optimal designs obtained with the sampling method in Figure 6.5 are in fact better for identifying the load parameters than their counterparts from the asymptotic method in the same Figure, when the prediction error is 10%.

From Figure 6.6, it is also interesting how the entropy for just 2 sensors with 1% error is lower than the entropy of 20 sensors with 10% error. This shows how important the effect of the prediction error is in the identification. The fact that in the later case we have much more sensors does not imply that their information is also much more useful, since every one of those 20 sensors has a 10% error associated with its reading and therefore can make the identification much more difficult. This implies that good quality in the data is much more important than quantity.

Finally, from Figure 6.6 we also observe the intuitive result from the Bayesian theory that as the number of sensors increases the expected posterior entropy decreases. This is expected since more sensors means more data and therefore better parameter inference with less uncertainty in the identified parameters. We also see that this decrease in the posterior entropy is sharper for the first added sensors and then it drops. This happens because if many sensors are already placed in the beam, adding one more does not contribute very much to the total information from all sensors. Whereas when there is a small number of sensors placed on the beam, a new sensor has a large contribution to the overall information. This issue has been studied in detail by also introducing correlated prediction error models with non-diagonal prediction error covariance matrices $\Sigma_{\underline{e}}(\underline{\delta})$ by Papadimitriou and Lombaert (2012).

6.6 Conclusions

This chapter presented a novel evidence-based interpretation to Optimal Experimental Design, based on a new improved equation for the objective function. This improved equation uses Monte Carlo sampling for 1 term only instead of 2 terms, and calculates the other term analytically. As expected, this was demonstrated numerically to reduce the variance of the estimator and hence increase its quality. It turns out that the evidence is the crucial factor for OED, and the optimal design is the one that can maintain a good fit with the data despite their uncertainty. Designs that give a good fit with only some of the data and fail with others are not preferred.

Both the sampling and asymptotic methods to OED were developed, and objective functions used in

different works were put under the same framework. It was shown that the asymptotic method requires only the first derivatives of the model w.r.t. $\underline{\theta}$ evaluated at different values of $\underline{\theta}$ for given design variables $\underline{\delta}$. The sampling method requires no derivatives, but only point-evaluation of the model. Using either the KL divergence between posterior and prior or the posterior entropy as objective function (both averaged over the data) result in the same optimal design since they only differ by a constant term which is the prior entropy. An important theorem was developed for the asymptotic method, explaining the effect of the Gaussian prior variances of the model parameters. This theorem proves that the Gaussian prior variances act as weighting factors in the identification of the parameters with the highest variance. Hence, the optimal design favors the identification of the parameters with the highest prior variance.

We also presented a general probabilistic framework for OED where the design variables themselves are treated as uncertain variables, and showed that existing formulations with deterministic design variables can be derived as special cases of the more general formulation. Even though not explored in this work, the probabilistic framework can be used to find optimal designs that are robust to changes in the design variables. The framework was also extended to handle nuisance parameters that enter the formulation but we do not wish to update them, but rather include their uncertainty in the optimal design formulation.

Also note that we never introduced a utility function (Lindley, 1956) that depends on the data and then have to take the expectation over the data because they are uncertain. Instead, this expectation occurred naturally by using the Mutual Information between the parameters and data as the objective function. This shows that the known theory of maximizing the expected utility function, when the utility function is taken as the KL-Divergence between the posterior and prior PDFs of the parameters (parameter inference), can be derived under the theory of maximizing the Mutual Information between the parameters and data when the design variables are deterministic.

Finally, a numerical application was presented using a simple beam model in order to demonstrate: 1) The variance reduction achieved using the new improved formula for the sampling objective function estimator, 2) The key difference between the sampling and asymptotic methods regarding the magnitude of the prediction error and the number of data. When small prediction errors are considered, of the order of 1% of the model and smaller, then the two methods seem to agree and the asymptotic method is preferred over the sampling method because it is computationally less expensive. As the prediction error magnitude was increased, the two methods started to diverge because the asymptotic method loses its accuracy. However, increasing the number of data (sensors) helped the asymptotic method regain its accuracy and led to agreement with the sampling method even for larger prediction errors.

Finally, we can conclude that the asymptotic method is reliable when the prediction error magnitude is very small, or the number of data is very large. Plots like Figure 6.6 can be used to decide how many sensors are sufficient for a given prediction error magnitude. Otherwise, if the number of sensors is small or the prediction error magnitude is large, it is mandatory to use the sampling method.

Chapter 7

Bayesian optimal sensor placement for modal identification of civil infrastructures

Original publication:
Argyris, C., Papadimitriou, C. and Panetsos, P. "Bayesian Optimal Sensor Placement for Modal Identification of Civil Infrastructures". In: *Journal of Smart Cities*, 2(2): 69-86, (2016).

Abstract

A Bayesian optimal experimental design (OED) method is proposed in this work for estimating the best locations of sensors in structures so that the measured data are most informative for estimating reliably the structural modes. The information contained in the data is measured by the Kullback-Leibler (K-L) divergence between the prior and posterior distribution of the model parameters taken in modal identification to be the modal coordinates. The optimal sensor placement that maximizes the expected K-L divergence is shown also to minimize the information entropy of the posterior distribution. Unidentifiability issues observed in existing formulations when the number of sensors is less than the number of identified modes, are resolved using a non-uniform prior in the Bayesian OED. An insightful analysis is presented that demonstrates the effect of the variances of Bayesian priors on the optimal design. For dense mesh finite element models, sensor clustering phenomena are avoided by integrating in the methodology spatially correlated prediction error models. A heuristic forward sequential sensor placement algorithm and a stochastic optimization algorithm are used to solve the optimization problem in the continuous physical domain of variation of the sensor locations. The theoretical developments and algorithms are applied for the optimal sensor placement design along the deck of a 537 m concrete bridge.

7.1 Introduction

Experimental measurements from civil infrastructures, including buildings, bridges, offshore structures, wind turbines, industrial facilities, are often used to obtain the modes of these structures. The modes are useful in structural performance evaluation (Simoen et al., 2013a), finite element model updating (Beck and Katafygiotis, 1998; Yuen, 2012, 2010) and model-based structural health monitoring (Lam et al., 2004; Vanik et al., 2000). A number of methodologies have been developed in the past to optimize the location of sensors in order to maximize the information contained in the measurements for identifying the structural modes. Among them, methods based on information theory are used to make rational decisions consistent with the information provided by the measurements.

This chapter concentrates on optimal sensor placement design methods for modal identification based on information theory. In the past, notable contributions to the sensor placement problem for modal identification have been provided by the effective independence (EFI) method (Kammer, 1991) for uni-axial and triaxial sensors (Li et al., 2009). Information theory measures, based on scalar measures of the Fisher Information Matrix (FIM) (Shah and Udwadia, 1978; Udwadia, 1994) and on information entropy (Papadimitriou et al., 2000; Yuen et al., 2001; Ye and Ni, 2012), proposed in the past for structural parameter estimation problems, have been extended to be used for modal identification (Papadimitriou and Lombaert, 2012) as well. An optimal sensor placement design for modal identification based on FIM was proposed by Kammer (2005). The information entropy measure for parameter estimation introduced by Papadimitriou et al. (2000), was extended by Papadimitriou (2004) to obtain useful expressions of the information entropy as a function of the number of sensors, and was applied to modal identification problems by Papadimitriou and Lombaert (2012). In particular, the information entropy measures the uncertainty in the posterior distribution of the model parameters to be identified. The posterior distribution is obtained from a Bayesian analysis that makes use of the prior distribution of the model parameters. Up to now, the effect of the prior distribution in the optimal design has not been adequately explored.

Applying the EFI and information entropy techniques to dense finite element models, the problem of sensor clustering is manifested. The source of the problem is the failure to take into account the redundant information provided from neighborhood sensor locations. The problem was adequately resolved using spatially correlated prediction error models in the Bayesian formulation to exclude redundant information from neighboring sensors (Papadimitriou and Lombaert, 2012). Stephan (2012) has also effectively tackled the issue of information redundancy between sensors by introducing a measure of information redundancy.

Optimization algorithms have also been proposed to find the optimal sensor locations. For structural parameter estimation problems multiple local/global optima make the solution of the optimization problem very challenging. Heuristic algorithms such as the backward and forward sequential sensor placement (BSSP and FSSP) algorithms have been proposed to drastically reduce the computational effort (Kammer, 2005; Papadimitriou, 2004). The optimal sensor placement strategies based on the information entropy and FSSP was applied to bridge, towers and timber structures for optimizing the location of

uniaxial and triaxial sensors (Chow et al., 2011). These algorithms are shown to be quite accurate. However, the use of the FSSP algorithm requires that identifiability issues are resolved for small number of sensors. Unidentifiability issues arise from the aforementioned methods for modal identification when the number of sensors placed in the structure is less than the number of identified modes. This is due to the fact that the FIM becomes singular. Yuen and Kuok (2015) noted that introducing non-uniform priors in the Bayesian posterior used in the information entropy measures resolves the problem. For uniform priors, Papadimitriou and Lombaert (2012) obtained reasonable sensor placement designs by excluding the zero eigenvalues from the product of the eigenvalues of the FIM used for computing the determinant of the FIM that is required in information entropy formulations.

In this chapter we revisit the problem of sensor placement for modal identification. We formulate the optimal experimental design problem based on the theory developed in Chapter 6, and especially on the asymptotic method of Section 6.4. The theory is applied to the case of optimizing the location of sensors in a real bridge. With the aid of the information contained in the prior PDF of the model parameters, the posterior covariance matrix is non-singular for non-uniform priors and the optimal sensor placement problem can be carried out also for the case where the number of sensors is less than the number of modes.

A novelty in this chapter is to study the effect of the Gaussian prior uncertainties on the optimal sensor design. For this, the theorem developed in Section 6.4.2 is used. The importance of spatial correlation in the prediction error is also pointed out as the means of avoiding sensor clustering phenomena for finite element models used to simulate civil infrastructures.

Theoretical developments are demonstrated by designing the optimal locations of a number of sensors for a 537m long concrete bridge using a dense finite element mesh of approximately 830000 degrees of freedom (DOF). The optimization is formulated in the physical continuous space of the design variables. This avoids the large discrete design space that arises from the extremely large number of possible nodal positions due to dense FE meshes. Multiple local optima are revealed that make the optimization problem very challenging. It is demonstrated that the computationally efficient heuristic FSSP algorithm provides accurate solutions when compared to stochastic algorithms such as the covariance matrix adaptation (CMA-ES) (Hansen et al., 2003) able to estimate the global optimum with substantially higher computational cost. Useful results are obtained which help guide experimentalists as to how the proposed method can be used for designing the optimal sensor locations in practical applications. In particular, we demonstrate that designs are also obtained for the very important case of number of sensors which are less than the number of modes by using the information contained in the prior distribution. In particular, this is important in designing the location of the reference sensors in a multiple sensor configuration set up experiment conducted with a limited number of reference and roving sensors in order to obtain the modal frequencies and reliably assemble the mode shapes from multiple setups. The effectiveness of the methodology is illustrated by designing the optimal location of two reference sensors (one transverse and one vertical) for the bridge. Finally, we draw attention to the sensor clustering issue and propose methods to avoid it and finally we show the effect of uncertainty in the prior distribution on the optimal design.

7.2 Bayesian optimal experimental design

The Bayesian optimal experimental design formulation was developed in detail in Chapter 6. Specifically, the asymptotic method developed in Section 6.4 is used. Since our model is linear w.r.t. the parameters to be identified from the experiment, the simplifications of Section 6.4.1 for linear models are used. Therefore, the objective function is the posterior entropy which depends only on the design variables. The posterior entropy for the special case of a linear model:

$$\underline{g}(\underline{\theta}, \underline{\delta}) = A(\underline{\delta})\underline{\theta} + \underline{b}(\underline{\delta}) \tag{7.1}$$

was given by Equations (6.58) and (6.59) in Section 6.4.1 which are repeated here for convenience:

$$U(\underline{\delta}) = H(p(\underline{\theta}|\underline{y}, \underline{\delta})) = \frac{1}{2}\log((2\pi e)^k |\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta}), \underline{\delta})}|) \tag{7.2}$$

where the posterior covariance matrix is independent of $\underline{\theta}$ and $\underline{y}$:

$$\Sigma_{p(\underline{\theta}|\underline{y}^i = \underline{g}(\underline{\theta}^i, \underline{\delta}), \underline{\delta})} = [Q_L(\underline{\delta}) + \Sigma_{p(\underline{\theta})}^{-1}]^{-1} \tag{7.3}$$

where $Q_L(\underline{\delta})$ is the FIM

$$Q_L(\underline{\delta}) = A(\underline{\delta})^T \Sigma_e(\underline{\delta})^{-1} A(\underline{\delta}) \tag{7.4}$$

Remind that since we have a linear model the asymptotic method is exact because all of its assumptions are true, as was proven in Section 6.4.1.

7.2.1 Optimal designs

The optimal experimental design problem is formulated as finding the values $\underline{\delta}_{opt}$ of the design variables $\underline{\delta}$ that minimize the posterior entropy $H(p(\underline{\theta}|\underline{y}, \underline{\delta}))$:

$$\underline{\delta}_{opt} = \arg\min_{\underline{\delta}} U(\underline{\delta}) = \arg\min_{\underline{\delta}} H(p(\underline{\theta}|\underline{y}, \underline{\delta})) \tag{7.5}$$

Optimal experimental design problems involving the design of the position of sensors often result in multiple local and global solutions. This will be also evident in the results presented in the application section. Also the gradient of the objective function with respect to the design variables in most cases of practical interest cannot be evaluated analytically. To avoid premature convergence to a local optimum and the evaluation of sensitivities of the utility function with respect to the design variables, stochastic optimization algorithms can be used to find the optimum. Herein the CMA-ES algorithm (Hansen et al., 2003) is used for solving the minimization problem, requiring only evaluation of the objective function at different values of the design variables. For this, the problem is formulated as a continuous optimization problem where the design variables are related to the coordinates of the sensors along the physical domain of the structure. To account, however, of curved and disconnected one-dimensional or two-dimensional domains, as well as to take into account the different types of sensors (vertical or transverse) that can be placed along the curved domain, a mapping technique can be used to map the physical design space

into a regular one-dimensional parent domain. The optimization is then conveniently carried out in the parent domain. The method will be illustrated for one-dimensional curved domains in the application section.

Heuristic algorithms have also been proposed to provide sub-optimal solutions. Notable are the forward and backward sequential sensor placement algorithms (FSSP and BSSP) in which the optimization is carried out sequentially. In the FSSP algorithm (Papadimitriou, 2004) the optimization is carried out sequentially for a single sensor, say the $i+1$ sensor, given that i sensors have already been placed in their optimal locations. This procedure is repeated for $i = 1, \ldots, N_0$, where N_0 is the total number of sensors to be placed in the structure. The heuristic algorithms have been shown to be effective and provided for several problems near optimal solutions (Papadimitriou and Lombaert, 2012). The FSSP algorithm is used here for investigating its effectiveness for the optimal sensor placement problem for modal identification.

7.3 Optimal sensor placement for modal identification

7.3.1 Formulation for modal identification

The optimal experimental design methodology is next implemented in structural dynamics for optimally placing the sensors in the structure for modal identification. Considering a linear finite element model of a structure, the equations of motion are given by

$$M\underline{\ddot{u}} + \tilde{C}\underline{\dot{u}} + K\underline{u} = \underline{f} \tag{7.6}$$

where M, $\tilde{C}$ and $K \in R^{n\times n}$ are the mass, damping and stiffness matrices, respectively, $\underline{u} \in R^n$ is the displacement vector and $\underline{f}$ is the forcing vector. Using modal analysis and assuming classically damped modes, the response displacement and acceleration vectors are given by $\underline{u} = \Phi\underline{\xi}$ and $\underline{\ddot{u}} = \Phi\underline{\ddot{\xi}}$, respectively, where $\Phi = [\underline{\phi}_1, \ldots, \underline{\phi}_m] \in R^{n\times m}$ is the matrix of mode shapes involving m contributing modes ($m \leq n$) that can be obtained by solving the eigenvalue problem $K\Phi = \Lambda M\Phi$, Λ is the diagonal matrix of eigenvalues, $\underline{\xi} = [\xi_1, \ldots, \xi_m]^T \in R^m$ is the vector of modal coordinates satisfying

$$\ddot{\xi}_r + 2\zeta_r\omega_r\dot{\xi}_r + \omega_r^2\xi_r = \underline{\phi}_r^T\underline{f} \tag{7.7}$$

$r = 1, \ldots, m$, and ζ_r is the modal damping ratio. The strain vector is given by a similar expression $\underline{\varepsilon} = E\underline{\xi}$, where E depends through the finite element modeling on the elements in the mode shape matrix Φ.

The problem of estimating the modal coordinate vector $\underline{\xi}$ or $\underline{\ddot{\xi}}$ using displacement/strain or acceleration measurements is investigated. The modal coordinates $\underline{\xi}$ or $\underline{\ddot{\xi}}$ contain the modal properties (modal frequencies, modal damping ratios, participation factors). The objective in modal identification is to place sensors (displacement, acceleration and/or strain sensors) so that the information contained in the measured data is sufficient to estimate the modal coordinate vectors$\underline{\xi}$ or $\underline{\ddot{\xi}}$, depending on the sensor type

used. Introducing the parameter set $\underline{\theta}$ to be either $\underline{\xi}$ or $\underline{\ddot{\xi}}$ and denoting by $\underline{g}(\underline{\theta};\underline{\delta}) \in R^{N_0}$ the response quantity that is measured by the sensors, one has the following equation between the modal model predictions and the parameter set $\underline{\theta}$

$$\underline{g}(\underline{\theta};\underline{\delta}) = \Phi(\underline{\delta})\underline{\theta} = L(\underline{\delta})\Phi\underline{\theta} \tag{7.8}$$

where the matrix $L(\underline{\delta}) \in R^{N_0 \times n}$ is the observation matrix and maps the n model DOF to the N_0 measured positions. The matrix $L(\underline{\delta})$ depends on the location vector $\underline{\delta}$ defining the locations of the sensors in the structure. If the measured positions coincide with the DOF of the model then the matrix $L(\underline{\delta})$ is comprised of zeros and ones. For the general case for which the measured locations do not coincide with the DOF of the FE model, the matrix $L(\underline{\delta})$ depends on the interpolation scheme used to obtain the response within a finite element in terms of the finite element nodal responses. The above formulation allows sensors to be placed in any point in the structure, not only at nodal points. Also it gives the flexibility to covert the optimization problem for estimating the design variables to a continuous optimization problem over the physical domain of the structure. For strain measurements the aforementioned analysis is the same provided that the mode shape matrix Φ is replaced by the matrix E.

The model equation in (7.8) is the same as the one used in (7.1) with $A(\underline{\delta}) = L(\underline{\delta})\Phi$ and $\underline{b}(\underline{\delta}) = \underline{0}$ so the objective function of (7.2) and (7.3) applies with the FIM of (7.4) given by

$$Q_L(\underline{\delta}) = \{L(\underline{\delta})\Phi\}^T \Sigma_e(\underline{\delta})^{-1} \{L(\underline{\delta})\Phi\} \tag{7.9}$$

in terms of the mode shape components at the measured locations.

Based on the form (7.9), and using the dimensions of the matrices $[L(\underline{\delta})\Phi]^T \in R^{m \times N_0}$, $\Sigma^{-1}(\underline{\delta};\underline{\sigma}) \in R^{N_0 \times N_0}$ and $L(\underline{\delta})\Phi \in R^{N_0 \times m}$, the FIM is a non-singular matrix product only if the number of sensors N_0 is at least equal to the number of contributing modes m ($N_0 \geq m$). For $N_0 < m$, the matrix product is by construction singular and for uniform prior PDF the second term of the combined matrix of (7.3) is zero so the determinant of the posterior covariance matrix will be zero for any sensor configuration. Thus, for $N_0 < m$ the optimal sensor location problem cannot be performed for uniform prior PDF. This means that the information content in the measured data and the prior is not sufficient to estimate all the parameters simultaneously.

The Bayesian optimal experimental design formulation yields a nonsingular posterior covariance matrix for $N_0 < m$ only if the prior is a non-uniform distribution. The non-uniform prior, say Gaussian, yields a Hessian matrix $\Sigma^{-1}_{p(\underline{\theta})}$ (prior covariance) that is added to the FIM $Q_L(\underline{\delta})$ and makes the combined matrix $Q_L(\underline{\delta}) + \Sigma^{-1}_{p(\underline{\theta})}$ non-singular. The FIM in (7.9) has exactly the same form as the one proposed by Kammer (1991) for designing the optimal sensor locations using the EFI algorithm. The difference of the present Bayesian formulation with the EFI algorithm is in the use of the prior information for the model parameters which permits the design of optimal sensor locations for the case of number of sensors which is less than the number of modes. The contribution from the prior is the result of the application of the Bayesian optimal experimental design proposed herein based on minimizing the expected posterior entropy. Alternatively, Yuen and Kuok (2015) have also proposed a non-uniform prior on the information

entropy measure in order to solve this unidentifiability problem. For uniform prior and unidentifiable case, Papadimitriou and Lombaert (2012) proposed the sum of the log of the nonzero eigenvalues in the FIM to be maximized instead of the sum of the log of all eigenvalues. Herein these results are generalized to incorporate prior uncertainty in the parameter estimates. This procedure allows to systematically place the sensors optimally in the structure even for the unidentifiable case that arises for a small number of sensors.

7.3.2 Effect of prior uncertainty

Next a diagonal covariance $S = diag(s_1^2, \ldots, s_m^2)$ for the Gaussian prior is assumed and the effect of the values of the variances is examined. These variances control the prior uncertainty in the values of the model parameters. A theoretical result is provided that shows the effect of the assigned prior uncertainties on the optimal design.

For simplicity, the case of optimally placing one sensor is considered first and then generalized for the multiple sensor case. In the case of a single sensor, the FIM reduces to the form $Q_L(\underline{\delta}) = \sigma^{-2}\underline{\varphi}_i\underline{\varphi}_i^T$, where $\underline{\varphi}_i = [\Phi_{i1}, \Phi_{i2}, \ldots, \Phi_{im}]^T$ is a vector of dimension m that consists of the values of each mode shape at the sensor location denoted here as i. Note that for one sensor the prediction error covariance matrix $\Sigma_e(\underline{\delta})$ is scalar with $\Sigma_e(\underline{\delta}) = \sigma^2$. Using the following known result for a square matrix B and two vectors $\underline{u}$ and $\underline{v}$

$$\det(\underline{u}\underline{v}^T + B) = (1 + \underline{v}^T B^{-1}\underline{u}) \det(B) \tag{7.10}$$

then $\det[Q] \equiv \det[Q_L(\underline{\delta}) + \Sigma_{p(\underline{\theta})}^{-1}]$ in (7.3) takes the form

$$\det[Q] = \det[\sigma^{-2}\underline{\varphi}_i\underline{\varphi}_i^T + \Sigma_{p(\underline{\theta})}^{-1}] = [1 + \sigma^{-2}\underline{\varphi}_i^T\Sigma_{p(\underline{\theta})}\underline{\varphi}_i] \det(\Sigma_{p(\underline{\theta})}^{-1}) = [1 + \sigma^{-2}\sum_{k=1}^{m} s_k^2\Phi_{ik}^2] \det(\Sigma_{p(\underline{\theta})}^{-1}) \tag{7.11}$$

Note that the variance s_k^2 of the k-th parameter (modal coordinate) weights the contribution in the sum of the value of the k-th mode shape at the sensor location. The higher the value of the variance s_k^2, the higher the contribution of the k-th mode shape on the determinant. So it is evident that the optimal design will give preference to the modes that have higher prior uncertainty, i.e. higher s_k^2 values.

Next we complete the proof for the general case of arbitrary number of N_0 sensors. For this, we use the following known result

$$\det(UWV^T + B) = \det(W^{-1} + V^T B^{-1} U) \det(W) \det(B) \tag{7.12}$$

Setting for simplicity $\Psi = L(\underline{\delta})\Phi$, then $\det[Q] = \det[Q_L(\underline{\delta}) + \Sigma_{p(\underline{\theta})}^{-1}]$ takes the form

$$\begin{aligned}
\det[Q] &= \det[\Psi^T\Sigma_e(\underline{\delta})^{-1}\Psi + \Sigma_{p(\underline{\theta})}^{-1}] \\
&= \det[\Sigma_e(\underline{\delta}) + \Psi\Sigma_{p(\underline{\theta})}\Psi^T] \det[\Sigma_{p(\underline{\theta})}^{-1}] \det[\Sigma_e(\underline{\delta})^{-1}] \\
&= \det\left[\Sigma_e(\underline{\delta}) + \textstyle\sum_{k=1}^{m} s_k^2\psi_k(\underline{\delta})\psi_k^T(\underline{\delta})\right] \det[\Sigma_{p(\underline{\theta})}^{-1}] \det[\Sigma_e(\underline{\delta})^{-1}] \\
&= \det\left[\Sigma_e(\underline{\delta}) + \textstyle\sum_{k=1}^{m} s_k^2||\psi_k(\underline{\delta})||^2\tilde{\psi}_k(\underline{\delta})\tilde{\psi}_k^T(\underline{\delta})\right] \det[\Sigma_{p(\underline{\theta})}^{-1}] \det[\Sigma_e(\underline{\delta})^{-1}]
\end{aligned} \tag{7.13}$$

where $\psi_k(\underline{\delta})$ is the k-th mode shape evaluated at the sensor locations, $\tilde{\psi}_k(\underline{\delta})$ is the unit-normalized mode shape, $||\psi_k(\underline{\delta})||^2$ is the Euclidean norm square of the mode shape.

Equation (7.11) is a special case of equation (7.13) when only one sensor is used and Ψ becomes a row vector and its columns become scalars. In the multiple sensor case we are dealing with a sum of rank-one matrices over the model parameters, where each rank-one matrix $\psi_k(\underline{\delta})\psi_k^T(\underline{\delta})$ is formed from the k-th mode shape $\psi_k(\underline{\delta})$, weighted by the corresponding prior variance s_k^2 of that mode. So we see again that, similarly to the one sensor case, the Gaussian prior variance s_k^2 acts as weighting factor, this time by giving greater weight to the matrix $\psi_k(\underline{\delta})\psi_k^T(\underline{\delta})$ related to the k-th mode. The variance s_k^2 of the k-th mode weights the contribution in the sum of the value of the k-th rank-one matrix formed from the k-th mode shape that is evaluated at the sensor locations. The higher the value of the variance s_k^2, the higher the contribution of the rank-one matrix of the k-th mode shape on the $\det[Q]$. So the optimal design of sensor locations again gives preference to the modes that are assigned by a user to have higher prior uncertainty.

It should be noted that the idea of favoring one or more modes in the design of the optimal sensor configuration by appropriately selecting the prior variances can be useful in model updating and damage detection applications. Specifically, in damage detection and localization the proposed optimal sensor placement design can be used to increase the information about the damage location and size contained in the measured data by favoring, for example, local modes known to be more sensitive to local damage. To avoid reducing the robustness of the proposed optimal sensor placement algorithm for identifying the least favored modes, one may choose to optimally allocate a fraction of the available sensors in an effort to favor a small number of the contributing modes, ensuring that the optimal placement of the rest of the sensors maintain adequate levels of robustness.

7.4 Application

The proposed methodology is used to optimize the location of acceleration sensors placed at the deck level of the Metsovo bridge. Numerical results are shown for up to 30 sensors for the spatially uncorrelated and spatially correlated prediction error models. The effect of the spatial correlation as well as the effect of the selection of the uncertainty in the prior distribution on the optimal designs is also investigated. Finally, the optimal design of a small number of reference sensors to be used as common sensors in multiple configuration setups is illustrated.

7.4.1 Bridge description and FE model

The Metsovo bridge of Egnatia Motorway, schematically shown in Figure 7.1, is crossing the deep ravine of Metsovitikos river, 150m over the riverbed. This is the highest bridge of the Egnatia Motorway, with the height of the tallest pier equal to 110m. The total length of the bridge is 537m. The bridge has 4 spans of length 44.78m, 117.87m, 235m, 140m and 3 piers of which M1(45m) supports the boxbeam superstructure

through pot bearings (movable in both horizontal directions), while M2(110m) and M3(35m) piers connect monolithically to the structure.

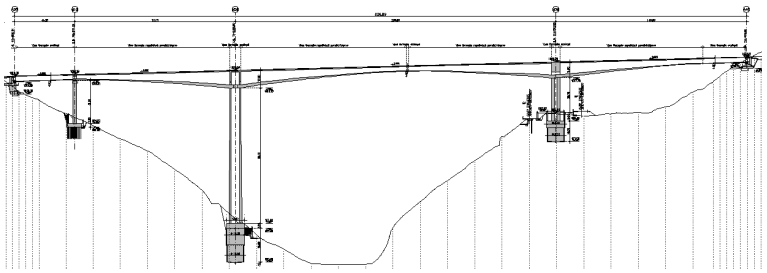


Figure 7.1: Longitudinal view of the Metsovo bridge.

A detailed FE model of the bridge is created using 3-dimensional tetrahedron quadratic Lagrange finite elements. The mesh is chosen to accurately predict the lowest twenty modal frequencies and mode shapes of the bridge. The model also considers the interaction with soil by modeling the soil stiffness with solid blocks surrounding the piers and abutments. The FE model has 830,000 DOF. The size of the elements in the mesh is controlled by the thickness of the deck and piers box-like cross-sections. The typical element length is of the order of the thickness of the deck cross-section. The complex 3D geometry of the bridge was designed in SolidWorks and the FE model was created and solved in COMSOL Multiphysics by importing the 3D geometry from SolidWorks. The modulus of elasticity for the concrete were taken from design considerations to be 37e9 Pa and 34e9 Pa for the deck and the piers, respectively. The modulus of elasticity for the solid blocks was taken to be 9e9 Pa. The stiffness and mass matrices of the finite element model are used to obtain the mass normalized mode shapes to be used in the optimal sensor placement methodology.

7.4.2 Optimal sensor placement

The purpose of the OED is to optimally locate a number of acceleration sensors such that one gets the most informative data for identifying the lowest 10 modes. Using the FE model, the lowest 10 mode shapes needed in the design of the optimal sensor locations were computed and stored. These modes consist of 5 transverse modes, with mode shapes that deform the bridge and deck in the transverse direction, and 5 vertical modes mode, with mode shapes that bend the deck in the vertical direction. The transverse mode numbers are $\{1, 2, 4, 6, 8\}$, while the bending mode number are $\{3, 5, 7, 9, 10\}$. The optimal locations of sensors are obtained by minimizing the posterior entropy. Herein, results will be presented in terms of the information entropy measure. It should be noted that in our case the number of parameters (modal coordinates) is $N_{\underline{\theta}} = m = 10$.

Continuous optimization in parent domain

Due to the fact that measurements on the bridge deck have to be taken without traffic interruptions, the sensors in the structure are only allowed to be placed along the pedestrian sidewalks (left and right). For demonstration purposes, in this study it will be assumed that sensors are placed along the one sidewalk marked in Figure 7.2 with red line. The line along the sidewalk over which sensors can be placed is a curved one and in order to perform the optimization problem along the curved line we develop a mapping of the curved line in a much simpler parent domain of a straight line. Each point in the physical curved line is mapped to a point in the straight parent line. To introduce such a mapping one can follow concepts developed in finite element analysis to map an arbitrary one-dimensional element in space in a parent element of specified length. It should be noted that the curved one-dimensional domains (red line) in Figure 7.2 consists of the straight elements that are the vertices of the neighbor tetrahedral finite elements used to model the bridge. So these straight vertices are mapped in the sequence that are encountered from left to the right edge of the bridge to a parent element so that the left corner of the red curve is mapped to the parent location 0, while the right corner of the red curve is mapped to the parent location 1. Specifically, the 0m to 537m curved line along the deck is mapped to the parent domain from 0 to 1.

To account for the different type of sensors that can be placed on the physical structures, in our case sensors measuring in the vertical and the transverse directions, the physical curved line with transverse sensors placed on it is mapped to a parent element from 0 to 1, while the physical curved line with vertical sensors placed on it is mapped to a parent element from 1 to 2. In this way one can efficiently handle the two types of transverse and vertical sensors without the need to distinguish them during the optimization. The optimization is performed in the parent domain with the design variables, indicating the location of the sensors in the parent domain, to vary from 0 to 2. A design value inside the parent domain [0,1] corresponds to a transverse sensor which is mapped through the aforementioned mapping to a point on the physical domain (red curve) in Figure 7.2. Similarly, a design value inside the parent domain [1,2] corresponds to a vertical sensor which is mapped to a point on the physical domain (red curve) in Figure 7.2.

The mode shape value at a point in the parent domain is obtained by the mode shape value at the corresponding point in the physical domain. Consider the mapping of points from 0 to 1 in the parent domain to the physical domain. Since the mapped point in the physical domain might not correspond to a node of the finite element model, the mode shape value is obtained by interpolation using the mode shape values at the two neighboring nodes of that point in the transverse direction. A similar procedure is used for points between 1 and 2 in the parent domain to find the mode shape component in the vertical direction.

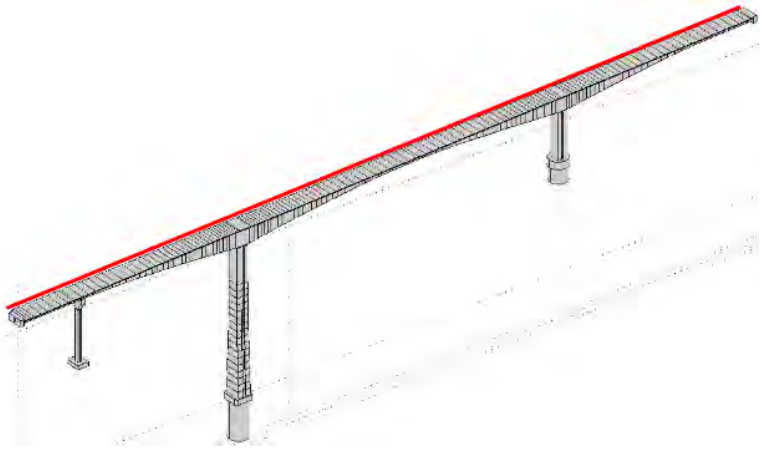


Figure 7.2: Optimization along pedestrian walkway of the deck.

Using this mapping strategy the optimization is performed in a more sophisticated way than what was used in the past. Instead of performing multiple optimizations for one sensor at a time which is predetermined to be either a transverse or vertical sensor, now it is allowed to optimize for several sensors simultaneously on the bridge pedestrian sideway, without predetermining the types of the sensors. Instead, it is left for the optimization to decide which of the sensors will be transverse and which will be vertical, depending on what region they lie in the parent domain.

So the optimization proceeds as follows. The design vector $\underline{\delta}$ is a vector with as many elements as the number of uniaxial sensors. Each element of this vector describes both the location and the type (transverse or vertical) of a sensor. Each element of the design vector $\underline{\delta}$ is a number on the interval [0 2]. This effectively allows for the optimization of many sensors simultaneously, which includes making the decision about the types of sensors as well along with their location. This is more realistic and gives more freedom to the approach than predetermining how many sensors of each type one should include in $\underline{\delta}$.

Note that the optimal sensor placement methodology can also be readily applied to design a set of triaxial instead of uniaxial sensors. In this case the information from each tri-axial sensor will be the combined information obtained from the three sensor components in each direction. The aforementioned concept can also be extended to design a combination of uniaxial (e.g. transverse and vertical) and triaxial sensors by mapping the physical curved line with triaxial sensors placed on it to a parent element from 2 to 3. In this way the three type of sensors can efficiently be handled. However, the triaxial sensors will always be preferred to the transverse or vertical uniaxial sensors since the information gain from a uniaxial sensor placed in an optimal position will always be less than the information gain from a triaxial sensor placed at the same position. To favor a uniaxial sensor in relation to a triaxial sensor one has to introduce additional constrains that penalize the placement of triaxial sensors by taking into account the extra cost of the sensor and the information gain a triaxial sensor will provide in relation to a uniaxial

sensor. However, this complicates the optimization, with the subject falling outside the scope of the present work. The previous concepts can be extended to design a combination of different uniaxial and triaxial sensor types such as acceleration, displacement and strain sensors.

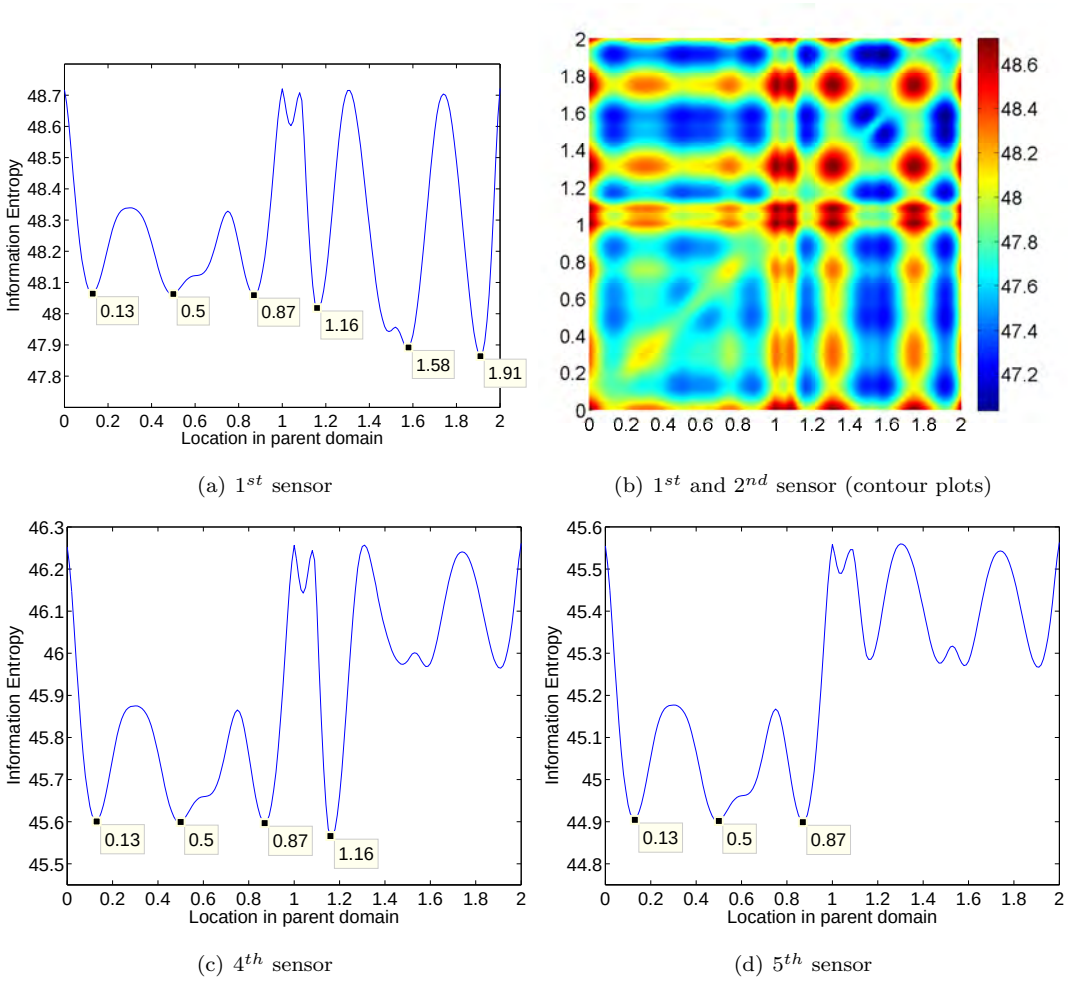
Numerical results for spatially uncorrelated prediction error model

An uncorrelated prediction error model is used with diagonal covariance matrix $\Sigma(\underline{\delta};\underline{\sigma}) = \sigma^2 I_{N_0}$, where $I_{N_0} \in R^{N_0 \times N_0}$ is the identity matrix, with the value of the single prediction error parameter chosen to be $\sigma = 0.01$. A Gaussian prior with relatively large uncertainties is used. As a result, the posterior covariance matrix is non-singular and the design can proceed for any number of sensors. The covariance matrix of the Gaussian prior is set to $S = 10^3 I_m$, with $s_1^2 = \ldots = s_m^2 = 10^3$.

Results for up to $N_0 = 9$ sensors are first obtained. Note that in this case the number of sensors is less than the number of modes ($m = 10$) and so the FIM is singular. The FSSP method is used to obtain the optimal sensor locations. It should be noted, however, that the sensor locations obtained by the heuristic FSSP method are verified that are the global optima, also obtained by the CMA-ES global optimizer (Hansen et al., 2003).

The optimal location of the first sensor corresponds to the minimum of the information entropy. The information entropy as a function of the location of the sensor is drawn in Figure 7.3(a) at the parent domain. It can be seen that there are 9 local minima. Figure 7.3(b) shows the contour plots of the information entropy as a function of two sensor locations in the parent domain. A large number of local optima is also observed (blue color). The number of optima are expected to increase as the dimension of the design space increases. Figures 7.3(a) and 7.3(b) confirm that there is a large number of local optima and so gradient-based optimization methods will be trapped to a local optimum and will not be able to obtain a global optimum. Stochastic algorithms such as CMA-ES have much higher chances to pinpoint the global optimum.

For the case of optimizing the location of the first sensor, a vertical sensor was obtained rather than a transverse one. The two minima at 1.58 and 1.91 in the parent domain in Figure 7.3(a) are smaller than all the rest, which implies that for the 1^{st} sensor it is best to be a vertical sensor in one of these two locations which correspond to distances on the bridge of 310m and 487m, respectively, from the left end. The minimum at 1.91 (487m) is slightly smaller, but the difference is negligible for practical purposes.



(a) 1^{st} sensor

(b) 1^{st} and 2^{nd} sensor (contour plots)

(c) 4^{th} sensor

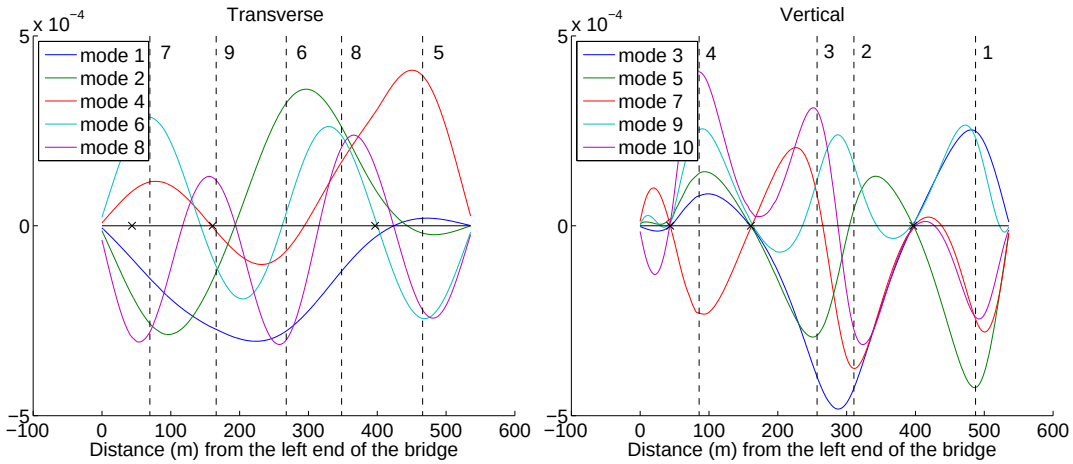
(d) 5^{th} sensor

Figure 7.3: Information entropy vs the location of a sensor in parent domain

Another interesting observation in Figure 7.3(a) is that at the two ends of the bridge (at points 0, 1 and 2 in the parent domain) the value of the posterior entropy is 48.72, which is exactly the information entropy of the Gaussian prior. This makes sense, because in these points the mode shapes have zero components and the FIM is zero, and therefore the only contribution to the posterior comes from the prior. There are also other points which result in the posterior entropy being almost equal to the prior entropy, which correspond to vertical sensors being placed in points of zero vertical response (locations 1.08, 1.3 and 1.74 in the parent domain) for all vertical mode shapes. These are points in the deck that are above the piers where the vertical motion of the bridge is almost restrained. As one would intuitively expect, the information gain from these designs (which is the difference between the prior and posterior entropies) is zero.

The information entropy as a function of the location of the fourth sensor in the parent domain given that the first three sensors have been placed in their optimal position is shown in Figure 7.3(c). A similar

plot for the information entropy as a function of the position of the 5^{th} sensor given that the first four sensors have been placed in their optimal position, is shown in Figure 7.3(d). The optimal locations of the first nine sensors are shown in Figure 7.4(a) and 7.4(b) for both the transverse and vertical sensors. Also, information of the optimal locations of sensors and the corresponding information entropy values is given in Table 7.1.



(a) Transverse sensors

(b) Vertical sensors

Figure 7.4: Optimal locations of transverse and vertical sensors (Case of 9 sensors with uncorrelated prediction error). The symbol x denotes the location of the piers along the deck.

Table 7.1: Optimal location in the physical domain and minimum information entropy for the first nine sensors ($N_0 < m$). V=Vertical, T=Transverse.

Sensor	1	2	3	4	5	6	7	8	9
Type	V	V	V	V	T	T	T	T	T
Optimal Location (m)	487	310	257	85	466	267	69	348	166
Information Entropy	47.86	47.03	46.26	45.57	44.89	44.23	43.58	43.02	42.63

Note that the optimal location of the first two sensors is made up of the location of the global and the next local optimum predicted in Figure 7.3(a). Also, from the results in Figure 7.4(b) the first 4 sensors are vertical sensors. This is because, as can be seen from the Figures 7.4(a) and 7.4(b), vertical modes have slightly higher displacements and therefore contribute more than the transverse modes to the FIM, and therefore to the posterior covariance matrix. Also the vertical modes have more "convenient" points where several modes have high responses compared to the transverse modes. These result in preferring the vertical modes compared to the transverse modes. Note that all modes are treated equally by selecting the prior uncertainty to be the same for all modal coordinates. Finally, results in Figure 7.4(a) suggest

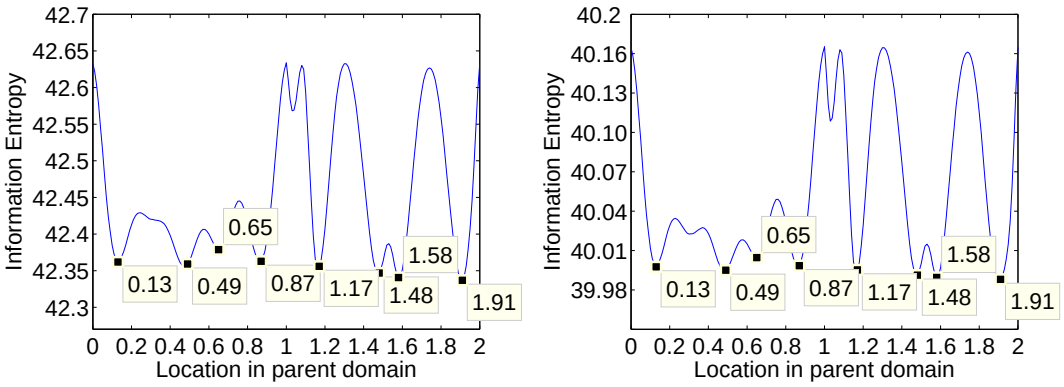
that the last 5 sensors (5^{th} to 9^{th} sensors) are selected to measure in the transverse direction.

From the results in Figure 7.3(c) of the information entropy as a function of the location of the 4^{th} sensor, given that the first 3 sensors are placed at their optimal location, it is noticed that the vertical sensor at 85m is slightly preferred from the three transverse sensors at 69m, 267m and 466m since the minimum value of 45.57 in the information entropy for the vertical sensor is slightly smaller than the local minimum value of 45.58 for the information entropy for the three transverse sensors. From the results in Figure 7.3(d) of the information entropy as a function of the location of the 5^{th} sensor, given that the first 4 sensors are placed at their optimal location, it can be observed that any of the locations 69, 267 and 466m for a transverse sensor are candidates. These sensor locations are the optimal sensor locations for the 5^{th}, 6^{th} and 7^{th} sensor. The optimal sensors for the 8^{th} and 9^{th} sensor are transverse ones, with optimal locations shown in Figure 7.4(a) and Table 7.1.

Observing the optimal sensor locations in relation to the mode shapes drawn in Figures 7.4(a) and 7.4(b), one should note that the results are reasonable for placing sensors in the suggested vertical or transverse locations since in these locations the mode shape components correspond, in general, to their higher values.

For up to nine sensors it is observed that placing two or more sensors in the same position corresponds to the worst sensor location. This is confirmed also by the contour plots in Figure 7.3(b) where the placing of the two sensors in the same position is not preferred (red colors). Also it is confirmed by the plots in Figure 7.3 where it is clear that when a new sensor location coincides with an already placed sensor location it gives large values of the information entropy as compared to the optimal one. So in the singular FIM case with $N_0 < m$, the problem of sensor clustering due to uncorrelated prediction error does not occur.

The optimal location and type (transverse or vertical) of the next 11 sensors (10^{th} to 20^{th}) is also considered. The design is performed using the FSSP algorithm. However, for selected number of sensors, the accuracy of the results obtained from the FSSP algorithm is confirmed by running also the CMA-ES algorithm. Both algorithms provide the same estimates. The information entropy as a function of the location of the 10^{th} sensor given that the first 9 sensors are placed at their optimal positions (shown in Table 7.1) is presented in Figure 7.5(a). Figure 7.5(b) gives similar information but for optimizing the location of the 19^{th} sensor. The optimal sensor locations for the sensors from 10^{th} to 20^{th}, their type (vertical or transverse) and the minimum information entropy are also given in Table 7.2 and in Figure 7.6.



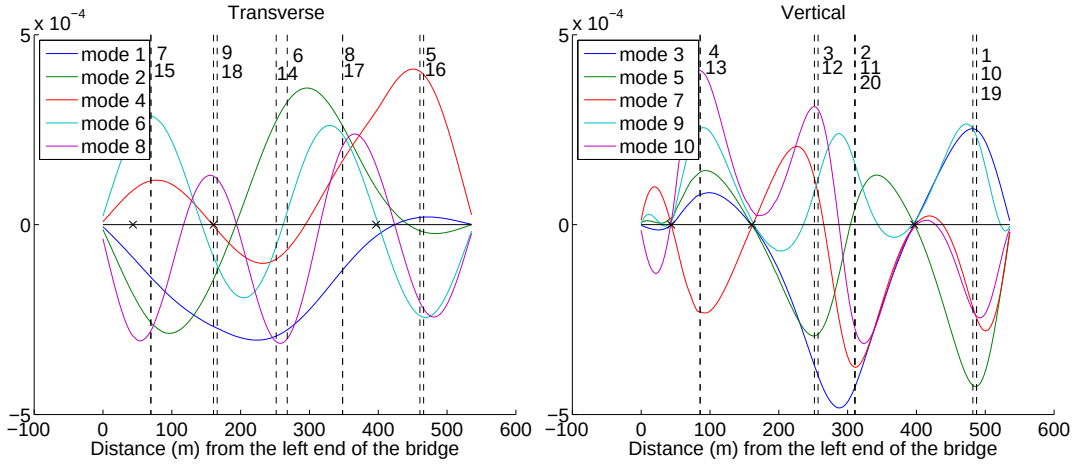
(a) 10^{th} sensor (b) 19^{th} sensor

Figure 7.5: Information entropy vs the location of a sensor in parent domain.

Table 7.2: Optimal location in the physical domain and minimum information entropy for the 10^{th} up to the 20^{th} sensor ($N_0 \geq m$). V=Vertical, T=Transverse, OL=Optimal Location (m), IE=Information Entropy.

Sensor	10	11	12	13	14	15	16	17	18	19	20
Type	V	V	V	V	T	T	T	T	T	V	V
OL (m)	487	310	251	85	251	69	460	348	160	482	310
IE	42.33	42.04	41.75	41.47	41.19	40.91	40.64	40.39	40.17	39.98	39.80

It is clear from the Figure 7.5(a) that the optimal location of the 10^{th} sensor coincides with the location of the 1^{st} sensor. Actually the information entropy as function of the location of the 10^{th} sensor is qualitatively similar to the information entropy for one sensor in Figure 7.3(a) as a function of its location. Comparing Figures 7.5(a) and 7.3(a), the local/global optimal appear at the same locations which suggests that the optimal locations of the next nine sensors will be close to the optimal locations of the first nine sensors. This is confirmed by comparing also the results in Table 7.2 with the results in Table 7.1. Comparisons clearly demonstrate (see also Figure 7.6) that the optimal locations of the sensors 9^{th} to 20^{th} coincide or they are very close to the optimal locations estimated for the first 9 sensors. This sensor clustering is due to the incorrect assumption of the spatially uncorrelated prediction errors (Papadimitriou and Lombaert, 2012). However, it seems that the sensor design obtained with uncorrelated prediction errors are very reasonable and intuitive for the first nine sensors which correspond to the case of singular FIM with $N_0 < m$. To correct the problem of sensor clustering for more than 9 sensors one has to introduce spatial correlation in the prediction errors.



(a) Transverse sensors

(b) Vertical sensors

Figure 7.6: Optimal locations of transverse and vertical sensors (Case of 20 sensors with uncorrelated prediction error). The symbol x denotes the location of the piers along the deck.

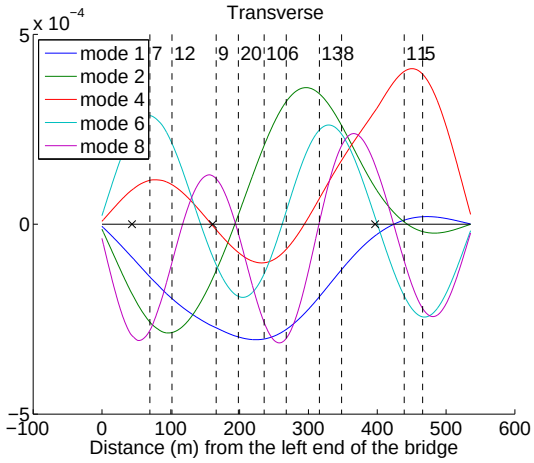
Numerical results for spatially correlated prediction error model

A spatially correlated prediction error model is assumed next. For this the covariance $\Sigma(\underline{\delta};\underline{\sigma})$ of the prediction error at the sensor locations is selected to be non-diagonal with the (i,j) element of the form $\Sigma_{ij}(\underline{\delta};\underline{\sigma}) = \sigma^2 R(\delta_i - \delta_j)$, where $R(\delta_i - \delta_j)$ is the spatial correlation structure of the prediction error, δ_i and δ_j are the locations of the i and j sensors, and σ^2 is the strength of the prediction error. An exponentially decaying correlation structure of the form $R(\delta_i - \delta_j) = \exp(-|\delta_i - \delta_j|/\lambda)$ is selected, where λ is the correlation length.

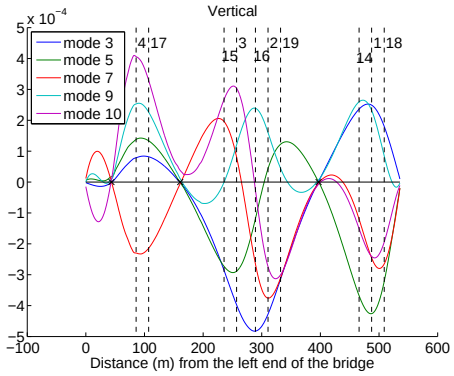
Since in this formulation we allow for each sensor to be either a transverse or vertical sensor, correlation is limited to sensors of the same type. That is, correlation exists between any two transverse sensors or any two vertical sensors, but not between a transverse and a vertical sensor. The described model allows for two different correlation length parameters (or correlation functions in general) to be used for the transverse and vertical sensors, respectively. In the numerical results that follow, the correlation parameters are chosen to be $\sigma = 0.01$ and $\lambda_{trans} = \lambda_{vert} = 10m$ or $\lambda_{trans} = \lambda_{vert} = 20m$.

Optimal sensor placement results for the correlated prediction error models are shown in Figure 7.7 for two different correlation length of 10m and 20m respectively. The Gaussian prior is selected to be the same as in the uncorrelated prediction error case. Results have been derived using the FSSP algorithm and their accuracy has been also confirmed for representative sensor cases using the CMA-ES algorithm. Comparing with the results of the uncorrelated case in Figure 7.6 it can be observed that the optimal sensor locations for the first 9 sensors are the same as the uncorrelated prediction error case. For more than 9 sensors, the sensor clustering problem is not present in the spatially correlated prediction error case. The optimal locations of sensors 10^{th} to 20^{th} are not close to the locations obtained for the first 9

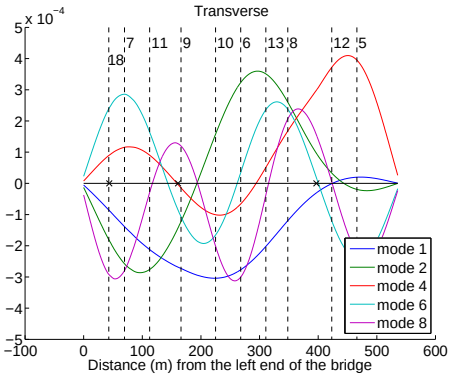
sensors. In fact, they are more uniformly distributed in some of the areas of the bridge deck. Comparing Figures 7.7(a,b) with Figures 7.7(c,d) it can be observed that the spacing of the sensors tends to increase as one increases the correlation length from 10m to 20m. This is consistent with the theoretical results obtained in Papadimitriou and Lombaert (2012).



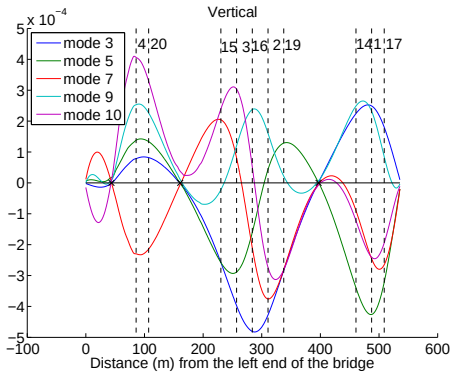
(a) Transverse sensors ($\lambda = 10m$)



(b) Vertical sensors ($\lambda = 10m$)



(c) Transverse sensors ($\lambda = 20m$)



(d) Vertical sensors ($\lambda = 20m$)

Figure 7.7: Optimal sensor locations for the case of 20 sensors with spatially correlated prediction error. The symbol x denotes the location of the piers along the deck.

In Figure 7.8 the optimal information entropy as a function of the number of sensors is shown for the uncorrelated and correlated prediction error cases and for 1 up to 30 sensors. The 0 sensor case corresponds to the case where no sensors are placed and so the information entropy is that of the Gaussian prior PDF. As expected, we notice that the optimal value of the information entropy for a given number of sensors decreases as the number of sensors increases. For up to 9 sensors the entropies are identical between the uncorrelated and correlated prediction error models. From the 10^{th} sensor and on, the correlated prediction error models lead to more information entropy (less information) and this entropy

increases with the correlation length. This is due to the fact that the information provided by neighbor sensors within the correlation length assumed is not significantly different and so this results in a drop of the total information for the same number of sensors. Also, the curves with high correlation suggest that the information provided by adding sensors in the structure is decreasing and eventually after a number of sensors there is no significant information offered by additional sensors. This plot can be used to decide on the number of sensors to be placed in the structure, given the correlation length. The uncorrelated prediction error models provide misleading results since continuing adding sensors in the structure has the effect of gaining additional information, independent of the number of added sensors, which is counter-intuitive. Finally, the CMA-ES algorithm is also used to design the optimal sensor locations and the resulting information entropy values for representative sensor cases, shown in Figure 7.8 for correlation length 20m, match exactly the information entropy values obtained using the FSSP algorithm, confirming in this case the accuracy of FSSP algorithm.

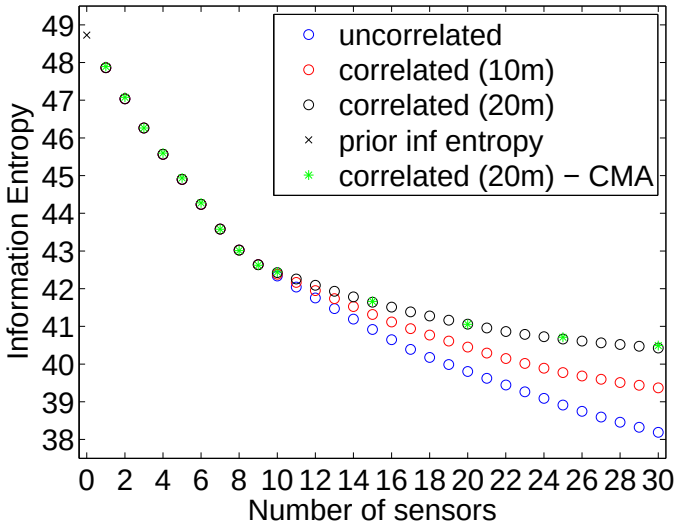


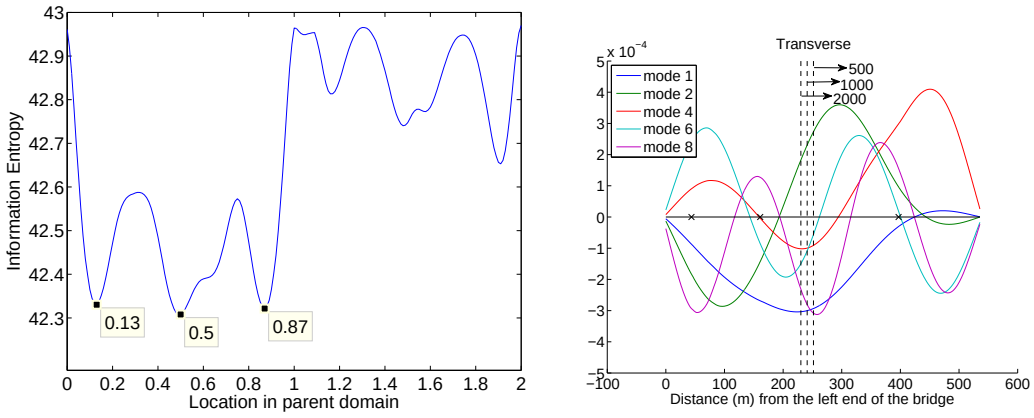
Figure 7.8: Optimal information entropy vs number of sensors.

7.4.3 Effect of prior uncertainty on optimal sensor placement

The results obtained so far correspond to an isotropic Gaussian prior where all prior uncertainties in the parameters were selected to be the same. This gives equal weight to all parameters as far as the prior is concerned, so there is no preference of a specific mode over another. The Gaussian prior covariance matrix is a modeling choice that depends on user preference. Therefore, it can be fully manipulated according to the needs. In Section 6.4.2 it was proved that by giving larger prior uncertainties to some specific parameters (modes in this case) we are essentially giving more weight in these modes in the selection of the optimal design. The insightful result of Section 6.4.2 states that a Gaussian prior can be used

as a means to perform more sophisticated OED, where we give preference to some selected modes over others. Different Gaussian prior variances for the different modes get transferred to the posterior and result in different optimal designs, favoring the identification of modes with the largest prior variances. The Bayesian framework for OED provides the means to fully quantify this preference of some modes (or parameters for identification in general) over others through the prior.

In order to illustrate this, the simple case of one sensor is examined. We would like to give more weight to the identification of the modal coordinates of the transverse modes. For this we lower the variances of the bending modes from 1000 to 100, and keep the variances of the transverse modes at 1000. The information entropy as a function of the position of the 1^{st} sensor is shown in Figure 7.9(a). We see that the optimal design for the 1^{st} sensor is now a transverse sensor in one of the already found optimal locations for transverse sensors, shown in Figure 7.3(d). In fact, the plot for one sensor in Figure 7.9(a) is now qualitatively very similar to the plot of the 5^{th} sensor shown in Figure 7.3(d), where the first transverse sensor appears as optimal. With this change in the prior variances giving weight to the transverse modes, the optimal designs have a preference for the transverse sensors now to show up as the 1^{st} sensor and not the 5^{th}.



(a) (b)

Figure 7.9: (a) Information entropy of 1^{st} sensor vs location in parent domain using Gaussian prior variance $s^2 = 1000$ for transverse modes and $s^2 = 100$ for vertical modes. (b) Optimal locations for 1^{st} sensor for prior variance of 1^{st} mode equal to 500, 1000 and 2000. All other prior variances equal to 100.

Consider next the case of giving preference to the first modal coordinate which is transverse. The design is performed by setting all prior variances, except the first, equal to 100, while we set the prior variance of the first modal coordinate to different values s_1^2 =300, 500, 1000 and 2000. For $\sigma_1^2 = 300$ or for smaller values of s_1^2, the optimal sensor is selected to be a vertical sensor. For values of s_1^2 =500, 100 and 2000, the transverse sensor is selected. The location of the transverse sensor for these higher values is shown in Figure 7.9(b). It is clear that increasing the value of s_1^2 the sensor is located closer to the location where the first mode has its maximum absolute deflection. This is consistent with the

theoretical result of Section 6.4.2 which states that increasing the prior variance of the first mode gives more and more weight to the identification of the first mode, making it more important in the optimal sensor placement design.

7.4.4 Optimal sensor placement of reference sensors

The OED is next used to address the important problem of selecting the optimal location of a few reference sensors in a multiple sensor configuration set up experiment conducted with a limited number of reference and roving sensors in order to obtain the modal frequencies and assemble the mode shapes from the multiple setups. The number of reference sensors is in most cases significantly smaller than the number of modes to be identified. It is important in this case that the reference sensors, common in most setups, contain the maximum possible information for all modes that are planned to be identified. Wrong locations of the reference sensors may degrade the modal information for one or more modes, degrading the accuracy of the corresponding assembled mode shape since such accuracy is based solely on the information contained in the reference sensors.

The effectiveness of the methodology is illustrated by designing the optimal locations of one vertical and one transverse reference sensor for identifying the lowest 10 modes of the bridge. It is clear from the results in Figure 7.3(a) that the best location of the 1^{st} sensor is at 487m measuring along the vertical direction. This is also taken as the location of the vertical reference sensor. Figure 7.4(b) demonstrates that the design of the reference sensor at 487m is rational since the deflection of all five vertical mode shapes is high. To design the location of the 1^{st} transverse reference sensor one could use the results in Figure 7.9(a), obtained for the transverse modes after selecting the prior uncertainty to give in the optimal design preferential treatment to these modes over the vertical modes. It is clear that the transverse reference sensor can be selected as the location 268m (parent location 0.5) that corresponds to the minimum information entropy in Figure 7.9(a). From Figure 7.4(a) it is seen that the transverse reference sensor location 268m corresponds to high deflections of four out of the five transverse modes. One of the transverse modes has relatively small deflection, a problem that arises from the trade-off that has to be made in the design to get maximum information from all five modes. A solution to this problem is to place a second transverse sensor on the structure at position 70m (the second best local optimum in Figure 7.4(a)). In practice, using more reference sensors that the minimum required (two in this case) is a good way to make sure that important information from reference sensors will not be lost. Concluding, the vertical and transverse reference sensors in experiments with multi sensor configuration setups could be 487m and 268m (and/or 70m), respectively.

7.5 Conclusions

An optimal sensor placement design for modal identification based on the posterior entropy as a measure of the information contained in the data is presented. Using non-uniform priors, the Bayesian OED

allows for optimal sensor placement to be performed even for the case when the number of sensors is less than the number of identified modes. This is important when designing the optimal locations of a very small number of reference sensors for the purpose of assembling the mode shapes using reference and roving sensors in multiple sensor configuration setups. In this study the effect of the Gaussian prior on the optimal design was thoroughly investigated. Insightful analytical expressions were derived to show that larger uncertainty in the prior of a subset of modal coordinates can be used to give preference in this subset in the optimal design of the sensor locations. The prior variances for all modes to be identified can be altered to weight the importance of different modes in the design, favoring a number of modes against the rest of the modes. The prior is the users' choice and it can be used in different ways to achieve different results, which is one of the strengths of Bayesian OED.

The methodology was applied to a 537m long reinforced concrete bridge in order to design the optimal sensor configuration for identifying the lowest 10 modes. The optimization was performed in the continuous space of the design variables, through appropriate mapping from the physical space to a parent domain. A large number of local optima were observed that result in a challenging optimization problem. The problem is overcome using computationally efficient heuristic FSSP algorithms. The accuracy of the FSSP algorithm was confirmed using the computationally demanding CMA-ES algorithm. A thorough investigation of the effect of correlated and uncorrelated prediction error models was also performed. The design for a smaller number of sensors than the number of modes was shown to be the same for spatially uncorrelated and correlated prediction error models. Rational and intuitive results were obtained. For more sensors than the number of modes the spatially correlated prediction error model gave intuitively reasonable results, avoiding sensor clustering observed for uncorrelated prediction error models.

The proposed method offers a useful decision tool for designing the sensor locations in a structure in order to obtain the maximum information for reliable modal identification of civil infrastructures and industrial facilities using vibration measurements.

Chapter 8

Bayesian optimal sensor location for crack identification using strain measurements

8.1 Introduction

This paper investigates the optimal experimental design problem related to finding the optimal location, number and density of strain sensors in a loaded plate for the goal of accurate and reliable identification of a crack using strain measurements. Then the optimal design is validated using simulated strain data to perform Bayesian crack identification. In both cases a computational FE model of a cracked plate under static loading is used as the forward model. Those two subjects were also studied in (Gaitanaros et al., 2010), but using different theoretical foundations and computational tools.

The objective of the present study is twofold. Firstly, a Bayesian formulation for the optimal design of strain sensor locations for crack identification is presented based on the expected KL-divergence measure. Previous work addressing the issue of optimally locating a given number of sensors in a structure has been carried out by several investigators. In particular, information theory based approaches (Kammer, 1991; Kirkegaard and Brincker, 1994; Udwadia, 1994; Shah and Udwadia, 1978; Qureshi et al., 1980; Ucinski, 2000; Alana, 2010; Chow et al., 2011) have been developed to provide rational solutions to several issues encountered in the problem of selecting the optimal sensor configuration. The optimal sensor configuration is selected as the one that maximizes some norm (determinant or trace) of the Fisher information matrix (FIM). In other studies (Heredia-Zavoni and Esteva, 1998; Heredia-Zavoni et al., 1999) the optimal sensor configuration has been chosen as the one that minimizes the expected Bayesian loss function involving the trace of the inverse of the FIM. A Bayesian framework to optimal sensor location for structural health monitoring (SHM) has also been introduced in (Flynn and Todd, 2010). The optimal configuration is

chosen to optimize (maximize or minimize) a Bayesian risk-based performance metric related to the probability of damage detection or false alarm of all regions of the structure. The information entropy, measuring the uncertainty in the model parameter estimates, was also introduced (Papadimitriou et al., 2000) for designing optimal sensor configurations. It was shown (Papadimitriou, 2004; Papadimitriou and Lombaert, 2012) that, asymptotically for very large number of data, the information entropy depends on the determinant of the FIM, justifying the use of the determinant instead of the trace or other scalar measures of FIM in previous approaches.

Herein, the KL-divergence is used to measure the quality of information that can be extracted from the data used to detect a crack. The optimal strain sensor design is obtained as the one that maximizes the expected KL-divergence (information gain) from the posterior to the prior PDFs of the crack parameters. In contrast with works using asymptotic approximations to the objective function (Gaitanaros et al., 2010; Papadimitriou, 2004; Papadimitriou and Lombaert, 2012), herein the more accurate sampling approach is used where the objective function integral is approximated with a Monte Carlo sum (Ryan, 2003; Huan and Marzouk, 2012, 2013). The change in the optimal design is investigated when different a priori knowledge about the area of the crack location is taken into account, through the prior PDF.

Secondly, a methodology for the estimation of the crack parameters based on a Bayesian system identification methodology is presented. The crack parameters may include crack location, size and orientation. Their values are estimated using measured data from a structure subjected to static loading. The location, size and orientation of the crack is inferred from the most probable values of the crack parameters obtained as the ones that maximize the posterior probability distribution of the parameters given the measured data.

The problem of crack detection in structures has received much attention over the years because of its profound importance in structural health monitoring. Early detection of cracks is a key element for preventing catastrophic failure and prolonging the life of structures. Crack identification information can be used for developing cost-effective maintenance procedures for structures, improving their safety and reducing their maintenance and rehabilitation costs, in a whole-life cost basis. Current inspection techniques involve complex, time-consuming procedures, which can be very labor-intensive and expensive. A fast, low-cost built-in structural health monitoring system involving a sensor array along with fast processing techniques is needed to overcome the shortcomings of the current inspection techniques.

Model-based methods rely on some forward model simulating the behavior of the structure with a crack, where the parameters of the model describe the crack. The most common tool for modeling the forward problem is the finite element method (FEM) (Teughels and Roeck, 2005; He et al., 2001; Horibe and Takahashi, 2007; Stavroulakis and Antes, 1998; Rus et al., 2005; Moore et al., 2011; Burczynski et al., 2004; Krawczuk, 2002). Recently, the extended FE method (XFEM) has also been introduced as a way to model discontinuities with several important advantages such as avoiding re-meshing for various flaw scenarios (Moës et al., 1999; Sukumar et al., 2001). Works using the XFEM method as the forward model to perform flaw detection include (Jung et al., 2013; Jung and Taciroglu, 2014; Nanthakumar et al., 2013, 2014; Sun et al., 2014, 2013).

From the numerical perspective of locating the flaw from measured data, the problem can be treated deterministically or probabilistically. In the deterministic approach some optimization algorithm is used to solve the inverse problem of estimating the flaw parameters by minimizing a measure of fit with the experimental data by adjusting the flaw parameters (Cawley and Adams, 1979; Chang and Liu, 2003; Gadala and McCullough, 1999; Ostachowicz, 2008). Hybrid optimization methods based on evolutionary strategies and gradient-based techniques were adopted in (Gaitanaros et al., 2010; Xu and Liu, 2002) in order to overcome the well-known problems of the gradient-based methods which in many cases simply arise from the unavailability of analytical gradients for the problem at hand. Genetic algorithms (GA) (Burczynski et al., 2004) coupled with the XFEM method have been proposed (Rabinovich et al., 2007, 2009) to perform the required optimization which are gradient-free. The XFEM-GA combination has been extended to detect flaws of various shapes, but by using a priori knowledge about the shape of the flaw (Waisman et al., 2010), and then further improved to detect flaws of any shape without a priori knowledge (Chatzi et al., 2011).

Deterministic methods in inverse problems such as model-based flaw identification only provide a point estimate of the model parameters and do not take various sources of uncertainty into account. Practically, modeling and measurement uncertainties are present and may have a significant effect in the results of the flaw detection. Probabilistic methods are able to take these uncertainties into account and provide the answer about the parameters in the form of probability density functions instead of point estimates. Herein, the Bayesian approach to probabilistic modeling is utilized, which uses probability as a way of quantifying the plausibilities associated with the values of the parameters of models given the observed data (Beck and Katafygiotis, 1998; Katafygiotis et al., 1998; Beck, 2010). The Bayesian framework for parameter estimation has been used in (Gaitanaros et al., 2010; Moore et al., 2011; Yin et al., 2010) for detecting single flaws while it has also been applied to detect multiple flaws without knowing their exact number a priori (Yan et al., 2015). For that purpose, the detection of an unknown number of flaws was proposed to be formulated as a Bayesian model selection problem (Gaitanaros et al., 2010; Beck and Yuen, 2004).

The presentation in this work is organized as follows. In Section 2, the Bayesian crack identification methodology is presented for the general case of a cracked structure based on strain measurements. In Section 3, a Bayesian formulation for the design of the optimal strain sensor locations for crack identification based on the expected KL-divergence measure is presented. The effectiveness of the proposed identification and optimal sensor location methodologies is illustrated in Section 4 by applying them to a FE model of a thin plate with a crack subjected to distributed loading. The simulated data are generated by adding noise to the FE model predictions in order to simulate the effect of measurement and model errors. Crack identification is performed for several cases of grids, numbers of sensors and noise levels. Optimal strain sensor locations are derived, and the effect of using different types of sensor grids is examined. Also the influence of the prediction error correlation model is studied. Finally, the conclusions are summarized in Section 5.

8.2 Bayesian crack identification

Consider a structure with a crack subjected to far field static loading (e.g. distributed stress, force, etc.). The objective is to identify the crack location, size and orientation using measured data such as strain measurements. We define a vector $\boldsymbol{\theta} \in R^{N_{\boldsymbol{\theta}}}$ of the parameters that describe the crack, namely the (X_c, Y_c) coordinates of one of the crack tips, crack length L and orientation ϕ. Therefore $N_{\boldsymbol{\theta}} = 4$ and the problem of crack identification is equivalent to the problem of estimating $\boldsymbol{\theta} = \{X_c, Y_c, L, \phi\}^T$.

A Bayesian parameter estimation framework is used to estimate the value of $\boldsymbol{\theta}$ and its associated uncertainty, using information from the data set $\boldsymbol{D}(\boldsymbol{d}) = \{\hat{g}^i(\boldsymbol{d}),\ i = 1, \ldots, N\}$ which consists of N readings of strain sensors located in the cracked structure at positions described by the design variables $\boldsymbol{d}$. Let M_m represent a model parametrized by $\boldsymbol{\theta}$, simulating the behavior of the cracked plate. This model predicts the strains at the locations described by $\boldsymbol{d}$ for a given crack configuration described by $\boldsymbol{\theta}$, namely $\boldsymbol{g}(\boldsymbol{d}, \boldsymbol{\theta})$. Herein, the model is associated with the solution of the stress and strain field in a cracked plate, and is described in detail in Section 8.4.1.

In order to proceed with the Bayesian formulation for parameter estimation, one can use the prediction error equation:

$$\hat{\boldsymbol{g}}(\boldsymbol{d}) = \boldsymbol{g}(\boldsymbol{d}, \boldsymbol{\theta}) + \boldsymbol{e}(\boldsymbol{d}) \tag{8.1}$$

where $\boldsymbol{e}(\boldsymbol{d})$ is the prediction error vector which is a probabilistic quantity that describes the discrepancy between the model predictions $\boldsymbol{g}(\boldsymbol{d}, \boldsymbol{\theta})$ and experimental data $\hat{\boldsymbol{g}}(\boldsymbol{d})$. The prediction error is due to modeling errors and measurement noise, and is commonly treated as a Gaussian zero-mean random vector with some user-specified covariance matrix, namely $\boldsymbol{e}(\boldsymbol{d}) \approx N(\boldsymbol{0}, \Sigma_e(\boldsymbol{d}))$. The correlation structure of its covariance matrix can depend on the sensor locations through the design variables $\boldsymbol{d}$. For the purposes of parameter estimation a diagonal (uncorrelated) prediction error covariance matrix is chosen. It does not depend on $\boldsymbol{d}$, $\Sigma_e(\boldsymbol{d}) = \Sigma_e(\sigma_i) = diag(\sigma_1^2, \ldots, \sigma_N^2)$, where σ_i^2 is the variance assigned to the i-th component of the model prediction strain vector $g^i(\boldsymbol{d}, \boldsymbol{\theta})$. Next it is assumed for simplification that the variances for each model prediction component are equal and the covariance matrix becomes: $\Sigma_e(\sigma_i) = \Sigma_e(\sigma) = \sigma^2 I$. For prediction error formulations where the error depends on the model prediction the identity matrix I can be substituted by a diagonal matrix of model predictions, $\Sigma_e(\sigma) = \sigma^2 diag(\boldsymbol{g}(\boldsymbol{d}, \boldsymbol{\theta}))$. A more sophisticated study of the prediction error for crack identification problems can be found in (Gaitanaros et al., 2010). Therefore, there is just one additional parameter added due to the prediction error, and the augmented parameter vector becomes: $\boldsymbol{\theta} = \{X_c, Y_c, L, \phi, \sigma\}$. So the identification procedure includes the identification of the prediction error parameter σ simultaneously with the identification of the crack parameters.

According to the Bayesian system identification methodology (Beck and Katafygiotis, 1998), the values of $\boldsymbol{\theta}$ are modeled by a PDF that quantifies the plausibility of each possible value, given the data set $\hat{\boldsymbol{g}}(\boldsymbol{d})$. From the Bayesian theorem, this PDF is the posterior PDF given by:

$$p(\boldsymbol{\theta}|\hat{\boldsymbol{g}}(\boldsymbol{d})) = \frac{\pi(\boldsymbol{\theta}) p(\hat{\boldsymbol{g}}(\boldsymbol{d})|(\boldsymbol{d}, \boldsymbol{\theta}))}{p(\hat{\boldsymbol{g}}(\boldsymbol{d}))} \tag{8.2}$$

From Equation (8.1) it follows that the experimental data vector $\hat{\boldsymbol{g}}(\boldsymbol{d})$ also follows a Gaussian distribution which means that the likelihood function is Gaussian (when viewed as a function of the data). The prior PDF for the model and prediction error parameters $\pi(\boldsymbol{\theta})$ is assumed to be Uniform, and the evidence term is a constant, therefore the posterior PDF can be written as:

$$p(\boldsymbol{\theta}|\hat{\boldsymbol{g}}(\boldsymbol{d})) = c\, p(\hat{\boldsymbol{g}}(\boldsymbol{d})|(\boldsymbol{d},\boldsymbol{\theta})) = c\, N(\hat{\boldsymbol{g}}(\boldsymbol{d})|\boldsymbol{g}(\boldsymbol{d},\boldsymbol{\theta}),\sigma^2 I) \tag{8.3}$$

where c is a constant such that the posterior PDF integrates to 1 and the notation $N(\boldsymbol{x}|\boldsymbol{\mu},\Sigma)$ denotes that the random vector $\boldsymbol{x}$ follows the multivariate Normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. We see that the model evaluation enters the formulation through the likelihood function, and especially through its mean. The posterior PDF, and hence the likelihood of Equation (8.3) is viewed as a function of $\boldsymbol{\theta}$. Accordingly, the model prediction $\boldsymbol{g}(\boldsymbol{d},\boldsymbol{\theta})$ is also viewed as a function of $\boldsymbol{\theta}$. For simple models $\boldsymbol{g}(\boldsymbol{d},\boldsymbol{\theta})$ is an analytical function of $\boldsymbol{\theta}$ and a closed-form expression can be derived for the posterior PDF. However, for numerical models, like FE models, there is no explicit formula for $\boldsymbol{g}(\boldsymbol{d},\boldsymbol{\theta})$ and it can only be evaluated point-wise for specific values of $\boldsymbol{\theta}$. Then in order to explore the posterior PDF, one can resort to stochastic simulation methods which work by drawing many $\boldsymbol{\theta}$ samples from the posterior PDF. Then, one can use those samples as a representation of the actual posterior PDF and perform statistical inference based on the samples. Herein, the TMCMC algorithm (Ching and Chen, 2007) is used to draw samples from the posterior, which requires the ratio between two posterior values for two $\boldsymbol{\theta}$ samples. This even removes the need to know the normalizing constant c because one has from Equation (8.3) for the ratio:

$$\frac{p(\boldsymbol{\theta}_1|\hat{\boldsymbol{g}}(\boldsymbol{d}))}{p(\boldsymbol{\theta}_2|\hat{\boldsymbol{g}}(\boldsymbol{d}))} = \frac{p(\hat{\boldsymbol{g}}(\boldsymbol{d})|(\boldsymbol{d},\boldsymbol{\theta}_1))}{p(\hat{\boldsymbol{g}}(\boldsymbol{d})|(\boldsymbol{d},\boldsymbol{\theta}_2))} = \frac{N(\hat{\boldsymbol{g}}(\boldsymbol{d})|\boldsymbol{g}(\boldsymbol{d},\boldsymbol{\theta}_1),\sigma^2 I)}{N(\hat{\boldsymbol{g}}(\boldsymbol{d})|\boldsymbol{g}(\boldsymbol{d},\boldsymbol{\theta}_2),\sigma^2 I)} \tag{8.4}$$

Consequently, only a point-wise evaluation of the forward model for different values of the parameters $\boldsymbol{\theta}$ which serve as the means of the Gaussian likelihoods is required by the algorithm. Then the evaluation of the Gaussian densities is straightforward:

$$N(\hat{\boldsymbol{g}}(\boldsymbol{d})|\boldsymbol{g}(\boldsymbol{d},\boldsymbol{\theta}),\Sigma_e) = \frac{\exp\left(-\frac{1}{2}(\hat{\boldsymbol{g}}(\boldsymbol{d})-\boldsymbol{g}(\boldsymbol{d},\boldsymbol{\theta}))^T\Sigma_e^{-1}(\hat{\boldsymbol{g}}(\boldsymbol{d})-\boldsymbol{g}(\boldsymbol{d},\boldsymbol{\theta}))\right)}{\sqrt{|2\pi\Sigma_e|}} \tag{8.5}$$

The TMCMC algorithm draws samples from the important high-probability region of the parameter space by comparing posterior values for different parameters using the ratio of Equation (8.4) to decide if a new candidate $\boldsymbol{\theta}$ sample is going to be accepted or rejected. Higher posterior values are more likely to be accepted, which leads to $\boldsymbol{\theta}$ samples which give a better fit with the experimental data, as can be seen from the minimization of the quadratic form of Equation (8.5). More details about the TMCMC algorithm, its implementation and its application can be found in (Ching and Chen, 2007; Angelikopoulos et al., 2015; Hadjidoukas et al., 2015).

8.3 Bayesian optimal sensor placement

The posterior PDF of Equation (8.2) completely describes the updated state of knowledge about the parameters $\boldsymbol{\theta}$ in light of the experimental data $\hat{\boldsymbol{g}}(\boldsymbol{d})$, while the prior PDF $\pi(\boldsymbol{\theta})$ describes our initial state

of knowledge before the data were collected. Therefore, the "distance" between the prior and posterior PDF is the information gain about the parameters $\boldsymbol{\theta}$ that was caused by the data $\hat{\boldsymbol{g}}(\boldsymbol{d})$. However, both the prior and posterior are PDFs and one needs a suitable way to compare them and measure their information difference about $\boldsymbol{\theta}$. One commonly used way to measure the "distance" between two PDFs is the Kullback-Leibler divergence (KL-div) (Cover and Thomas, 2006) which is a scalar measure of the information gain about $\boldsymbol{\theta}$ when going from the prior to the posterior PDF. Being a scalar measure of information gain, the KL-div is a suitable measure to be used as an objective function for optimal sensor placement. It depends on the experimental data $\hat{\boldsymbol{g}}(\boldsymbol{d})$ through the posterior, and the experimental data depend on the design variables $\boldsymbol{d}$ since the sensor locations determine where the data is going to be coming from. Therefore, we can view the KL-div as a function of $\boldsymbol{d}$. Next the expectation is taken over all the possible data resulting in the Expected KL-div as our final objective function. In this way the KL-Div is averaged over all the possible data that can arise from the design $\boldsymbol{d}$, based on the likelihood and prior PDFs of our problem. This leads to the decision-theoretic expected utility approach of Lindley (Lindley, 1956) with the KL-Div between the posterior and prior PDF taken as the utility function.

The Expected KL-Div (objective function) is defined as:

$$U(\boldsymbol{d}) = E_{\hat{\boldsymbol{g}}(\boldsymbol{d})}\left[D_{KL}(p(\boldsymbol{\theta}|\hat{\boldsymbol{g}}(\boldsymbol{d}))\,||\,\pi(\boldsymbol{\theta}))\right] \tag{8.6}$$

$$= \int_{\hat{\boldsymbol{G}}} p(\hat{\boldsymbol{g}}(\boldsymbol{d})) \left[\int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\hat{\boldsymbol{g}}(\boldsymbol{d})) \log \frac{p(\boldsymbol{\theta}|\hat{\boldsymbol{g}}(\boldsymbol{d}))}{\pi(\boldsymbol{\theta})}\, d\boldsymbol{\theta}\right] d\hat{\boldsymbol{g}}(\boldsymbol{d}) \tag{8.7}$$

In general, the objective function of Equation (8.6) can only be estimated by Monte Carlo sampling. Ryan (Ryan, 2003) has proposed a Monte Carlo estimator which uses samples from the prior and likelihood PDFs to approximate it with a double sum:

$$U(\boldsymbol{d}) \approx \frac{1}{N}\sum_{i=1}^{N}\left\{\log p(\hat{\boldsymbol{g}}(\boldsymbol{d})_i|(\boldsymbol{d},\boldsymbol{\theta}_i)) - \log\left\{\frac{1}{M}\sum_{j=1}^{M} p(\hat{\boldsymbol{g}}(\boldsymbol{d})_i|(\boldsymbol{d},\boldsymbol{\theta}_{i,j}))\right\}\right\} \tag{8.8}$$

where samples $(\boldsymbol{\theta}_i, \hat{\boldsymbol{g}}(\boldsymbol{d})_i)|\boldsymbol{d}$ from the joint PDF $p(\boldsymbol{\theta}, \hat{\boldsymbol{g}}(\boldsymbol{d})|\boldsymbol{d}) = \pi(\boldsymbol{\theta})\, p(\hat{\boldsymbol{g}}(\boldsymbol{d})|(\boldsymbol{d},\boldsymbol{\theta}))$ were used to approximate the integrals with Monte Carlo sums (Ryan, 2003). This estimator of the objective function has also been used in (Huan and Marzouk, 2012, 2013) to perform optimal sensor placement with the help of surrogate models and stochastic optimization techniques.

Note that for each parameter sample from the prior PDF the evaluation of the likelihood includes one forward model run for the corresponding sample. Therefore, Equation (8.8) would require $N + NM$ forward model runs. In order to reduce that computational burden to just N model runs it is common practice to use the same set of prior samples for both the outer and inner sum, accepting a small reduction in the quality of the estimator (Huan and Marzouk, 2012, 2013):

$$U(\boldsymbol{d}) \approx \frac{1}{N}\sum_{i=1}^{N}\left\{\log p(\hat{\boldsymbol{g}}(\boldsymbol{d})_i|(\boldsymbol{d},\boldsymbol{\theta}_i)) - \log\left\{\frac{1}{N}\sum_{i=1}^{N} p(\hat{\boldsymbol{g}}(\boldsymbol{d})_i|(\boldsymbol{d},\boldsymbol{\theta}_i))\right\}\right\} \tag{8.9}$$

Therefore, in order to compute $U(\boldsymbol{d})$ one has to perform the following steps:

1. Draw N random parameter samples from the prior PDF $\pi(\boldsymbol{\theta}) \rightarrow \boldsymbol{\theta}_i$, for $i = 1, \ldots, N$.

2. For each $\boldsymbol{\theta}_i$ draw a random data sample from the likelihood function $p(\hat{\boldsymbol{g}}(\boldsymbol{d})|(\boldsymbol{d}, \boldsymbol{\theta})) \rightarrow \hat{\boldsymbol{g}}(\boldsymbol{d})_i$. This step requires the forward model run.

3. Evaluate the log-densities of the likelihoods required in Equation (8.9) and calculate the sums. Note that no further model runs are required for the second (inner) sum since the likelihood evaluations for different means (samples) require the model outputs that have already been carried out in order to draw the N data samples.

With Equation (8.9) providing a Monte Carlo estimate of the Expected KL-div (expected information gain) as a function of the design variables $\boldsymbol{d}$, one can use it as an objective function for performing optimal sensor placement. The objective is to find the sensor design $\boldsymbol{d}$ that maximizes the expected information gain.

8.4 Application

In this section the Bayesian optimal sensor placement and crack identification methodologies are applied to a FE model of a square plate with a crack. The strain field is computed using a FE model described in detail next.

8.4.1 Description of FE model

The FE model was developed in the commercial FE software ANSYS (ANSYS, 2013). The geometry of the model, shown in Figure 8.1, is a square steel plate of dimensions $1 \times 1\,m$ which is fixed in its bottom and left sides, while a static loading is distributed uniformly along the top side pointing upwards. The material behavior is linear elastic with modulus of elasticity $E = 210 \times 10^9 Pa$ and Poisson ratio $\nu = 0.3$, and the analysis type is static. The model is parametrized by the crack location, length and orientation as shown in Figure 8.1. The location of the crack is specified by the coordinates of one of the crack tips (X_c, Y_c), while the orientation is specified by the angle ϕ the crack makes with the bottom side of the plate.

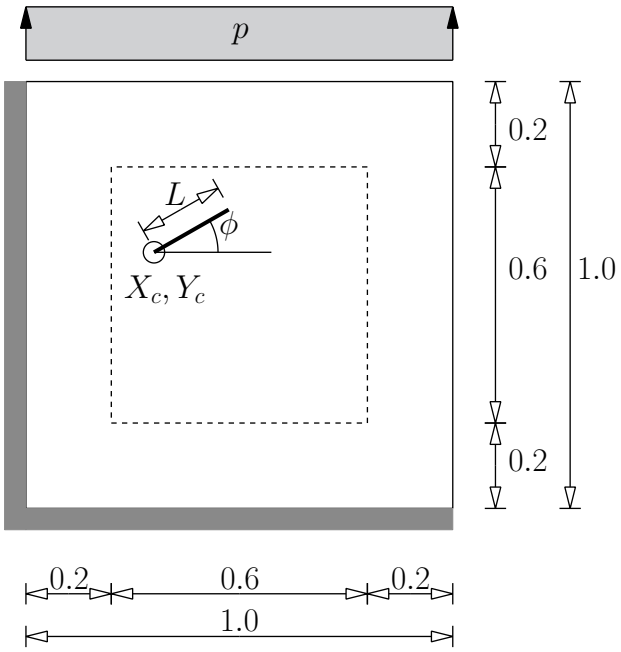
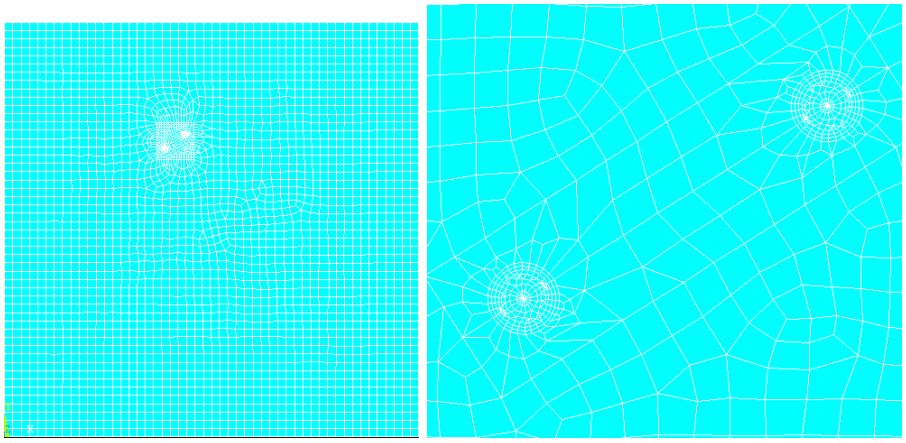


Figure 8.1: Description of model parameters, boundary conditions and loading.

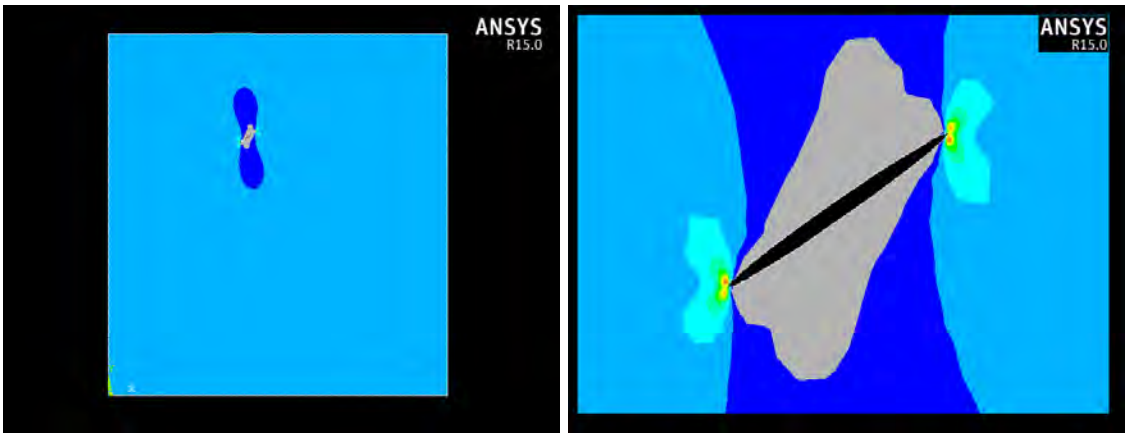
In this study, the system was modeled using 2D elements with 4 nodes having 2 DOFs at each node: the displacements in each direction. The entire mesh consists of about 3300 elements, 3300 nodes, and 6600 DOFs. The exact number of nodes depends on the crack configuration since re-meshing is required to be performed for each different crack location, orientation and size. The mesh for an arbitrary crack is shown in Figure 8.2.



(a) (b)

Figure 8.2: (a): Finite Element mesh. (b): Finite Element mesh zoomed in crack area.

For the calculation of the strain field around the crack the theory of Linear Elastic Fracture Mechanics was applied (ANSYS, 2013). The static solution of the FE model yields the displacement field at all FE nodes which in turn is transformed into the strain field at all FE nodes. The strain distribution field is depicted in Figure 8.3.



(a) (b)

Figure 8.3: (a): Strain field. (b): Strain field zoomed in crack area.

8.4.2 Optimal sensor placement

In this section the sampling algorithm is applied in order to find the optimal locations of strain sensors for crack identification. The sensors measure the plane strains on the surface of the plate in the vertical direction which is parallel with the loading. Three alternative types of sensor grids with uniformly distributed sensors are examined: 1) full square grid, 2) boundary square grid, and 3) circular grid. In order to make comparisons between the 3 grids we use the same number of sensors for each grid. A $N \times N$ full square grid has N^2 sensors. To maintain the same number of sensors in the other two sensor grids we place $N^2/4$ sensors per side in the boundary square grid and uniformly distribute the N^2 sensors along the circumference in the circular grid.

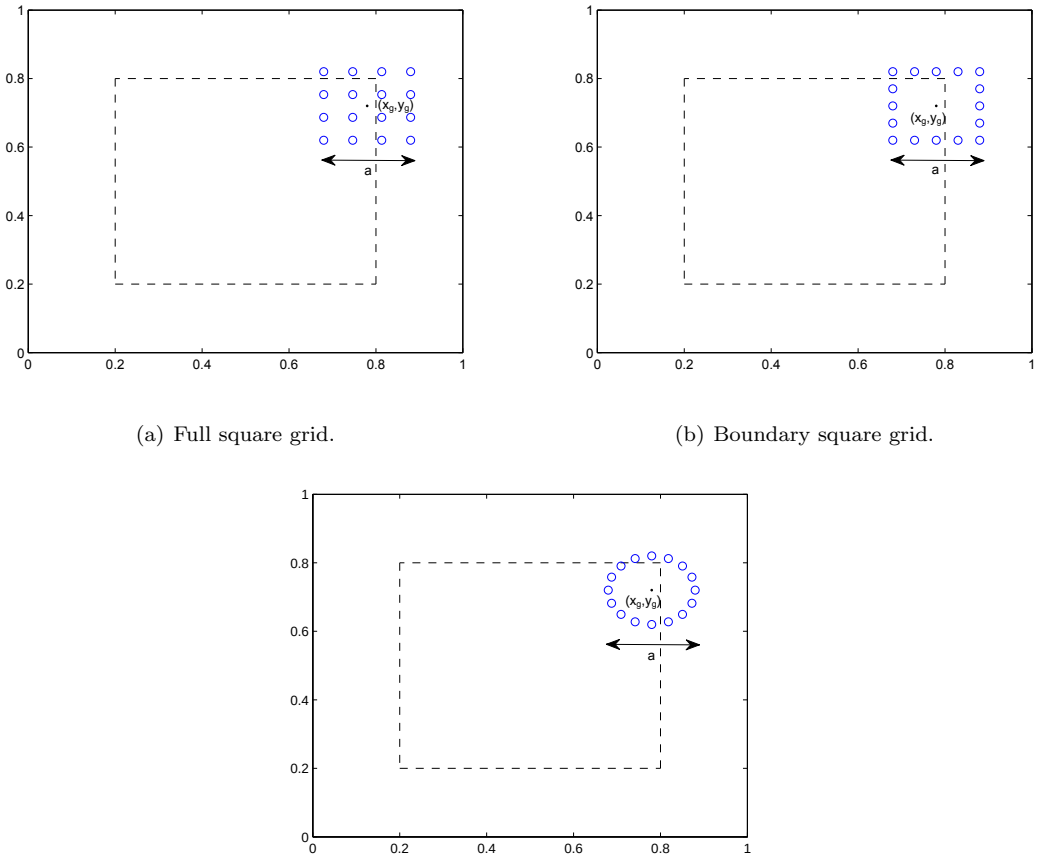
The goal is to optimally design the placement of sensors so that we monitor a crack that could be manifested inside a specific region Ω in the plate, with orientation values ranging from ϕ_{min} to ϕ_{max} and crack sizes ranging from L_{min} to L_{max}. This information is inserted in the analysis by introducing a uniform prior for the parameters of the crack. Specifically, the prior distribution of the location of the crack is assumed to be uniform within the domain Ω, while the prior distribution of the orientation angle and crack size are assumed to be uniform with supports the one dimensional domains $[\phi_{min}, \phi_{max}]$ and $[L_{min}, L_{max}]$, respectively.

For demonstration purposes, results are obtained for the following values of the crack parameters: $\Omega = [0.2, 0.8] \times [0.2, 0.8]$, $\phi_{min} = 0$, $\phi_{max} = 360$, $L_{min} = 40mm$ and $L_{max} = 60mm$. The optimal sensor

placement algorithm is applied using a number of 1000 samples generated from the prior distribution of the crack parameters. This requires as many as 1000 model runs that are performed offline. The number of crack parameter samples is fixed for different values of the design variables $\boldsymbol{d}$.

Three prediction error (likelihood) models are examined, one uncorrelated and two correlated models. The uncorrelated prediction error model has a diagonal covariance matrix $\Sigma = \sigma^2 I$ independent of the design variables d. The correlated model has a non-diagonal covariance matrix of the form $\Sigma_{ij} = \sigma^2 R(d_{ij})$. The spatial correlation structure $R(d_{ij})$ of the prediction error was chosen to be exponential $R(d_{ij}) = \exp(-d_{ij}/\lambda)$, where d_{ij} is the distance between the sensors i and j and λ is the correlation length. In both models σ represents the strength of the prediction error and was chosen to be 1% of the model prediction. In order to investigate the effect of the correlation length, two cases were examined: $\lambda = 100mm$ and $\lambda = 200mm$.

For a fixed number of sensors the design variables in $\boldsymbol{d}$ are the location of the grid and its size. The location of the grid is defined by the location of its center (x_g, y_g) and its size is defined by the length a of the side for the square grids and the diameter a for the circle. Thus, the design vector is $\boldsymbol{d} = (x_g, y_g, a)$.



(a) Full square grid.

(b) Boundary square grid.

(c) Circular grid.

Figure 8.4: Design variables for the three grid types.

Preliminary optimal sensor placement results have verified the obvious idea that the optimal location of the grid is the center of the plate. This is expected since the prior domain for the crack location is the interior square $[0.2, 0.8] \times [0.2, 0.8]$ which is also centered at the center of the plate. With this prior domain it would not be sensible for the grids to be optimal anywhere else but the center of the plate. Therefore, the 3 sensor grids were fixed at the center of the plate such that only the effect of the grid size was examined, for a fixed number of sensors each time. Accordingly, for each number of sensors the objective function (expected KL-divergence) was evaluated as a function of the grid size. This analysis was repeated for each one of the 3 grid types and prediction error correlation models. The optimal grid size is the one which maximizes the objective function.

The maximum of the utility function for given number of sensors is plotted as a function of the number of sensors. In Figure 8.5, nine (9) lines are presented corresponding to different grid types and correlation models. The optimal grid size at which this maximum is attained for a fixed number of sensors is shown in

Table 8.1 for the uncorrelated prediction error model. The analysis has been performed for the following numbers of sensors: N^2 for $N = 2, \ldots, 11$.

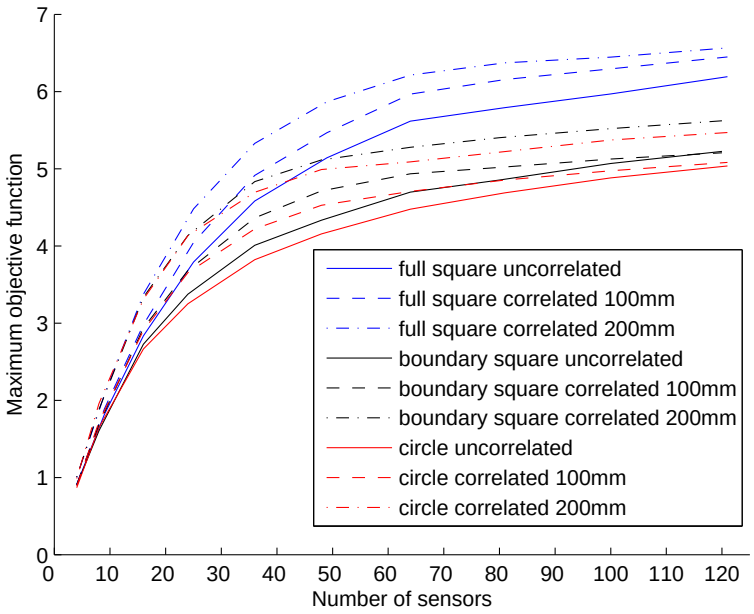


Figure 8.5: Objective function maxima versus number of sensors for each grid case and correlation case.

It can be seen very clearly that for all numbers of sensors the full square grid is the best of the 3 grids for all three prediction error models considered. The second best is the boundary square grid, while the circular grid is quite close to the boundary square grid. The grids provide similar information for small numbers of sensors (< 16) while for larger numbers of sensors (> 16) the difference in the information that each grid provides is evident. The full square grid outperforms the other two grids for the same number of sensors.

It can be seem from Figure 8.5 that the information gain increases an the number of sensors, placed in their optimal locations, increases. This increase is steep for small numbers of sensors (approximately less than 60), signifying that the information gain is large, and slows down for sufficiently large numbers of sensors, signifying that the amount of information gained form the optimal grids with extra sensors is not considerable after a number of sensors has been optimally placed in the structure. The results in Figure 8.5 are important for selecting the type of the sensor grid as well as the number of sensors to be used by trading-off extra information with the cost of sensors and instrumentation. Regarding the prediction error correlation length, we see that as the correlation length increases the information gain slows down more quickly. This means that the same amount of information is gained with fewer number of sensors required for the less spatial correlated prediction error case.

The full square grid is better than the other two grids for the same number of sensors, but with a

different size. Table 8.1 shows the total information of Figure 8.5 together with the optimal grid sizes for each grid type for the uncorrelated prediction error model. Optimal sensor placement designs are shown for the cases of 36 and 81 sensors in Figure 8.6 for all 3 grid types. It can be seen that the size of the full square grid is larger than the size of the other 2 grids, consistently covering a bigger area of the plate, whereas the other two grids are optimal at smaller sizes. However, Figure 8.5 does not show at which grid size the optimum occurs for a given number of sensors.

Thus, configurations with sensors only in boundary square grids are sub-optimal compared with full square grids for the same number of sensors. Note that this result is true only on average, for the specific prior PDF assigned for the crack parameters. It is possible that for another area in the plate monitored for cracks the boundary square grid or the circular grid is better than the full square grid. However, if the crack can be anywhere in the domain Ω defined by the prior PDF, the full grid is better on average.

Table 8.1: Optimal objective values and grid sizes for the uncorrelated prediction error case.

No. of sensors	Full square		Boundary square		Circular	
	Objective	Size	Objective	Size	Objective	Size
4	0.90	0.20	0.90	0.22	0.86	0.28
9	1.82	0.36	1.60	0.36	1.64	0.36
16	2.83	0.40	2.72	0.40	2.66	0.44
25	3.79	0.46	3.37	0.40	3.24	0.46
36	4.58	0.51	4.00	0.44	3.82	0.48
49	5.14	0.51	4.40	0.42	4.15	0.48
64	5.61	0.53	4.72	0.44	4.47	0.51
81	5.79	0.53	4.85	0.44	4.68	0.51
100	5.96	0.55	5.06	0.44	4.88	0.48
121	6.19	0.55	5.22	0.42	5.03	0.51

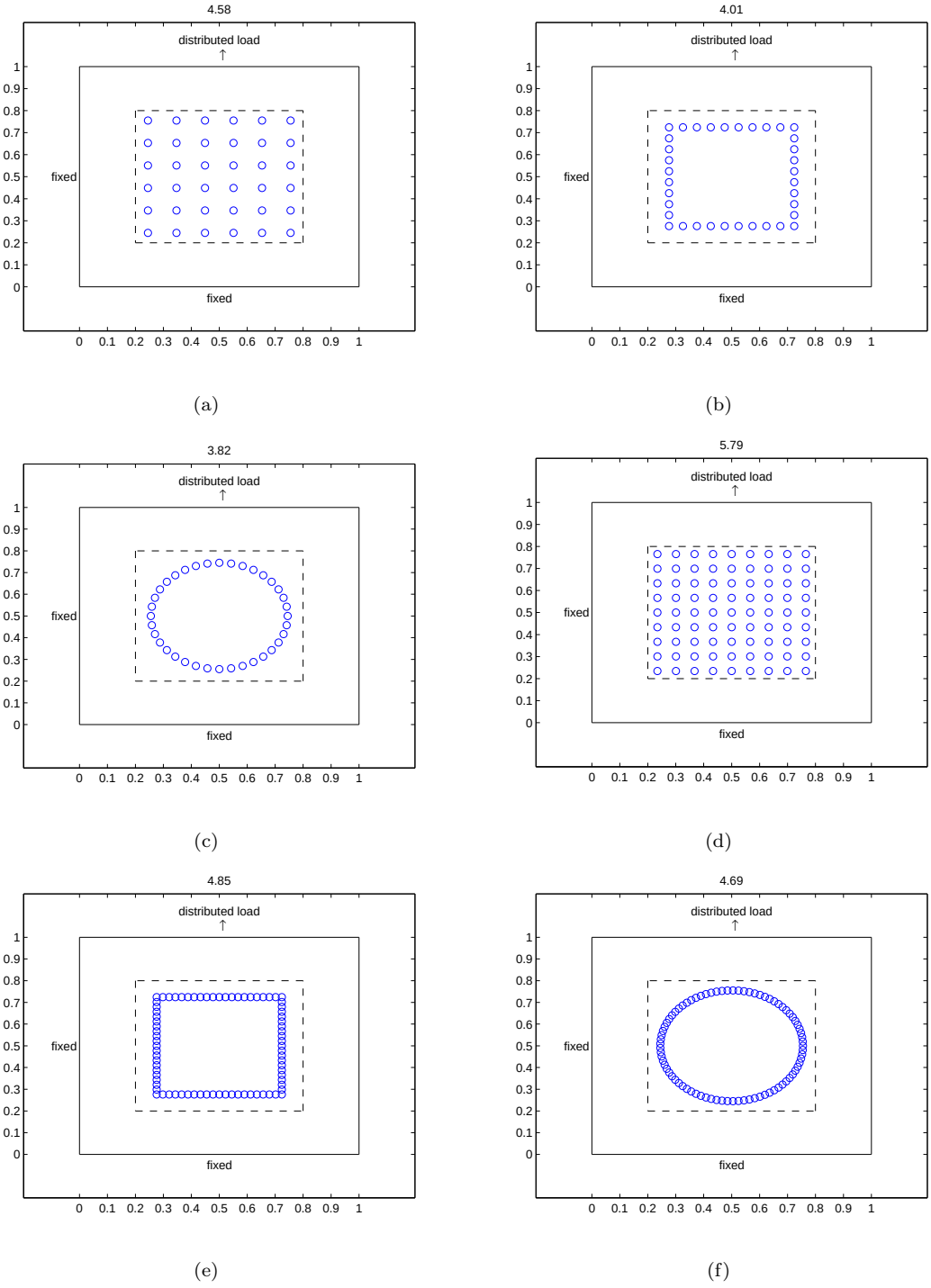
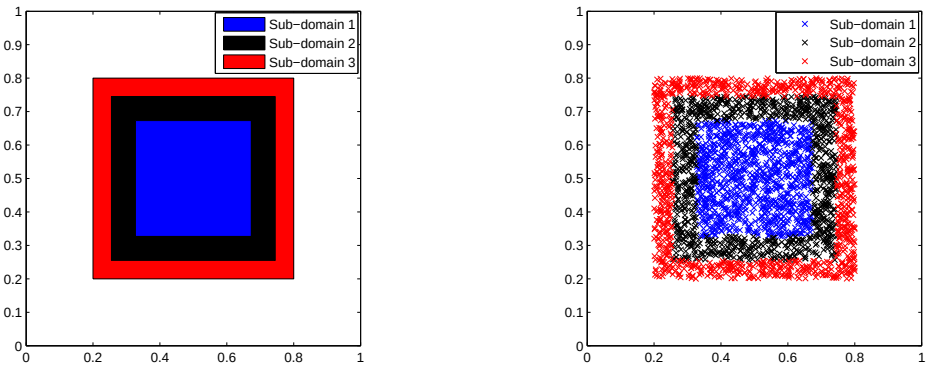


Figure 8.6: Optimal grids for 36 (a,b,c) and 81 (d,e,f) sensors for each grid type with objective values.

Performance of optimal sensor placement for monitoring cracks in sub-domains

The performance of the optimal sensor placement designs is evaluated for monitoring cracks in sub-domains $\Omega_i \subset \Omega$ of the domain Ω used to design the sensor configuration. For this reason, the prior domain Ω of the crack location is partitioned in the 3 sub-domains shown in Figure 8.7(a), representing the interior, middle, and exterior parts of the prior domain. The information gain for monitoring for a crack assuming that is contained in each one of the sub-domains Ω_i is obtained by evaluating the objective function using as support of the uniform prior PDF for the crack location each one of the three sub-domains. The samples generated from the uniform prior for each sub-domain are shown in Figure 8.7(b). A number of 1000 samples per sub-domain is used to evaluate the objective function for each sub-domain.

There is a subtle point here that needs to be properly addressed in order for the analysis to be correct. Since the prior PDF is directly involved in the objective function, in order to compare the objective function values between the different sub-domains, one needs to be sure that the prior PDFs are equivalent, that is, they have the same entropy. Otherwise the values from different sub-domains correspond to different prior PDFs and cannot be compared. For uniform prior PDFs in 2 dimensions (x, y coordinates) this means that the total area of each sub-domain must be the same with all the other sub-domains. Then the prior PDFs have the same entropy and the information gain for each sub-domain can be compared with the others. For this reason, the sizes of the sub-domains in Figure 8.7 cannot be arbitrary but were selected specifically such that the areas covered by the 3 sub-domains are equal. It is simple to prove that in order for the 3 sub-domains to have the same area the sides of the squares must satisfy the relations: $b = \sqrt{\frac{2}{3}}a$ and $c = \sqrt{\frac{1}{2}}b$ where $a > b > c$ are the sides of the outer, middle, and inner squares respectively. The requirement to have same areas in the sub-domains also makes sense from a computational point of view because the same number of samples is used to cover the same areas. Otherwise the objective function would be evaluated with a different accuracy in each sub-domain which is not desirable.



(a) (b)

Figure 8.7: The three equal-area sub-domains without (a) and with (b) uniformly distributed random samples.

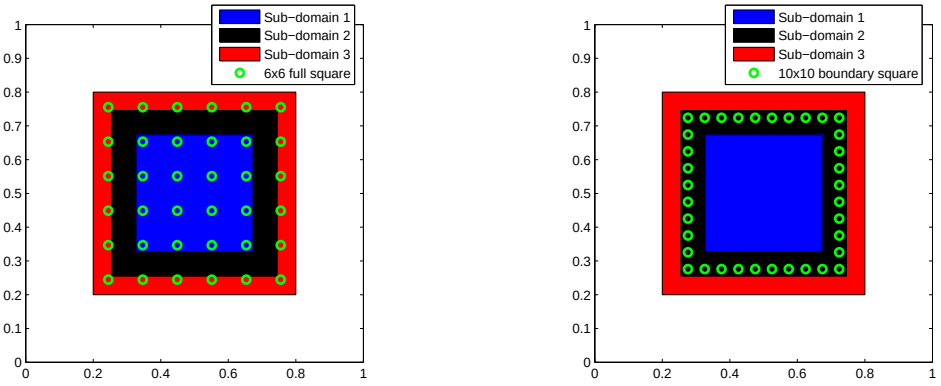
Results for the objective function for each sub-domain, for three different numbers of sensors are shown in Table 8.2 for the full square and the boundary square grids. Note that the objective function values of Table 8.2 cannot be compared with those of Table 8.1 because of the different prior PDFs used in Table 8.2. However, the values within Table 8.2 are suitable for comparison with each other because of the reason mentioned above.

The full square grid provides similar information gain for all three sub-domains with the highest information gain provided for sub-domain 1. This is expected since more of the sensors of the full square grid receive higher readings when the crack is in sub-domain 1 than in sub-domains 2 and 3, as shown in Figure 8.8(a). For the boundary square grid, the highest information gain is obtained for sub-domain 2, while the worst information is provided for sub-domain 1. By observing the optimal boundary square configuration in Figure 8.8(b), this is very reasonable because its sensors fully cover the area in the sub-domain 2, while they are further away from the sub-domain 1. Also note that the boundary square grid gives a higher information gain for sub-domain 2 than the information gain provided by the full square grid.

One needs to keep in mind that this analysis is done based on knowing that the crack is in each of the 3 sub-domains. This becomes clear in the case of sub-domain 2. Only by knowing that the crack is in sub-domain 2, the boundary square grid performs better than the full square grid. This is plausible, because a look at the two grids in Figures 8.8(a) and 8.8(b) makes obvious that the boundary square grid is much better suited for monitoring the cracks in sub-domain 2.

So, under different known information about the crack the optimal designs do change. However, this does not change the fact that if the prior knowledge of the crack is the entire $[0.2,\ 0.8] \times [0.2,\ 0.8]$ domain (union of all 3 sub-domains) then the full square grid is optimal, as was demonstrated earlier. One can see that optimality is domain-dependent with the full square grid giving greater average information gain

over the 3 sub-domains compared to the boundary square grid. Even though the boundary square grid is significantly better if it is known that the crack is in sub-domain 2, on average, over all 3 sub-domains it gives less information gain than the full square grid. Note that the average values over the 3 sub-domains correspond to those found earlier in Table 8.1 which were for the entire $[0.2, 0.8] \times [0.2, 0.8]$ prior PDF domain Ω.



(a) (b)

Figure 8.8: The three sub-domains along with the optimal: (a) 6×6 full square grid, (b) 10×10 boundary square grid.

Table 8.2: Information gain comparisons for the 3 sub-domains for the full square and boundary square grids.

Number of sensors	Objective function			
	Domain 1	Domain 2	Domain 3	Average
Full square				
36	4.74	4.57	4.30	4.53
49	5.40	5.20	4.80	5.13
64	5.80	5.70	5.40	5.63
Boundary square				
36	2.25	5.71	4.08	4.01
49	2.70	6.00	4.58	4.42
64	2.92	6.35	4.95	4.74

8.4.3 Crack identification

In this section crack identification results are presented using the proposed Bayesian parameter estimation methodology. The parameters $\boldsymbol{\theta} = \{X_c, Y_c, L, \phi\}^T$ of the model are those describing the crack tip location,

length and orientation. The optimal sensor grids obtained in the previous section are used for crack identification and parameter estimation. As in the previous analyses it is assumed that the strain sensors measure in the vertical direction, which is parallel with the loading.

The initial state of knowledge about the crack is represented by the same uniform prior PDF also used in the previous section, that is, the crack is located in the domain $\Omega = [0.2, 0.8] \times [0.2, 0.8]$ and can have any orientation $\phi \in [0, 360]$. However, in the identification process also cracks with a greater length are allowed, and therefore the uniform prior PDF bounds for the crack length were set to $10 - 200mm$.

Simulated data are used to demonstrate the framework. The simulated data are created by running the FE model for a fixed (known) crack location, size and orientation and then contaminating the resulting strains at the sensor locations by Gaussian noise. This noise is the source of uncertainty in the identification results, and accounts for the model error combined with the sensor measurement error. Based on the knowledge about the true crack parameters the identification results and their uncertainty for various grids and noise levels can be assessed and useful conclusions can be drawn.

The Bayesian parameter estimation was carried out using the TMCMC algorithm (Ching and Chen, 2007) with 1000 samples per TMCMC stage to populate the posterior PDF of the model parameters. Given the posterior samples, we estimate the most probable values of the model parameters along with their associated uncertainty in the form of 90% intervals.

Crack identification results are presented for two cases. A small crack of size $50mm$ and a larger crack of size $100mm$ which is expected to be easier to identify. For both cases the crack tip is located at (0.4, 0.6) and the crack has an orientation of 45 degrees.

We study the effect of the following in the identification results.

1. Number of sensors for a given grid type.

How does increasing or decreasing the number of sensors affect the crack identification and its uncertainty ? How many sensors are needed in order to identify a small crack ($50mm$) accurately ? How many for the larger crack ($100mm$) ?

2. Sensor.

Is it important if a full square grid is used instead of a boundary square grid or the opposite, for the same number of sensors ? This question was properly addressed in the optimal sensor placement section where it was shown that the full square grid is superior. It cannot be answered fully through the identification process because one has always to choose a particular crack to work with, and that crack may favor one grid over the other. So general conclusions can not be drawn unless a large number of cracks is examined. However, this would be impractical since too many TMCMC runs would have to be performed. Despite that, some illustrative identification results are obtained that compare the full square grid with the boundary square grid.

3. Level of measurement and model error.

How does the measurement + model error affect the identification and its uncertainty ? Naturally one would assume that more noise increases the inaccuracy of crack identification with a higher level of uncertainty.

Effect of number of sensors

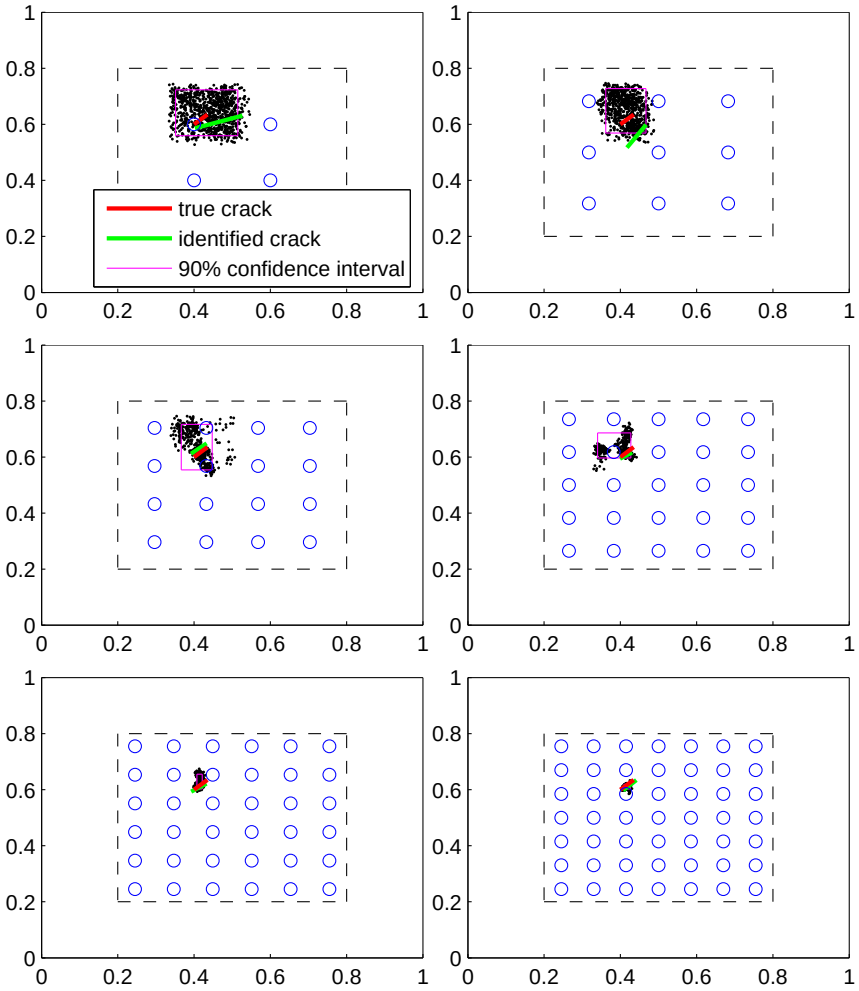


Figure 8.9: ID results for the 50mm crack with optimal full square grids from 2x2 to 7x7: Crack center uncertainty.

Parameter estimation results such as the most probable value and the 90% uncertainty interval for the location, the size and the orientation of the crack are presented in Figures 8.9 and 8.10 for different numbers of $N \times N$ sensors, N ranging from 2 to 7, for the full square grid. Results correspond to 1% noise. In these figures the identification results are presented visually in the plate, where the true crack (location, size and orientation) is presented along with the most probable identified crack, and the 90% uncertainty interval of the identified crack. The most probable crack is the one corresponding to the posterior sample which gives the maximum posterior PDF of the model parameters. The 90% interval

is a measure of the uncertainty associated with the identified crack. It provides information about the spread of uncertainty in the crack location, size and orientation based on the strain data. It is estimated by taking the 5% and 95% quantiles of the posterior samples and drawing two shapes in the plate based on them. The first shape in Figure 8.9 is a rectangle and shows the crack center uncertainty. This rectangle encloses 90% of the crack centers of the posterior samples (cracks) and makes up the majority of the area where the crack center could be located. The full set of the 1000 crack center samples is also shown by projecting the samples into the 2d parameter space (X_c, Y_c). The second shape shown in Figure 8.10 is composed of two circular arches, one small and one big radius that quantify the uncertainty in the crack length and orientation. The full set of the 1000 crack length-orientation samples are also shown by projecting the samples into the 2d parameter space (L, ϕ). In contrast to the 90% interval of the crack center, the position of the second shape in the plate is irrelevant, since only its size is important. However, a convenient location of the center of the arches was selected to be the bottom left corner of the first shape (Figure 8.9) for easy comparison of the uncertainty in the crack length with the uncertainty in the crack center location.

The figures described above give a qualitative visual assessment of the identification results. For a quantitative assessment, plots of the sample statistics as functions of the number of sensors for each of the 4 parameters are shown in Figure 8.11. The sample statistics include the mean, the most probable value and the 5% and 95% quantiles. Also the true values used to generate the simulated data are also presented in Figure 8.11.

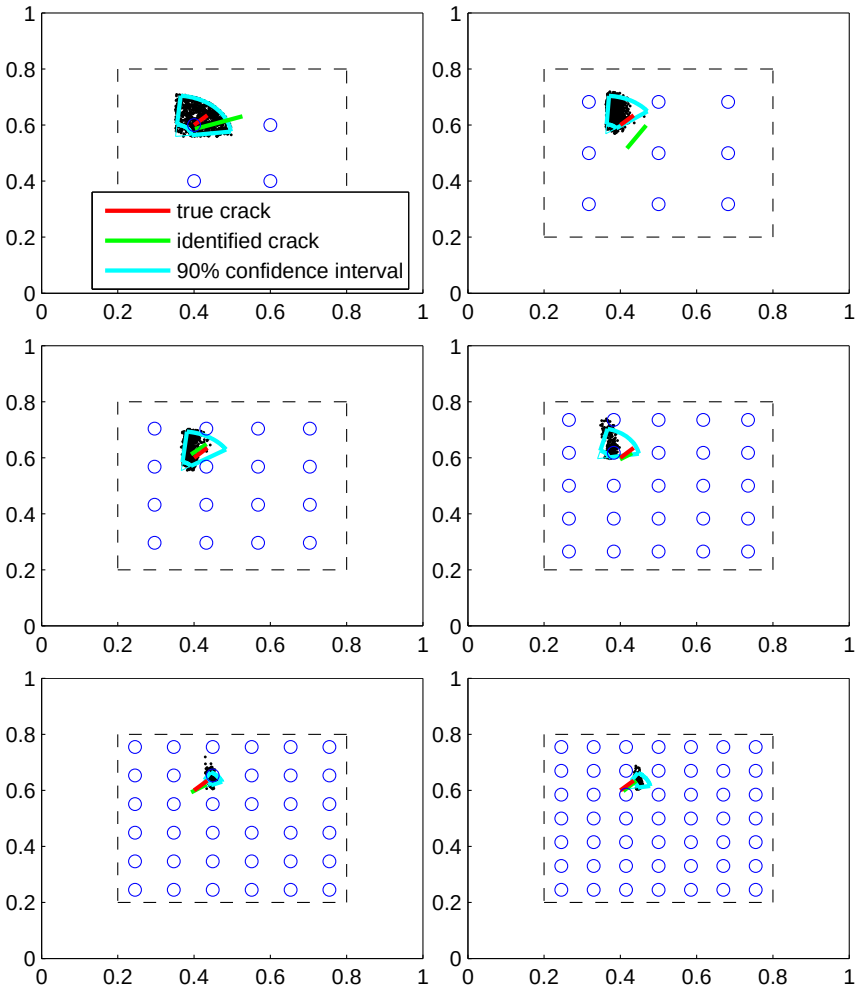


Figure 8.10: ID results for the 50mm crack with optimal full square grids from 2x2 to 7x7: Length and orientation uncertainty.

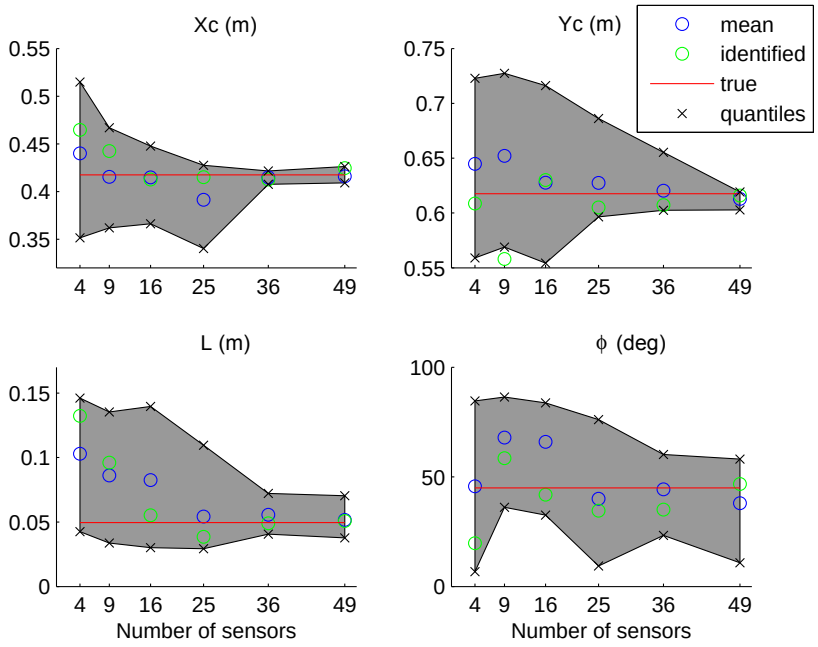


Figure 8.11: ID results for the 50mm crack with optimal full square grids from 2x2 to 7x7: Sample statistics vs number of sensors.

Sufficiently accurate estimates are obtained for sensor grids with N values as low as $N = 4$ (16 sensors). In particular, it is evident that for $N = 4$ the most probable crack (location and size) is very close to the actual (nominal) crack. However, the uncertainty intervals are quite large for both the crack center and length-orientation parameters. The uncertainty intervals shrink in size as more sensors are added, and the identified crack moves closer and closer to the parameters of the nominal $50mm$ crack. This is the expected behavior from the Bayesian theory of identification which promises improvement of the results as the number of data is increased. One can conclude that with the 6×6 full square grid the $50mm$ crack gets identified with small uncertainty. In general, the results in Figure 8.11 suggest that the uncertainty in the size and orientation is higher than the uncertainty in the crack center location.

Similar results for the larger crack of $100mm$ are presented in Figures 8.12, 8.13 and 8.14. Comparing with Figures 8.9, 8.10 and 8.11 it is clear that the identification of the $100mm$ crack is effective with less number of sensors than the ones required for the $50mm$ crack.

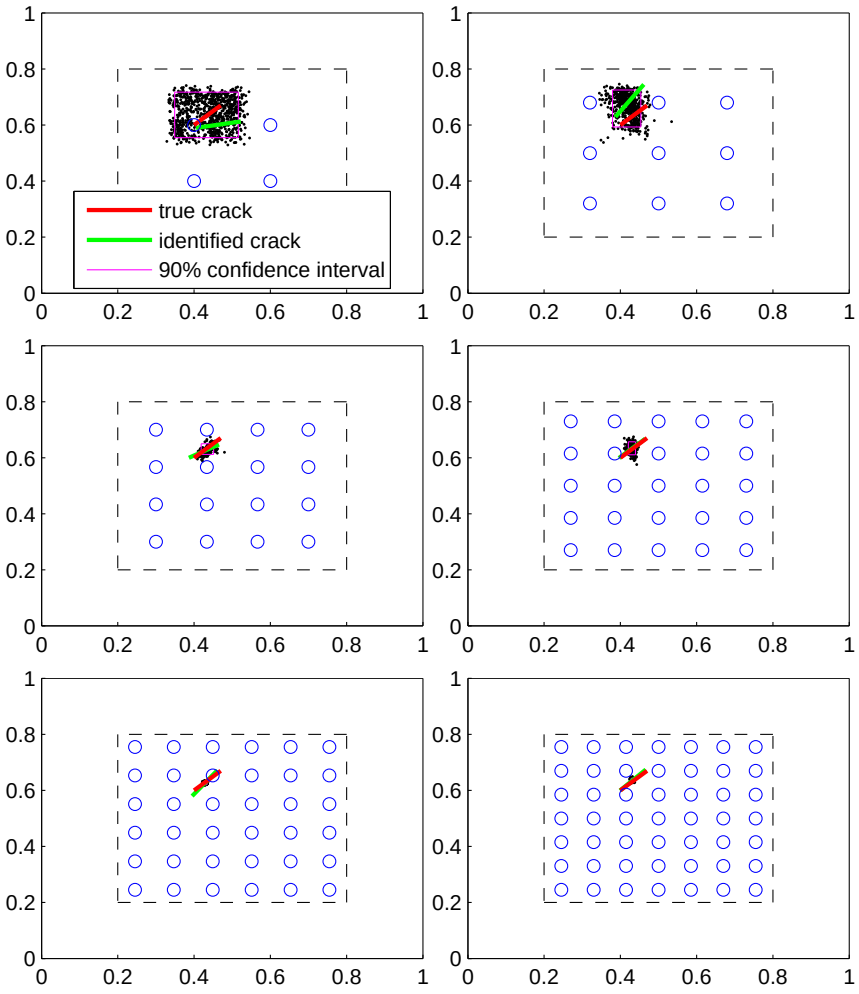


Figure 8.12: ID results for the 100mm crack with optimal full square grids from 2x2 to 7x7: Crack center uncertainty.

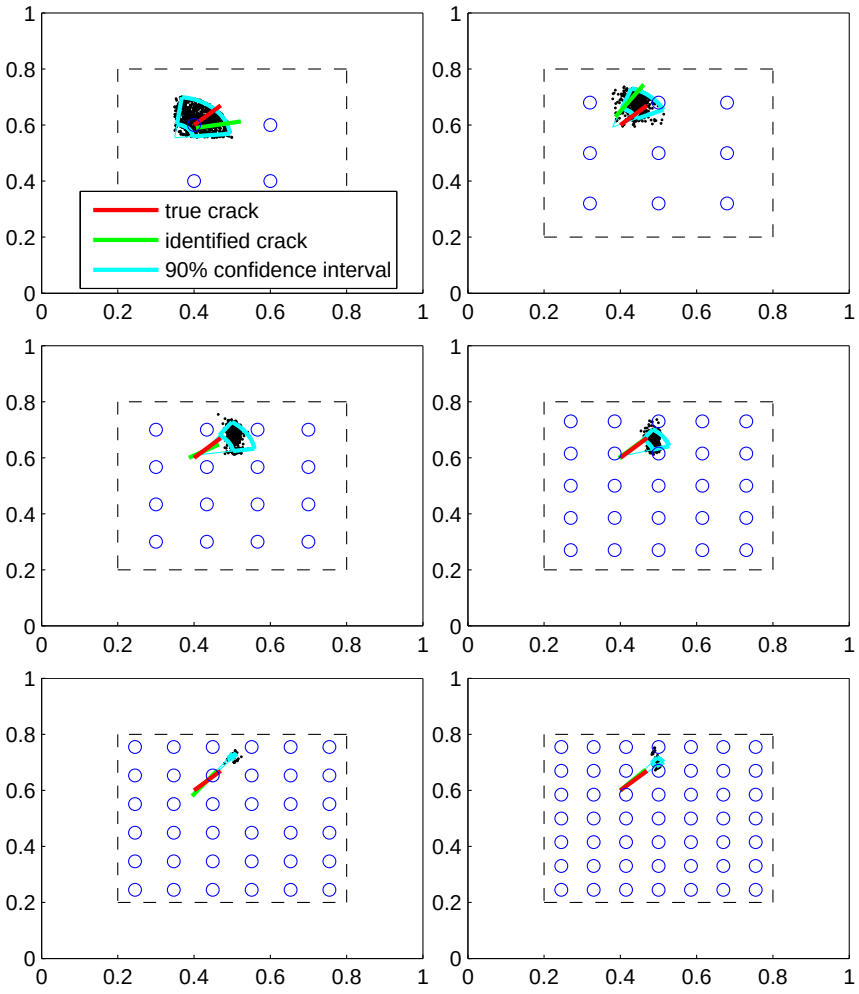


Figure 8.13: ID results for the 100mm crack with optimal full square grids from 2x2 to 7x7: Length and orientation uncertainty.

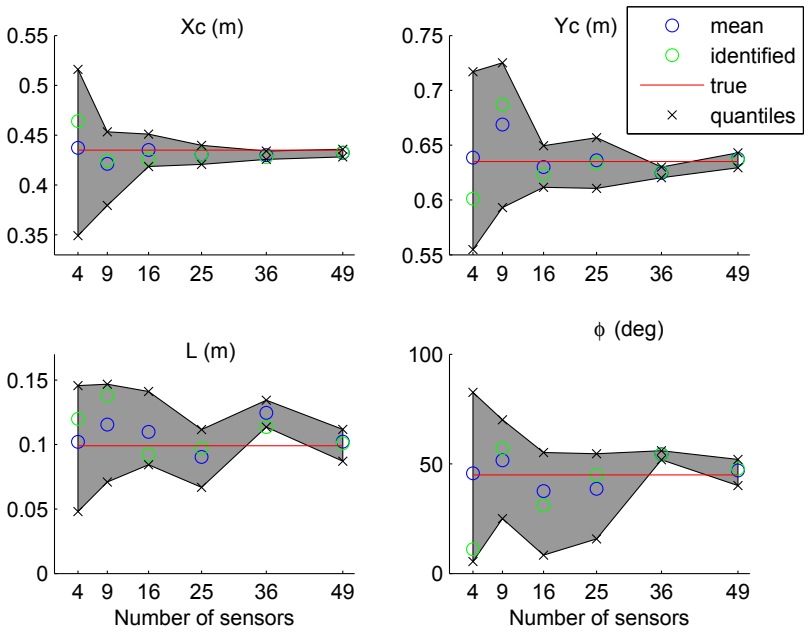


Figure 8.14: ID results for the 100mm crack with optimal full square grids from 2x2 to 7x7: Sample statistics vs number of sensors.

The $100mm$ crack is identified perfectly even from the 5×5 grid. Also from the 6×6 grid there is practically no uncertainty even in the length-orientation parameters, compared to the $50mm$ case where the uncertainty was larger. It is instructive to compare the 4×4 and 5×5 sensor grids for the $50mm$ and $100mm$ crack cases as far as the uncertainty intervals are concerned. This confirms the intuitive notion that a larger crack is easier to identify accurately with fewer sensors than a smaller crack. The $50mm$ crack required a 7×7 full grid (49 sensors) to pinpoint exactly its location, whereas the $100mm$ crack required a 5×5 grid (25 sensors, almost half those for the $50mm$ crack).

Summarizing the results of Figures 8.9 - 8.14 the following conclusions can be derived: 1) The quality of crack identification increases as the number of strain sensors increases. This applies to both the most probable identified crack and the uncertainty intervals for the crack parameters. 2) For the $50mm$ crack the 7×7 full square grid was required to identify the crack with small uncertainty whereas for the $100mm$ crack the 5×5 grid was sufficient. 3) The crack size and orientation parameters are harder to identify than the crack location parameters as seen by comparing their respective uncertainty intervals. The hardest parameter to identify is the crack orientation, since out of the 4 parameters this is the one with the greatest posterior uncertainty, especially for the $50mm$ crack. 4) The rate with which the posterior uncertainty decreases by the addition of more sensors is notably higher for the $100mm$ than for the $50mm$ crack. This shows that a larger crack is easier to identify.

Next, results are presented in Figures 8.15, 8.16 and 8.17 using the boundary square grid on the

50mm crack. Results should be compared to those obtained for the full square grid in Figures 8.9, 8.10 and 8.11.

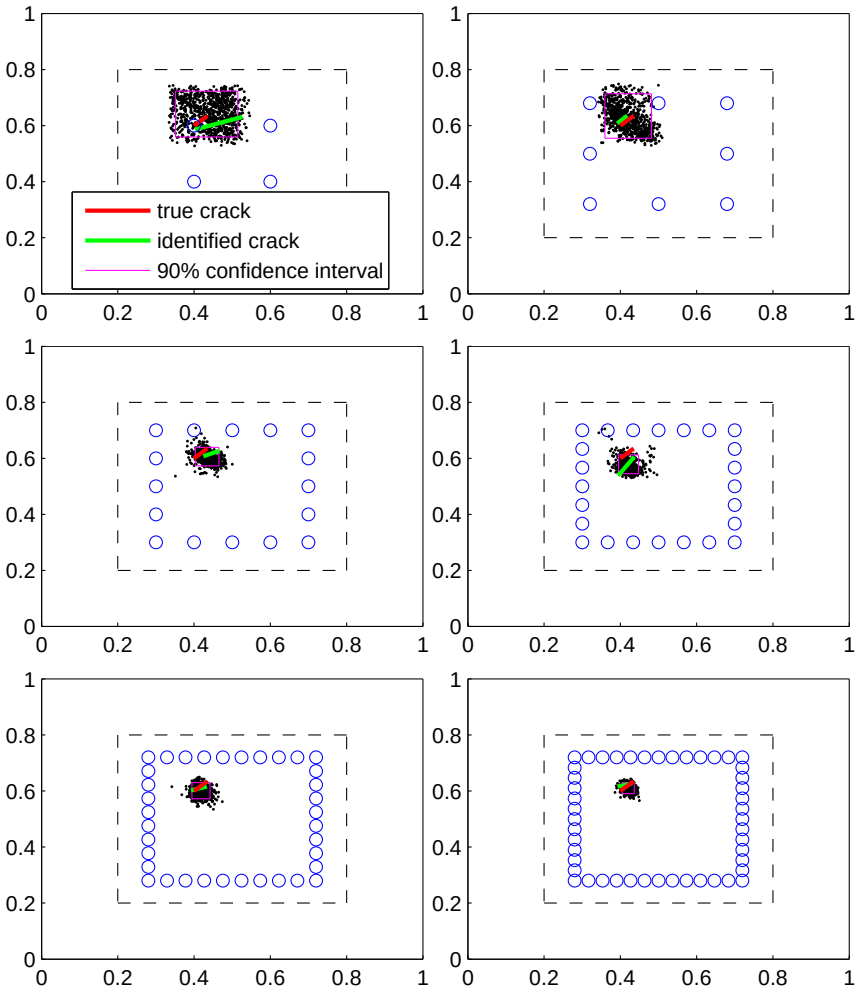


Figure 8.15: ID results for the 50mm crack with optimal boundary square grids from 2x2 to 13x13: Crack center uncertainty.

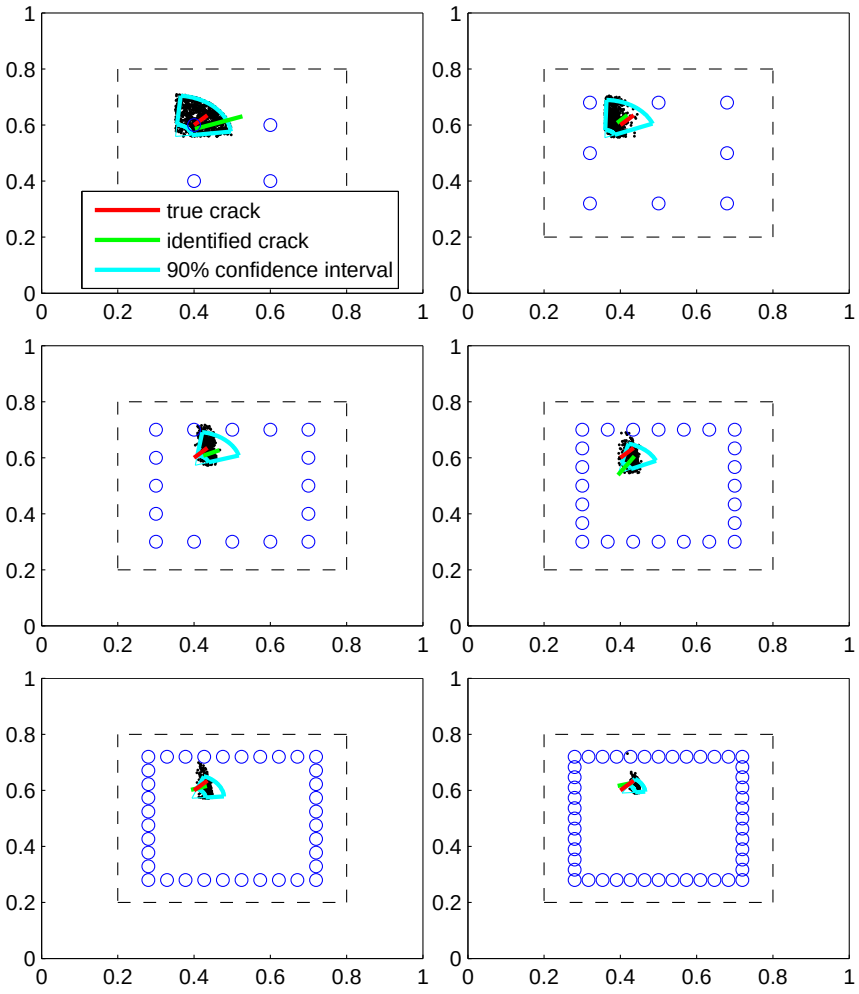


Figure 8.16: ID results for the 50mm crack with optimal boundary square grids from 2x2 to 13x13: Length and orientation uncertainty.

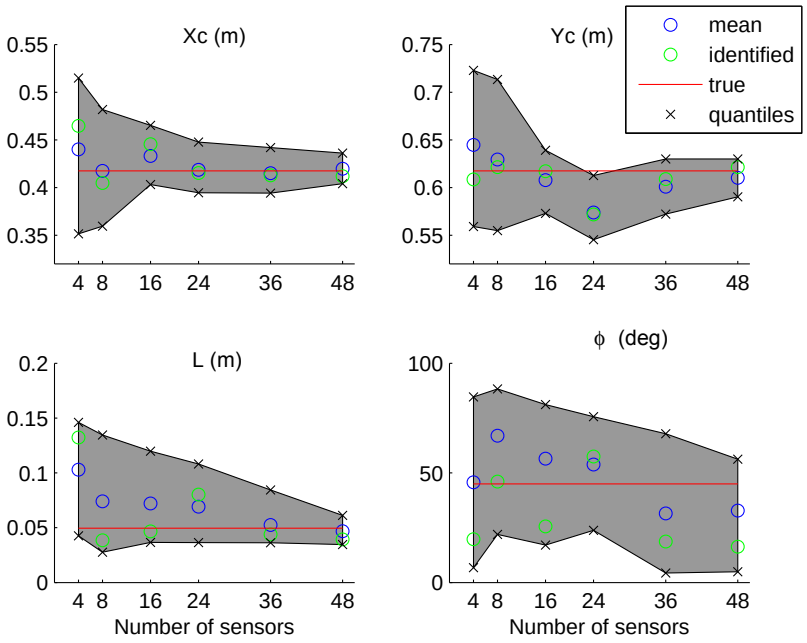
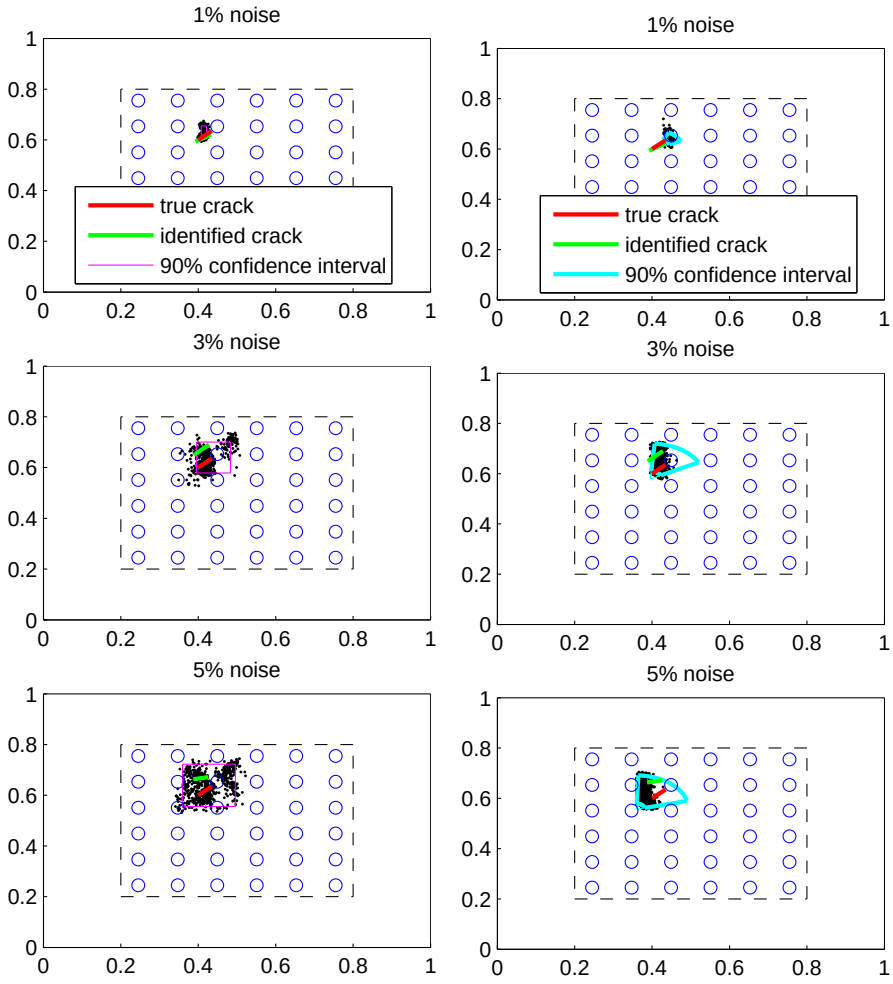


Figure 8.17: ID results for the 50mm crack with optimal boundary square grids from 2x2 to 13x13: Sample statistics vs number of sensors.

Here the boundary square grid was tested using the same number of sensors as in the full square grid case. Again, the grid sizes used were the optimal ones found for the specific number of sensors. In general, no significant differences appear between the identification from the boundary square and full square grids. However, the boundary square grid gives slightly worse results in terms of the crack location, which is expressed by a greater crack center uncertainty compared with the full square grid case. This shows that for the particular crack examined the two grids have almost equal performance, with the full square grid slightly outperforming the boundary square grid. Although it was shown that the full square grid is superior to the boundary square grid on average, over all possible cracks in the domain Ω, for the particular crack considered in this study, only a slight preference of the full square grid over the boundary square grid was observed.

Effect of measurement and model error level

Now the effect of the measurement and model error level on the identification results is investigated. Naturally one would expect that an increasing noise level would lead to a quality reduction in the identification. In order to demonstrate the effect of the noise, three different noise levels are used: 1%, 3% and 5%. In the following respective results the 6×6 full square grid was applied to the identification of a $50mm$ crack. The identification results for these cases are presented in Figures 8.18 and 8.19.



(a) (b)

Figure 8.18: ID results for the 50mm crack with optimal 6x6 full square grid. Left: Crack center uncertainty, Right: Length and orientation uncertainty.

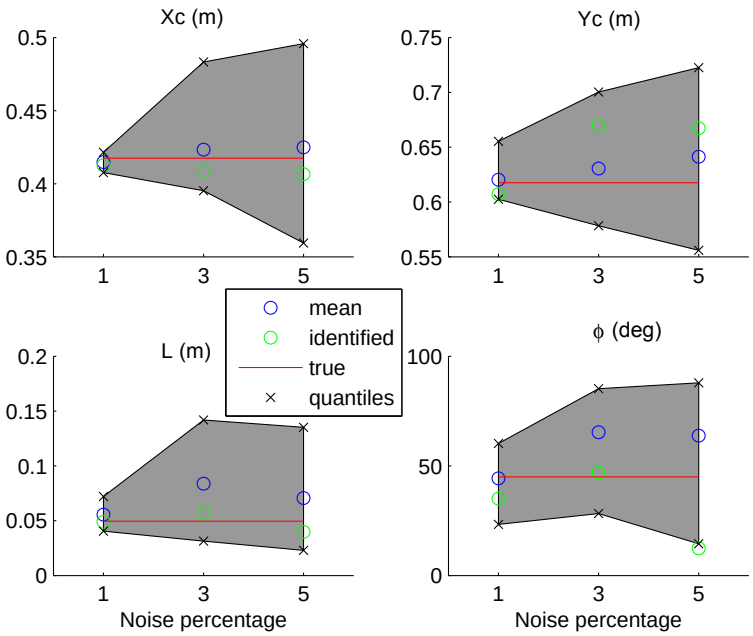


Figure 8.19: ID results for the 50mm crack with optimal 6x6 full square grid: Sample statistics vs noise level.

It is apparent from the results that the crack identification is sensitive to measurement/model errors since an increase from 1% to 5% noise has a significant impact on both the most probable crack parameters and the uncertainty intervals. In Figure 8.19 one can see the anticipated result that the uncertainty intervals increase as the noise in the data is increased. The quality of the identification starts to deteriorate rapidly, both for the optimal identified crack and its associated uncertainty.

8.5 Conclusions

A Bayesian optimal sensor placement methodology was presented for optimally designing the locations of strain sensors in a structure for the purpose of identifying a crack using strain measurements. The optimal locations of the strain sensors were derived based on the expected information gain (expected KL-divergence) quantity which measures the information gain from the prior to the posterior PDF in a Bayesian framework. Three types of sensor grids were examined and compared. A full square grid, a boundary square grid and a circular grid were compared. The full square grid turned out to be the best for identifying cracks in a square sub-domain of the plate. Optimal designs were obtained for an increasing number of sensors for all three grid types. It was shown how the optimal design can change when different a priori knowledge about the crack location is taken into account, through the Bayesian prior PDF. Also the prediction error correlation model was examined. It was demonstrated how including correlation

between the strain sensors results in less information gain from the measurements when increasing the number of sensors.

The effectiveness of the optimal sensor locations was investigated by estimating the location, size and orientation of a crack using a Bayesian parameter estimation (identification) methodology utilizing strain measurements. The methodology was used to estimate the location, size and orientation of cracks using information provided from strain measurements of optimally located sensors from a cracked plate subjected to distributed static loading. Because of the Bayesian nature of the identification, the uncertainty associated with the identified crack is also estimated. The results showed that the proposed identification methodology can efficiently detect and completely identify an existing crack using simple grids of sensors, even in the presence of measurement and model error, provided that this error is sufficiently small.

Parametric analyses were performed in order to study the effect of the number of sensors and noise level in the quality of the identification results. These analyses provided useful insights about the minimum number of sensors required to identify cracks of small and medium sizes accurately. It was also shown how the addition of more sensors improved not only the identified crack but its uncertainty as well. Results also verified the intuitive idea that a larger crack is easier to identify than a smaller crack, and this was quantified by the required number of sensors. It was shown how the crack orientation is the hardest parameter to identify compared to the size and location parameters. The effect of the noise level in the experimental data was also investigated. Results showed that the analysis is very sensitive to noise and an increase could deteriorate the identification quality significantly.

Chapter 9

Conclusions

The main theme of this thesis was to develop Bayesian methodologies for uncertainty quantification, propagation, model selection and optimal experimental design and apply them to real world engineering problems. The Bayesian methodologies for uncertainty quantification were applied in model updating problems of mechanics using experimental data obtained from field tests. Specifically, model updating was performed in a high-fidelity detailed finite element model of a bridge, where experimentally identified modal data were used to update the existing finite element model. The updating process refers to the parameters of the model which were the stiffnesses of the deck, piers and that of the soil. Bayesian parameter inference enabled us to learn about the possible values of those stiffnesses compared to where we would expect them to be, and also allowed us to check the effect of different modeling scenarios regarding the use of the modal data. The second application of model updating using real data was in a hanger of an arch bridge, where model updating allowed us to find the tension force that was developed in the hanger based on experimentally identified modal frequencies obtained from forced vibration tests. The Bayesian methods proved to be valuable in order to understand the bounds where the tension force could lie and the capability of different models to make accurate predictions. A third case of model updating was carried out this time using a non-linear model of a seismically isolated bridge. The Bayesian formulation was applied using experimentally measured time histories of various response quantities of the bridge-isolator system. The parameters of the non-linear model equations were identified along with their uncertainties, enabling for robust model predictions. Finally, a fourth model updating case was carried out using simulated data, in a finite element model of a plate with a crack. The methodology proved to be successful in detecting the crack using synthetic strain measurements, and also allowed us to investigate the effect of the number of sensors and noise levels in the crack detection results.

Bayesian model selection proved to be a valuable tool in understanding which models are the best in simulating the behavior of a structural component and which aspects of the model really matter, based on experimental data. Specifically, model selection was used to rank several model classes of a hanger cable and decide which is the best by utilizing information from measured modal frequencies of the hanger.

It turned out that the flexibility of the support conditions of the hanger were extremely important to its behavior and models with fixed supports were ruled out by Bayesian model selection, while flexible models were favored.

In the problem of optimal experimental design, important theoretical contributions were made to both the sampling and asymptotic approaches to estimate the objective function which provide further insight into the problem. Numerical case-studies demonstrated the methodologies and verified the theoretical findings. Optimal sensor locations were found for up to 20 acceleration sensors to be placed on the deck of a bridge for modal identification. The effect of the prediction error model on the sensor locations was found to be crucial, and the novel theorem about the effect of the Gaussian prior distribution was demonstrated numerically. Also, optimal strain sensor grids were designed for crack detection in a cracked plate under static loading. The methodologies proved to be useful in comparing the various sensor grids with each other and finding the optimal grid size and location in the plate in order to identify the crack which can exist in a specified domain in the plate.

Summarizing, the novel contributions of this thesis are:

1. Use of non-linear detailed finite element models to simulate hanger behavior under tensile loads, which fully take into account the support conditions of the hanger.

2. Comparison of complex finite element models of hangers and simple beam models using the concept of equivalent length in each of the two vibrational directions of the hanger.

3. Use of high-fidelity detailed finite element models of bridges which take the soil flexibility into account.

4. Use of Bayesian model selection in order to rationally decide about the importance of the soil flexibility.

5. Utilization of model-reduction techniques in order to make the Bayesian framework computationally feasible.

6. Application of the Bayesian theories for the estimation of parameters of non-linear models of seismically isolated bridges.

7. Theorem in the field of asymptotic Optimal Experimental Design which explains the effect of the Gaussian prior variances in the optimal design. The variances are proved to act as weighting factors for the identification of parameters with the largest variances. Therefore, the optimal design favors the identification of those parameters instead of treating them all equally with the same weight.

8. Improved estimator of the objective function of sampling Optimal Experimental Design. The estimator was simplified by calculating one term analytically, hence avoiding the need to approximate it with Monte Carlo sampling. This effectively leads to reduced variance (better quality) of the estimator.

9. New evidence-based interpretation of Optimal Experimental Design based on the improved estimator. The optimal design can be interpreted as the one which maximizes the entropy of the evidence. This means that the best design is the one which gives a good fit with the data on average over the uncertain data that may arise from each design.

10. Novel construction of the design variables in Optimal Sensor Placement which includes the type of sensor as well instead of its location only. With this formulation it is possible to include the type of sensor (orientation, measured quantity etc.) into the formulation as well and let the methodology decide what is the optimal sensor type and in which location.

11. Application of the novel theorem developed in Chapter 6 which demonstrates the role of the Gaussian prior variances as weighting factors for the corresponding parameters.

12. Comparison of sensor grids of various geometries for the purpose of Optimal Sensor Placement for crack identification using strain measurements.

Some directions for potential future research are:

- Improve the computational tools required to do Bayesian analysis, specifically stochastic simulation algorithms or methods to estimate multidimensional probability integrals efficiently.

- Improve or develop ways to estimate the evidence term of a Bayesian model efficiently and reliably, which is crucial for Bayesian model selection.

- In the problem of optimal experimental design, introduce new estimators of the objective function so that sampling is not required, or find ways to reduce the variance associated with the random samples required by Monte Carlo numerical integration.

- In the asymptotic approach of optimal experimental design find ways to make it work reasonably well even for larger prediction error without compromising the accuracy too much.

- If possible, derive some of the theoretical results of the asymptotic optimal experimental design framework (such as the weighting role of the Gaussian prior variances) under the sampling framework.

- Further research can be done regarding the way in which mode shape data are taken into account into the Bayesian model updating scheme. Correlation models can be used to introduce correlation between the components of the mode shape, or different probability distributions can be used to describe the quantity $1 - MAC_r^2(\underline{\theta})$.

Appendix A: Software for Experimental Modal Analysis

In this Appendix the software developed for experimental output-only (ambient) modal analysis is presented. The software is written in MATLAB and consists of four distinct and independent modules, namely: Data, Pre-processing, Modal Identification and Post-processing. The main menu of the program with the 4 modules is shown in Figure A.1.

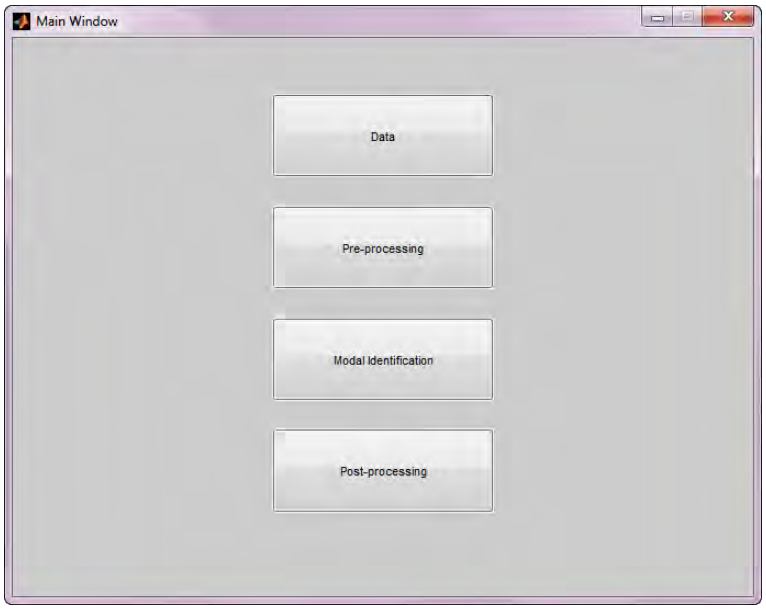


Figure A.1: Main window of the program.

The Data module is used to load data from .mat files into the software. The data can be:

1. The experimentally measured ambient acceleration time histories of multiple sensors measuring at the same time. The time histories of multiple sensors need to be arranged in the columns of a matrix and stored into a .mat file along with the discretization time and the sensor labels. Multiple

sets of measurements (.mat files) can be inserted into the program. This is the case when we are measuring a large structure with a small number of sensors and have to take repeated measurements at different locations of the structure (sensor configurations). An example of a .mat file containing the simultaneous measurements of 18 sensors is shown in Figure A.2.

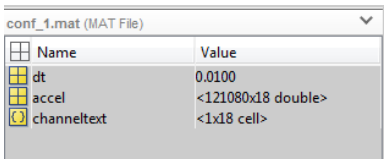


Figure A.2: Variables in a measurement .mat file: accel = matrix of simultaneous acceleration measurements in the columns, dt = time between measurements (inverse of sampling rate), channeltext = label of each sensor that corresponds to the columns of the accel matrix.

2. Modal identification results that had been saved from previous sessions and need not be produced again from measurements. In this case the user works directly in the Post-processing module.

3. Variables defining the geometry of the measured structure for the purpose of visualizing the mode shapes that result from modal identification. The geometry is a simple sketch of the structure that consists of points (nodes), directions (DOFs) and lines (elements) connecting the points. The matrices contained in the geometry .mat file describe the coordinates of the nodes, the node DOFs and the element connectivity of the elements with the nodes. An illustrative example of a geometry .mat file is shown in Figure A.3.

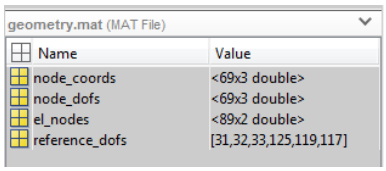


Figure A.3: Variables in a geometry .mat file: node_coords = node (X,Y,Z) coordinates in rows, node_dofs = DOFs in (X,Y,Z) directions of each node, el_nodes = nodes of each element, reference_dofs = common reference measured DOFs.

The Data module can also be used to convert measurements from another format (e.g. text files) to appropriate .mat files that can be inserted into the program. The Data insertion module is shown in Figure A.4.

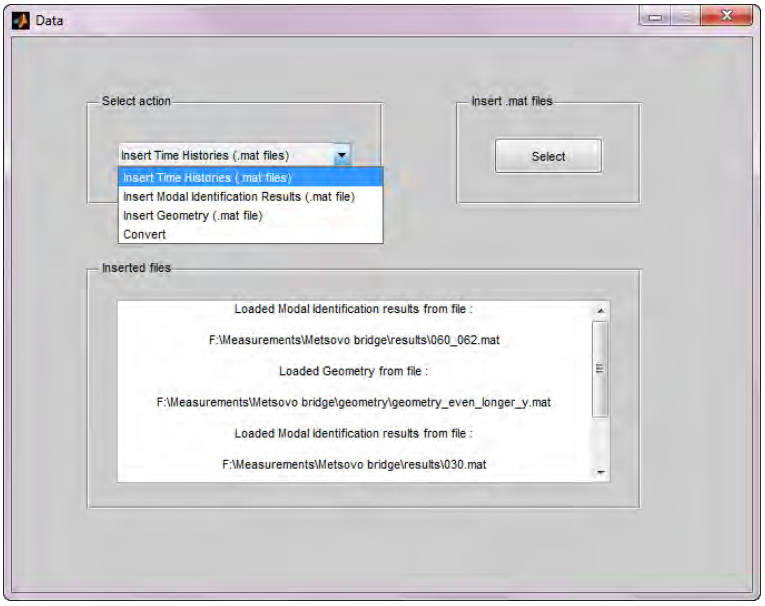


Figure A.4: Data insertion module.

In the Pre-processing stage the user can visually inspect the Power Spectral Density (PSD) and the Singular Value Spectrum (SVS) of the ambient acceleration time histories (Figures A.5 and A.6 respectively). Such visual inspection can provide valuable insight about the modal frequencies and damping ratios of the structure. It is important to note that the PSD and SV spectrums are not utilized in the Modal Identification procedure, rather they are only used to obtain a rough estimate of the natural frequencies of the measured structure. This step is important in the Modal Identification methodology used in the next step. Each sensor configuration can be inspected individually in the pre-processing step, and specific channels of a configuration can be selected or de-selected from being used in the Modal Identification process. This feature serves to potentially remove an unwanted sensor from the analysis because of possible bad recording quality. The method used to estimate the PSD of the signals is Welchs averaged periodogram method, and the parameters of the method can be adjusted by the user depending on the length of the signal, the sampling rate, the desired frequency resolution in the spectrum, and the desired variance of the estimated spectrum. The SV spectrum is obtained by calculating the singular values of the Cross Power Spectral Density matrix for each frequency in the spectrum. It provides more information than the PSD because it also takes into account the Cross Power Spectral Densities between the acceleration signals. Its main merit comes from the fact that it has the ability to separate the noise from the signal, and that it can reveal closely spaced modes that are not apparent in the PSD. This was proven to be the case for the second and third modes of the Metsovo bridge which were revealed by different singular values, as can be seen in Figure A.6.

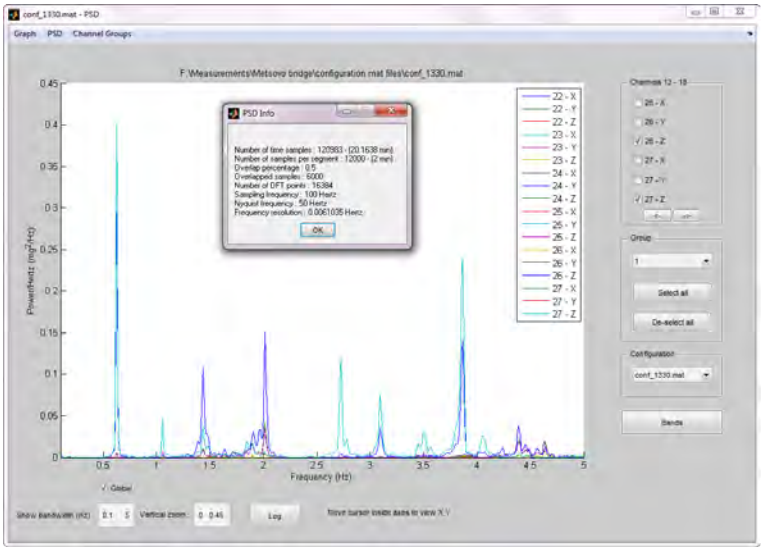


Figure A.5: Pre-processing: Auto Power Spectral Densities of multiple selected signals from a single sensor configuration, along with PSD algorithm information.

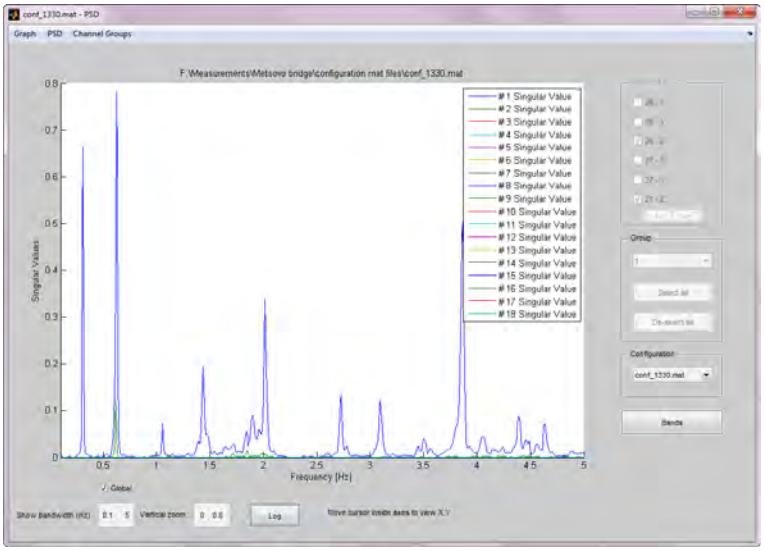


Figure A.6: Pre-processing: Singular Value Spectrums of multiple selected signals from a single sensor configuration.

The raw time histories of the ambient acceleration measurements can also be viewed for each of the inserted sensor configurations, as shown in Figure A.7.

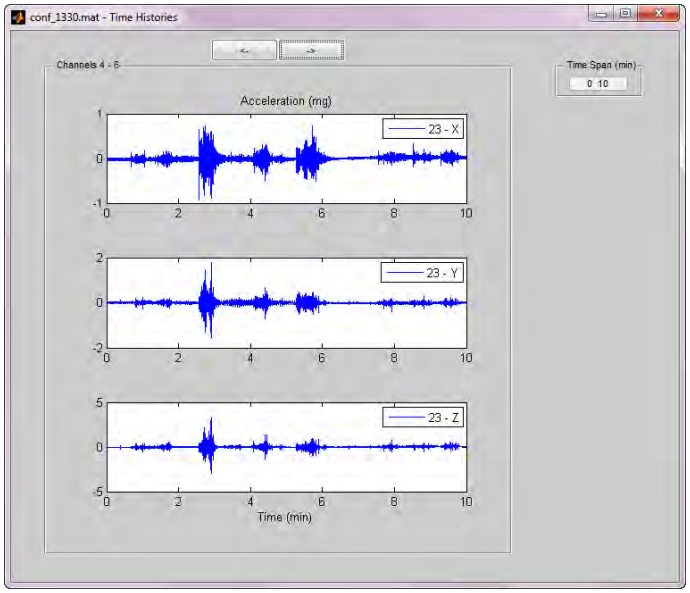


Figure A.7: Raw ambient acceleration time histories.

After obtaining an estimate of the natural frequencies of the structure from observing the Singular Value spectrum, the user can define the frequency bands which most probably contain the natural frequencies of the structure (Bands button in Figures A.5 and A.6). These frequency bands are going to be used by the Modal Identification module. The bands can be saved to be re-loaded in future sessions (Figure A.8).

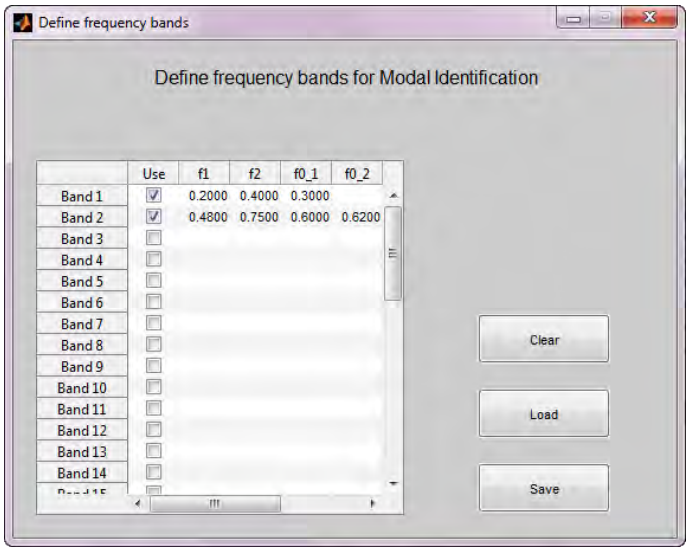
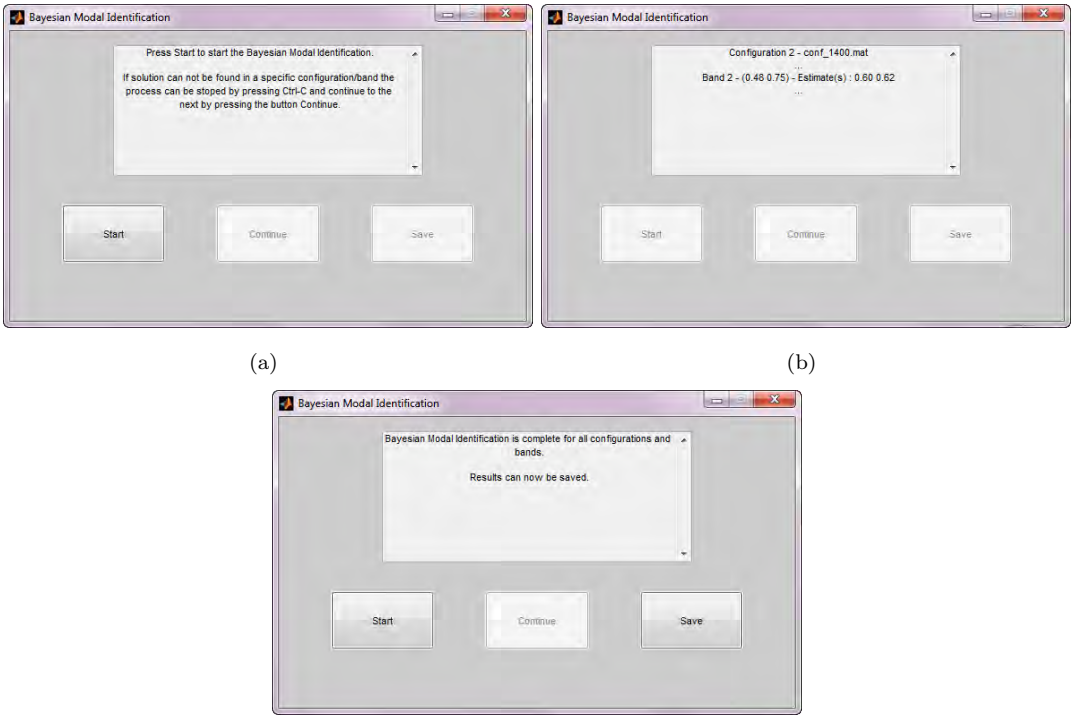


Figure A.8: Definition of frequency bands for Modal Identification.

The Modal Identification module uses a Bayesian methodology (Au, 2012a,b) in order to extract the modal frequencies, mode shapes, and modal damping ratios from the measured ambient acceleration time histories of each sensor configuration. The method is based on the Fast Fourier Transform (FFT) of the acceleration signals in specific bands of interest which are believed to contain the natural frequencies of the structure. The Bayesian nature of the methodology also provides the uncertainty in the estimates of the modal properties which can be thought of as experimental uncertainty. This information can be useful for model updating purposes using experimental modal properties, since it can separate the measurement error from the model error. Pictures from the Modal Identification module are shown in Figure A.9.



(a) (b)

(c)

Figure A.9: Modal Identification module.

After Modal Identification has been performed and the modal properties have been extracted from each configuration, the next step is to visualize the results in the Post-processing module (Figure A.10).

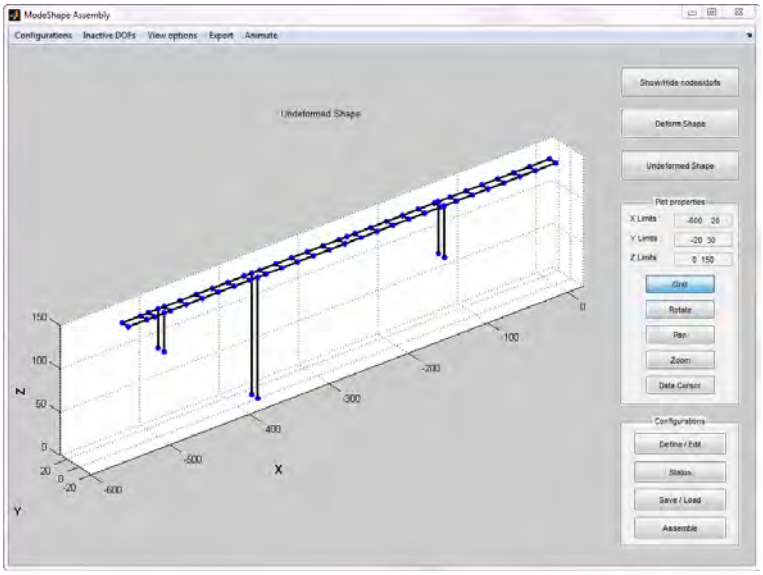


Figure A.10: Post-processing module main window.

In order to visualize the mode shapes it is necessary to combine all the local mode shapes identified from each configuration to produce the full mode shape at all measured degrees of freedom. The methodology used to achieve this is a variation of the method proposed by Au (2011) and is based on least squares fitting of each local mode shape with the full mode shape. The method is developed in detail in Section 4.4.3 and relies on common measured degrees of freedom among all configurations, called reference degrees of freedom.

For visual representation of the full mode shapes, it is also required that the user specifies the measured points of each configuration in the geometry of the structure. Therefore, it is necessary to describe at which point of the structure geometry and direction was each sensor measuring, for all configurations. Then the identified mode shape components of each configuration can be correctly matched with the appropriate point and direction in the geometry. This is done from the Define/Edit button of the Post-processing main window (Figures A.11, A.12 and A.13).

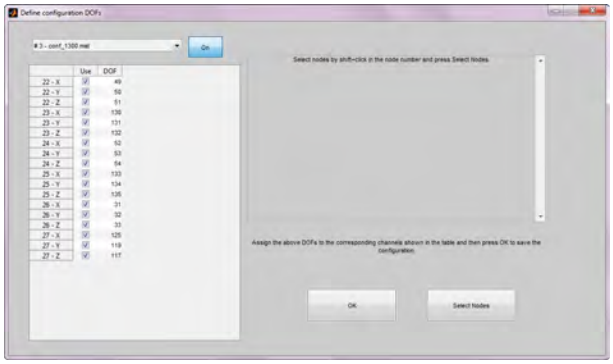


Figure A.11: Define/Edit configurations window.

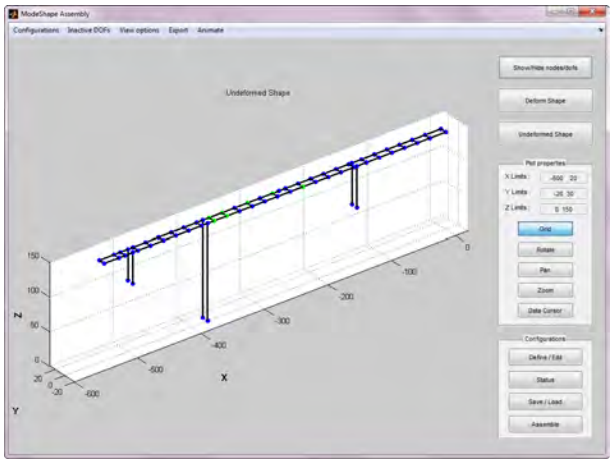


Figure A.12: Selection of geometry points where sensors were placed.

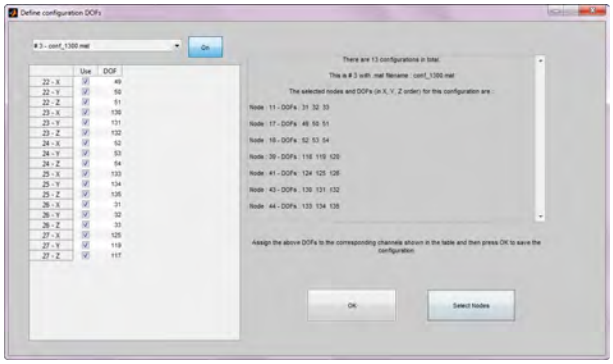


Figure A.13: Assign geometry points and directions (DOF) to each channel (sensor) for all configurations.

The user can exclude specific channels or even entire configurations if their measurements were found

to be unreliable in the Pre-processing analysis. There is also the possibility to save the channel-DOF associations for all configurations in order to load them in future sessions from the Save/Load button of the Post-processing main window. The status of all the sensor configurations can be viewed from the Status button of the main Post-processing window (Figures A.14 and A.15).

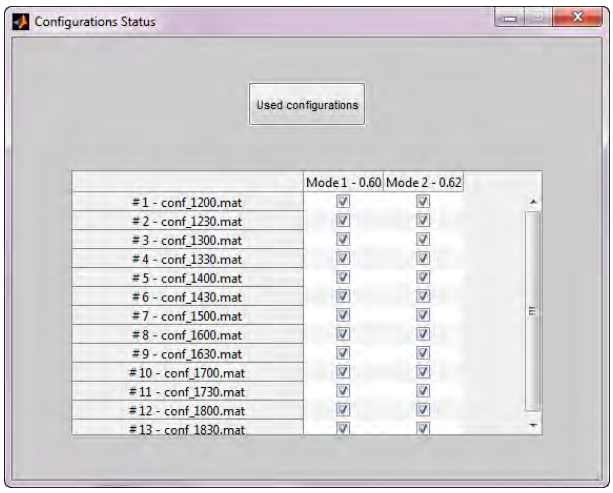


Figure A.14: Sensor configurations selected for mode shape assembly for specific modes.

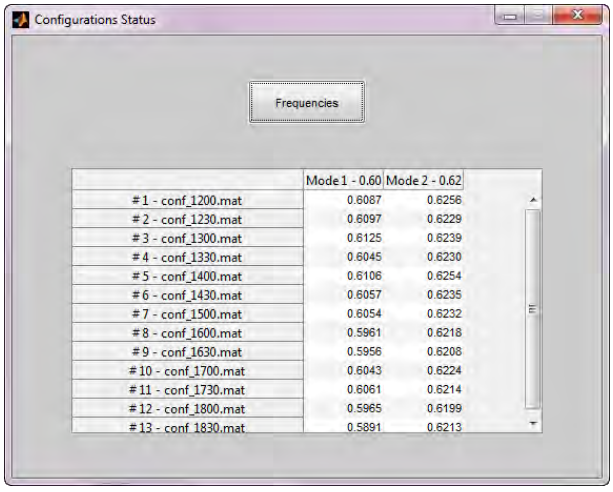


Figure A.15: Identified modal frequencies from each configuration.

After the channel-DOF associations have been defined, the mode shapes can be assembled from the Assemble button of the Post-processing main window. After the assembly of the mode shapes the user can view them from the Deform Shape button (Figures A.16, A.17 and A.18).

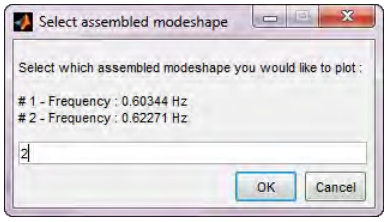


Figure A.16: Assembled mode shape selection for visualization.

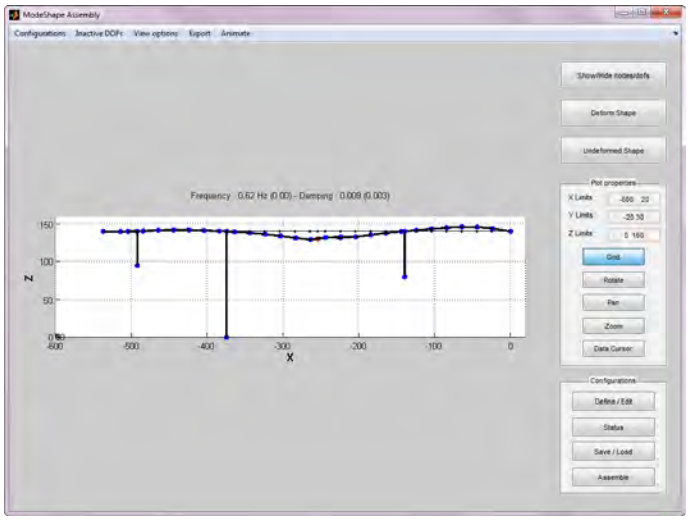


Figure A.17: First full mode shape.

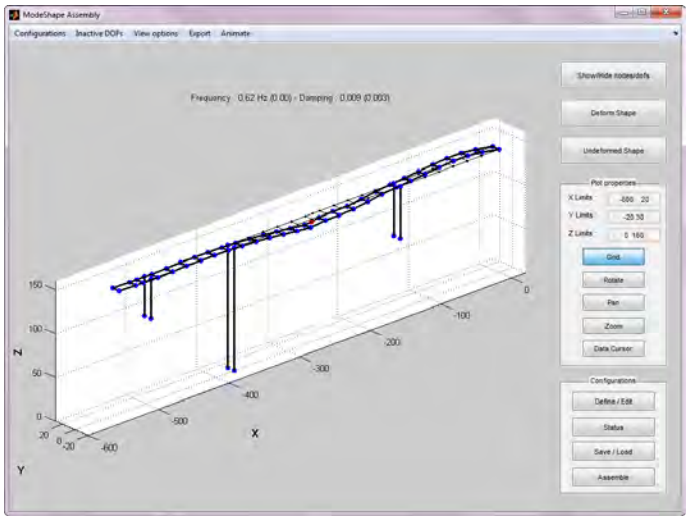


Figure A.18: Second full mode shape.

Usually not all geometry points were measured by a sensor, and those points have no associated mode shape component. However, for visualization purposes we would like to associate those points with some other measured points in order for them to deform as well. This is done from the Inactive DOFs menu of the main Post-processing window (Figure A.19).

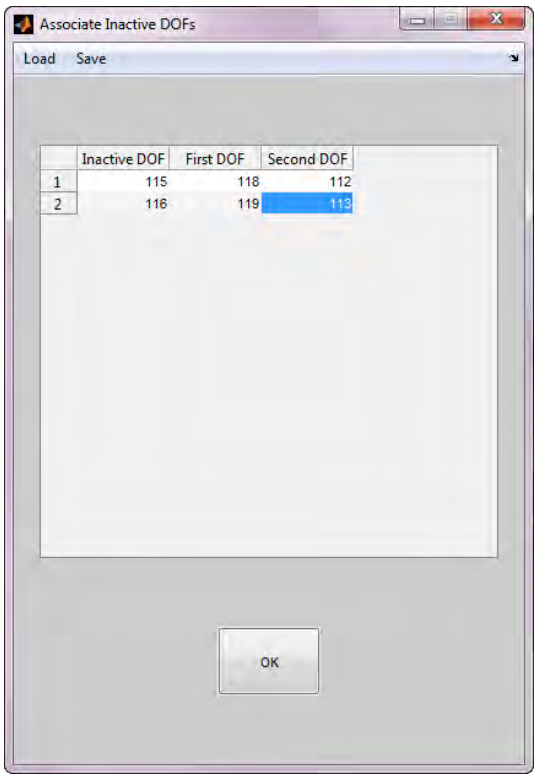


Figure A.19: Associate inactive DOFs with two active DOFs.

Then the inactive DOFs deform as the mean of the two associated DOFs. The inactive DOFs associations can be saved and loaded from the corresponding Save and Load buttons. The status of the inactive DOFs associations can be viewed from the Inactive DOFs menu of the main Post-processing window (Figure A.20).

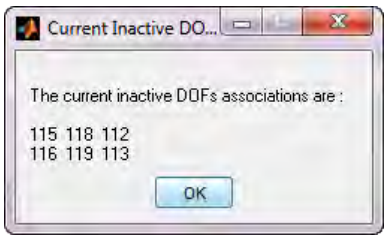


Figure A.20: Status of inactive DOFs associations.

The user can check the error of the assembled mode shapes which is the mismatch between the assembled mode shape and the identified mode shape at a particular sensor configuration. This can be done from the Configurations menu (Figures A.21 and A.22).

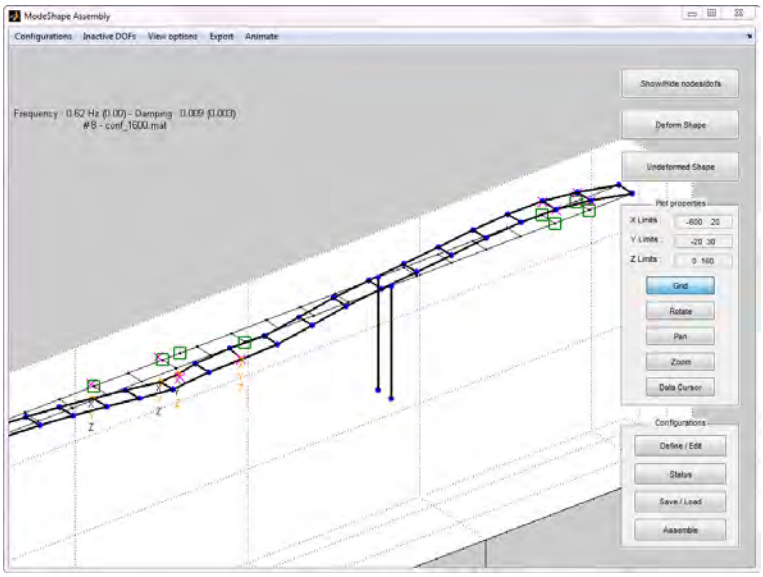


Figure A.21: Configuration 8 identified mode shape components and full assembled mode shape.

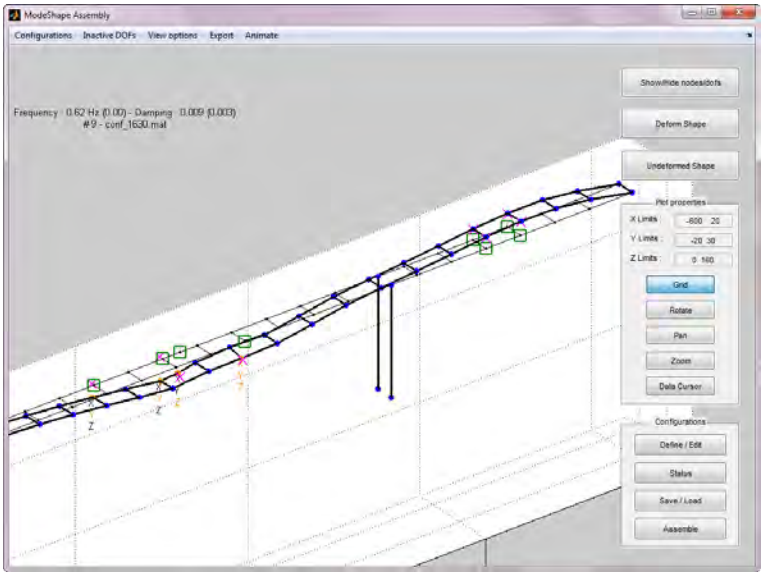


Figure A.22: Configuration 9 identified mode shape components and full assembled mode shape.

The reference sensors (common among all configurations) are marked with orange color, the locations

of the sensors of the corresponding configuration are shown with green squares, and the identified mode shape components are shown with the pink crosses.

Finally, a summary of all the results is presented from the Configurations → Details menu (Figure A.23).

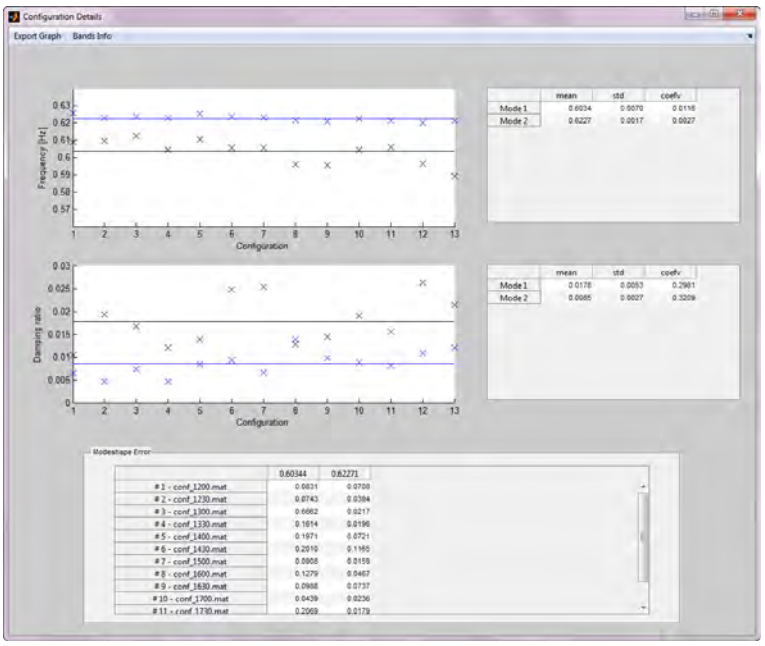
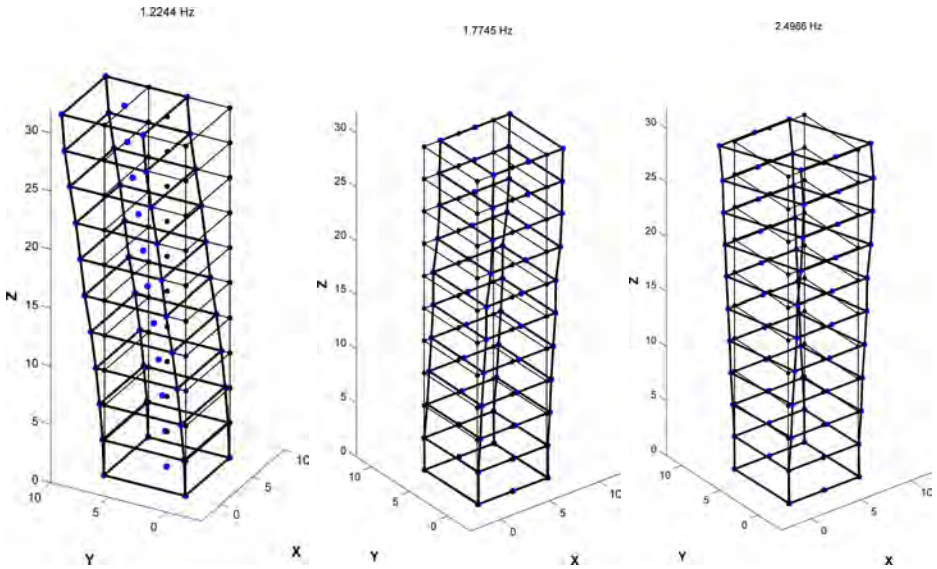


Figure A.23: Detailed results information: Statistics of identified modal frequencies and modal damping ratios over all configurations and percentage mode shape errors for all configurations and modes.

The results plots can be exported to files from the Export Graph menu. Other auxiliary post-processing actions can be performed from the menu bar of the main window, such as animate the mode shapes and export images and videos of the mode shapes to files.

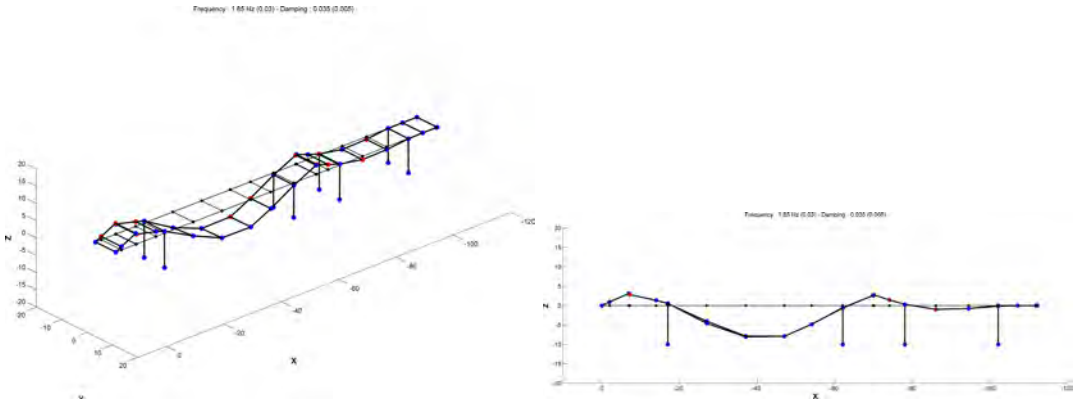
Illustrative mode shape results

In this Section we present illustrative identified mode shapes from various cases of real structures where experimental ambient acceleration data were used to perform modal identification. Representative identified assembled mode shapes are shown in Figures A.24 - A.28.



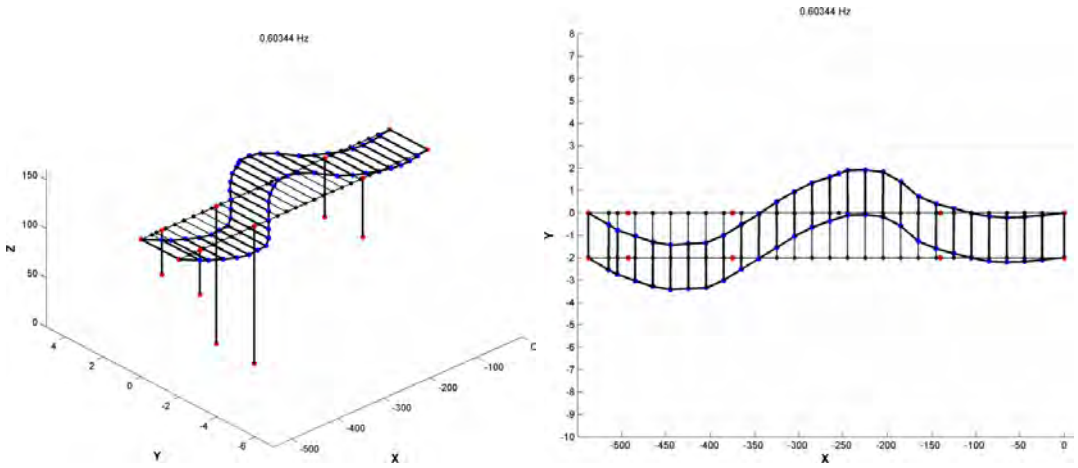
(a) First (b) Second (c) Third

Figure A.24: Experimentally identified mode shapes of a building located in Volos, Greece.



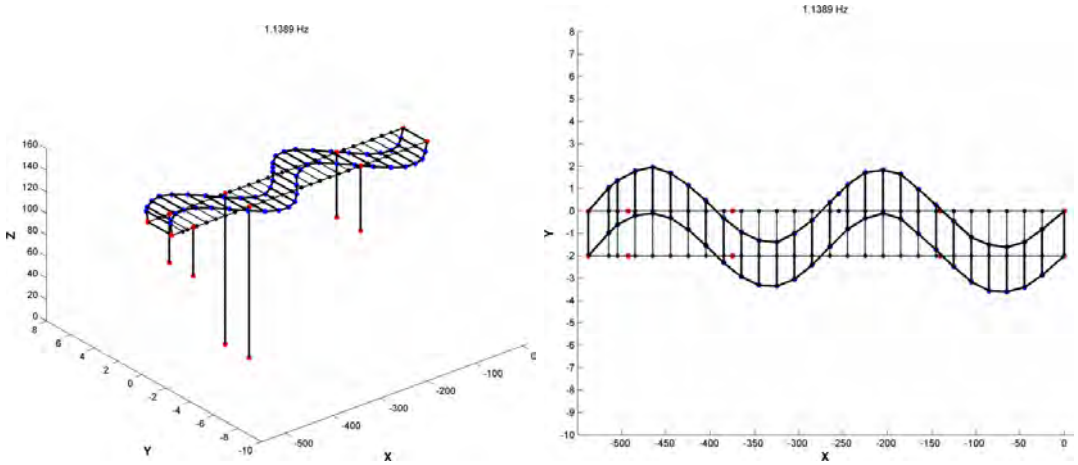
(a) 3D view (b) Side view

Figure A.25: First experimentally identified mode shape of a bridge located in Katerini, Greece.



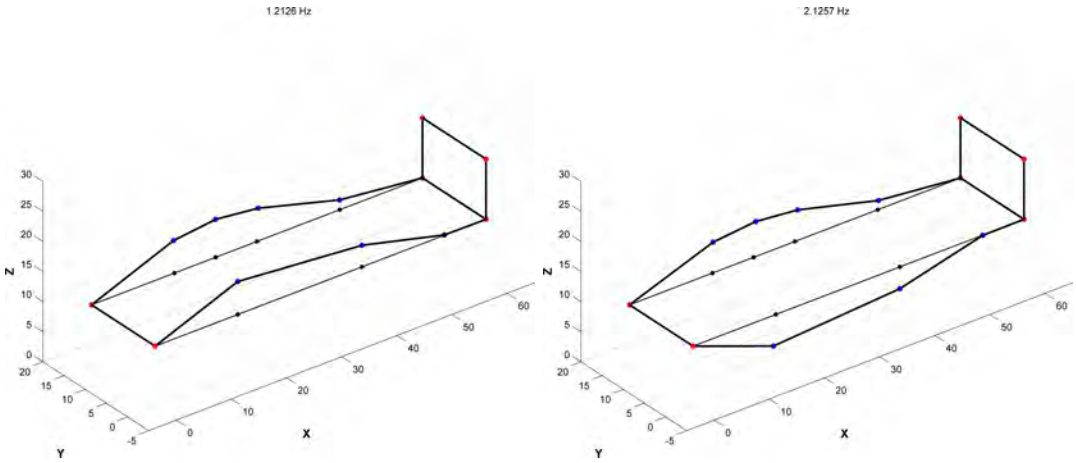
(a) 3D view (b) Top view

Figure A.26: Second identified mode shape of the Metsovo bridge.



(a) 3D view (b) Top view

Figure A.27: Sixth identified mode shape of the Metsovo bridge.

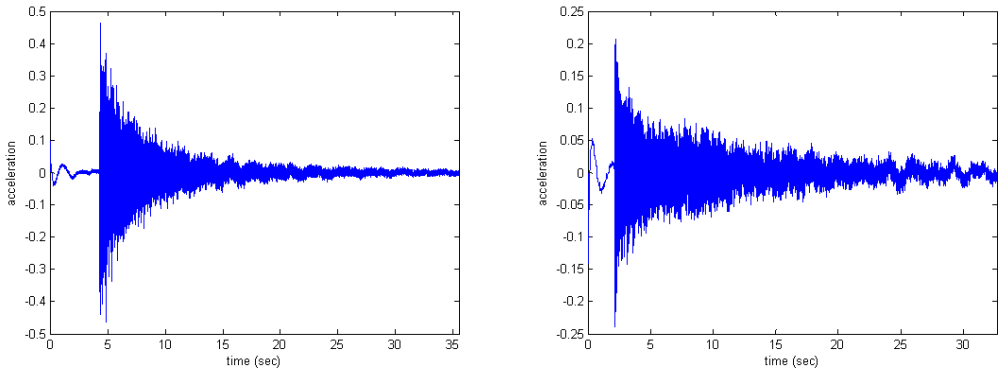


(a) First identified mode shape (b) Second identified mode shape

Figure A.28: First two identified mode shapes of a bridge in Palini, Athens, Greece.

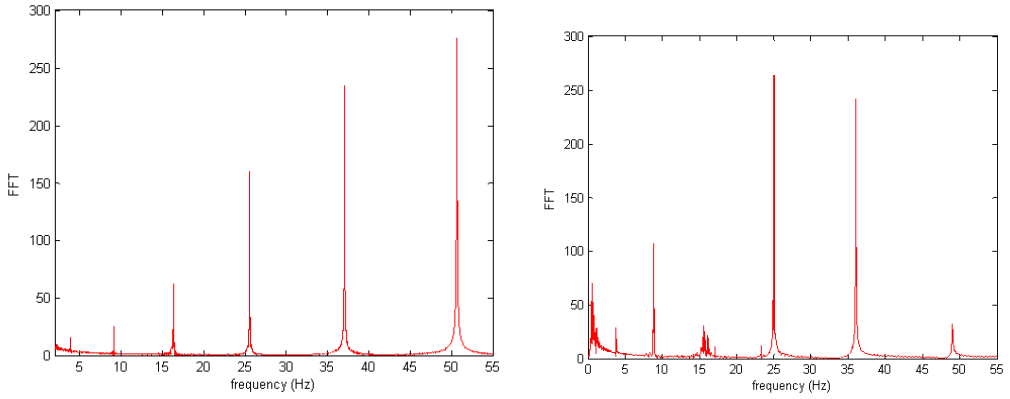
Appendix B: Experimental time histories

In this Appendix we present some selected experimental time histories obtained from the field tests used in Chapters 3 and 4. These time histories were the raw data that were used to extract the modal properties, which were eventually used as data in the Bayesian model updating. Illustrative forced vibration time histories used in Chapter 3 along with their Fourier transform are shown in Figures B.1 and B.2. Illustrative ambient acceleration time histories used in Chapter 4 are shown in Figure B.3.



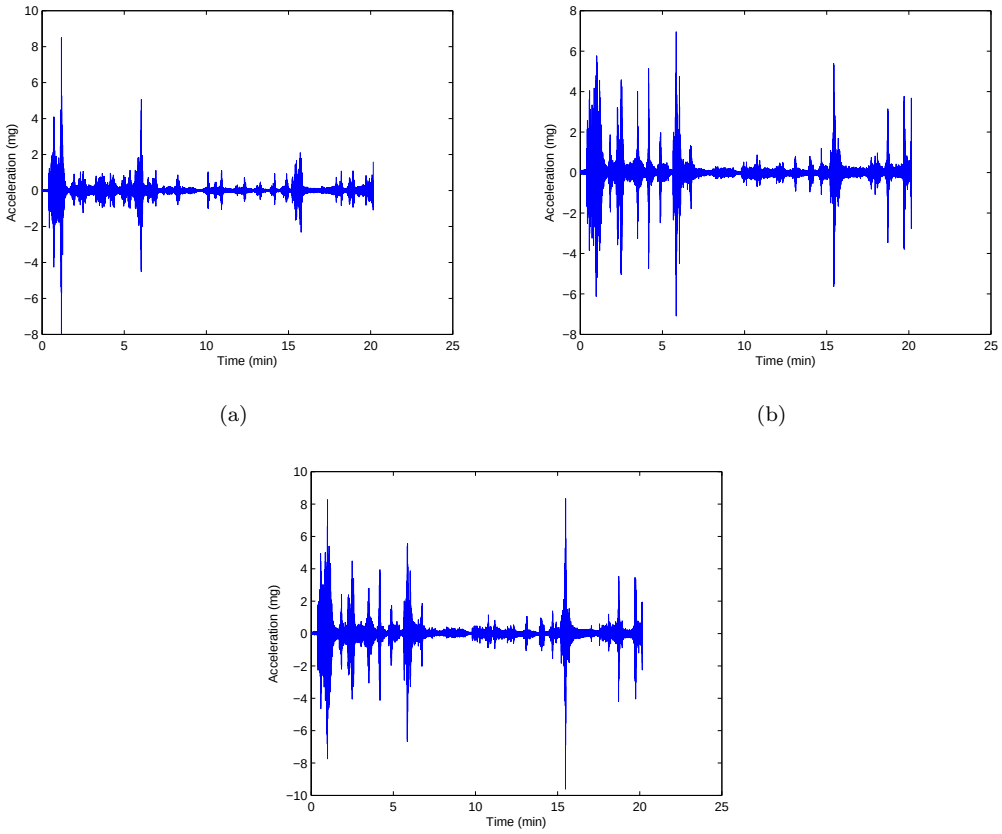
(a) (b)

Figure B.1: Hanger forced acceleration time histories.



(a) (b)

Figure B.2: Hanger forced acceleration Fast Fourier Transforms.



(a) (b)

(c)

Figure B.3: Selected ambient acceleration time histories from Metsovo bridge.

Bibliography

Alana, J.E. (2010). "Optimal measurement locations for parameter estimation of non linear distributed parameter systems". *Brazilian Journal of Chemical Engineering*, vol. 27(4): 627–642.

Angelikopoulos, P., Papadimitriou, C. and Koumoutsakos, P. (2012). "Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework". *The Journal of Chemical Physics*, vol. 137(14): 103–144.

Angelikopoulos, P., Papadimitriou, C. and Koumoutsakos, P. (2015). "X-TMCMC: Adaptive kriging for Bayesian inverse modeling". *Computer Methods in Applied Mechanics and Engineering*, vol. 289: 409–428.

ANSYS (2013). "ANSYS Mechanical APDL Theory Reference". *ANSYS Inc*, vol. Release15(November): 1 – 909.

Argyris, C., Papadimitriou, C. and Panetsos, P. (2016). "Bayesian optimal sensor placement for modal identification of civil infrastructures". *Journal of Smart Cities*, vol. 2(2): 69–86.

Au, S.K. (2011). "Assembling mode shapes by least squares". *Mechanical Systems and Signal Processing*, vol. 25(1): 163–179.

Au, S.K. (2012a). "Fast Bayesian ambient modal identification in the frequency domain, Part I: Posterior most probable value". *Mechanical Systems and Signal Processing*, vol. 26: 60–75.

Au, S.K. (2012b). "Fast Bayesian ambient modal identification in the frequency domain, Part II: Posterior uncertainty". *Mechanical Systems and Signal Processing*, vol. 26: 76–90.

Au, S.K. and Zhang, F.L. (2012). "Ambient modal identification of a primarysecondary structure by Fast Bayesian FFT method". *Mechanical Systems and Signal Processing*, vol. 28: 280–296.

Au, S.K., Zhang, F.L. and Ni, Y.C. (2013). "Bayesian operational modal analysis: Theory, computation, practice". *Computers and Structures*, vol. 126: 3–14.

Au, S.K., Zhang, F.L. and To, P. (2012). "Field observations on modal properties of two tall buildings under strong wind". *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 101: 12–23.

Barcilon, V. (1976). "Inverse problem for a vibrating beam". *Journal of Applied Mathematics and Physics*, vol. 27(3): 347–358.

Beck, J.L. (2010). "Bayesian system identification based on probability logic". *Structural Control and Health Monitoring*, vol. 17(7): 825–847.

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

Beck, J.L. and Katafygiotis, L.S. (1998). "Updating models and their uncertainties. I: Bayesian statistical framework". *Journal of Engineering Mechanics*, vol. 124(4): 463–467.

Beck, J.L. and Taflanidis, A. (2013). "Prior and posterior robust stochastic predictions for dynamical systems using probability logic". *International Journal for Uncertainty Quantification*, vol. 3 (4): 271–288.

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

Belleri, A. and Moaveni, B. (2015). "Identification of tensile forces in tie rods with unknown boundary conditions". In "7th International Conference on Intelligent Infrastructure", Turin, Italy.

Bellino, A., Garibaldi, L., Fasana, A. and Marchesiello, S. (2011). "Tension estimation of cables with different boundary conditions by means of the added mass technique". In "International Conference Surveillance 6", Compiegne, France.

Bellino, A., Marchesiello, S., Fasana, A. and Garibaldi, L. (2010). "Cable tension estimation by means of vibration response and moving mass technique". *Mecanique & Industries*, vol. 11(6): 505–512.

Bleistein, N. and Handelsman, R.A. (1975). *Asymptotic Expansions of Integrals*, vol. 50. Courier Corporation.

Bokaian, A. (1990). "Natural frequencies of beams under tensile axial loads". *Journal of Sound and Vibration*, vol. 142(3): 481–498.

Bungartz, H. and Griebel, M. (2004). "Sparse grids". *Acta Numerica*, vol. 13: 147–269.

Burczynski, T., Kus, W., Dlugosz, A. and Orantek, P. (2004). "Optimization and defect identification using distributed evolutionary algorithms". *Engineering Applications of Artificial Intelligence*, vol. 17(4): 337–344.

Cawley, P. and Adams, R.D. (1979). "The location of defects in structures from measurements of natural frequencies". *The Journal of Strain Analysis for Engineering Design*, vol. 14(2): 49–57.

Ceballos, M.A. and Prato, C.A. (2008). "Determination of the axial force on stay cables accounting for their bending stiffness and rotational end restraints by free vibration tests". *Journal of Sound and Vibration*, vol. 317(1-2): 127–141.

Chaloner, K. and Verdinelli, I. (1995). "Bayesian Experimental Design: A Review". *Statistical Science*, vol. 10(3): 273–304.

Chang, P.C. and Liu, S.C. (2003). "Recent Research in Nondestructive Evaluation of Civil Infrastructures". *Journal of Materials in Civil Engineering*, vol. 15(3): 298–304.

Chatzi, E.N., Hiriyur, B., Waisman, H. and Smyth, A.W. (2011). "Experimental application and enhancement of the XFEM-GA algorithm for the detection of flaws in structures". *Computers and Structures*, vol. 89(7-8): 556–570.

Cheung, S.H. and Beck, J.L. (2009). "Bayesian Model Updating Using Hybrid Monte Carlo Simulation with Application to Structural Dynamic Models with Many Uncertain Parameters". *Journal of Engineering Mechanics*, vol. 135(4): 243–255.

Ching, J. and Chen, Y.C. (2007). "Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging". *Journal of Engineering Mechanics*, vol. 133(7): 816–832.

Chow, H.M., Lam, H.F., Yin, T. and Au, S.K. (2011). "Optimal sensor configuration of a typical transmission tower for the purpose of structural model updating". *Structural Control and Health Monitoring*, vol. 18(3): 305–320.

Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory 2nd Edition*.

Ewins, D.J. (2000). *Modal testing : theory, practice, and application*. Research Studies Press.

Fang, Z. and Wang, J.q. (2012). "Practical Formula for Cable Tension Estimation by Vibration Method". *Journal of Bridge Engineering*, vol. 17(1): 161–164.

Flynn, E.B. and Todd, M.D. (2010). "A Bayesian approach to optimal sensor placement for structural health monitoring with application to active sensing". *Mechanical Systems and Signal Processing*, vol. 24: 891–903.

Friswell, M.I. and Mottershead, J.E. (1995). *Finite element model updating in structural dynamics*, vol. 38.

Gadala, M.S. and McCullough, A.D. (1999). "On the finite element analysis of inverse problems in fracture mechanics". *Engineering Computations*, vol. 16(4): 481–502.

Gaitanaros, S., Karaiskos, G., Papadimitriou, C. and Aravas, N. (2010). "A Bayesian methodology for crack identification in structures using strain measurements". *International Journal of Reliability and Safety*, vol. 4(2-3): 206–237.

Gerstner, T. and Griebel, M. (1998). "Numerical integration using sparse grids". *Numerical Algorithms*, vol. 18(3): 209–232.

Goller, B. and Schueller, G.I. (2011). "Investigation of model uncertainties in Bayesian structural model updating". *Journal of Sound and Vibration*, vol. 330(25): 6122–6136.

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*, vol. 284(1): 1–21.

Hansen, N., Muller, S.D. and Koumoutsakos, P. (2003). "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)". *Evolutionary Computation*, vol. 11(1): 1–18.

Hastings, W.K. (1970). "Monte carlo sampling methods using Markov chains and their applications". *Biometrika*, vol. 57(1): 97–109.

He, Y., Guo, D. and Chu, F. (2001). "Using genetic algorithms and finite element methods to detect shaft crack for rotor-bearing system". *Mathematics and Computers in Simulation*, vol. 57(1-2): 95–108.

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

Heredia-Zavoni, E., Montes-Iturrizaga, R. and Esteva, L. (1999). "Optimal instrumentation of structures on flexible base for system identification". *Earthquake Engineering & Structural Dynamics*, vol. 28(12): 1471–1482.

Heylen, W., Lammens, S. and Sas, P. (1998). *Modal analysis theory and testing*. Katholieke Universiteit Leuven, Faculty of Engineering, Dept. of Mechanical Engineering, Division of Production Engineering, Machine Design and Automation.

Horibe, T. and Takahashi, K. (2007). "Crack Identification in Beam Using Genetic Algorithm and Three Dimensional p-FEM". *Journal of Solid Mechanics and Materials Engineering*, vol. 1(7): 886–894.

Huan, X. and Marzouk, Y.M. (2012). "Gradient-based stochastic optimization methods in Bayesian experimental design". *arXiv.org*, vol. 4(6): 40.

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

Huang, Y.H., Fu, J.Y., Wang, R.H., Gan, Q. and Liu, A.R. (2015). "Unified Practical Formulas for Vibration-Based Method of Cable Tension Estimation". *Advances in Structural Engineering*, vol. 18(3): 405–422.

Humar, J.L. (2012). *Dynamics of structures*. CRC Press - Taylor & Francis Croup, 3 ed.

Jensen, H.A., Millas, E., Kusanovic, D. and Papadimitriou, C. (2014). "Model-Reduction Techniques for Bayesian Finite Element Model Updating Using Dynamic Response Data". *Computers and Structures*, vol. 279: 301–324.

Jung, J., Jeong, C. and Taciroglu, E. (2013). "Identification of a scatterer embedded in elastic heterogeneous media using dynamic XFEM". *Computer Methods in Applied Mechanics and Engineering*, vol. 259: 50–63.

Jung, J. and Taciroglu, E. (2014). "Modeling and identification of an arbitrarily shaped scatterer using dynamic XFEM with cubic splines". *Computer Methods in Applied Mechanics and Engineering*, vol. 278: 101–118.

Kammer, D.C. (1991). "Sensor Placement for On-Orbit Modal Identification and Correlation of Large Space Structures". *Journal of Guidance, Control, and Dynamics*, vol. 14(2)(August): 251–259.

Kammer, D.C. (2005). "Sensor set expansion for modal vibration testing". *Mechanical Systems and Signal Processing*, vol. 19(4): 700–713.

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

Katafygiotis, L.S., Papadimitriou, C. and Lam, H.F. (1998). "A probabilistic approach to structural model updating". *Soil Dynamics and Earthquake Engineering*, vol. 17(7-8): 495–507.

Kelantonis, V. (2010). *Design and dynamic analysis of the Metsovo bridge using Finite Elements*. Master thesis, University of Thessaly, Greece.

Kim, B.H. and Park, T. (2007). "Estimation of cable tension force using the frequency-based system identification method". *Journal of Sound and Vibration*, vol. 304(3-5): 660–676.

Kirkegaard, P. and Brincker, R. (1994). "On the optimal location of sensors for parametric identification of linear structural systems". *Mechanical Systems and Signal Processing*, vol. 8(6): 639–647.

Krawczuk, M. (2002). "Application of spectral beam finite element with a crack and iterative search technique for damage detection". *Finite Elements in Analysis and Design*, vol. 38(6): 537–548.

Lagomarsino, S. and Calderini, C. (2005). "The dynamical identification of the tensile force in ancient tie-rods". *Engineering Structures*, vol. 27(6): 846–856.

Lam, H.F., Katafygiotis, L.S. and Mickleborough, N.C. (2004). "Application of a Statistical Model Updating Approach on Phase I of the IASC-ASCE Structural Health Monitoring Benchmark Study". *Journal of Engineering Mechanics-Asce*, vol. 130(1): 34–48.

Lam, H.F., Yang, J. and Au, S.K. (2015). "Bayesian model updating of a coupled-slab system using field test data utilizing an enhanced Markov chain Monte Carlo simulation algorithm". *Engineering Structures*, vol. 102: 144–155.

Li, D.S., Li, H.N. and Fritzen, C.P. (2009). "A note on fast computation of effective independence through QR downdating for sensor placement". *Mechanical Systems and Signal Processing*, vol. 23(4): 1160–1168.

Lindley, D.V. (1956). "On a Measure of the Information Provided by an Experiment". *The Annals of Mathematical Statistics*, vol. 27(4): 986–1005.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). "Equation of state calculations by fast computing machines". *Journal Chemical Physics*, vol. 21(6): 1087–1092.

Moës, N., Dolbow, J. and Belytschko, T. (1999). "A finite element method for crack growth without remeshing". *International Journal for Numerical Methods in Engineering*, vol. 46(1): 131–150.

Moore, E.Z., Murphy, K.D. and Nichols, J.M. (2011). "Crack identification in a freely vibrating plate using Bayesian parameter estimation". *Mechanical Systems and Signal Processing*, vol. 25(6): 2125–2134.

Muto, M. and Beck, J.L. (2008). "Bayesian updating and model class selection for hysteretic structural models using stochastic simulation". *Journal of Vibration and Control*, vol. 14(1-2): 7–34.

Nam, H. and Nghia, N.T. (2011). "Estimation of cable tension using measured natural frequencies". In "Procedia Engineering", vol. 14. pp. 1510–1517.

Nanthakumar, S.S., Lahmer, T. and Rabczuk, T. (2013). "Detection of flaws in piezoelectric structures using extended FEM". *International Journal for Numerical Methods in Engineering*, vol. 96(6): 373–389.

Nanthakumar, S.S., Lahmer, T. and Rabczuk, T. (2014). "Detection of multiple flaws in piezoelectric structures using XFEM and level sets". *Computer Methods in Applied Mechanics and Engineering*, vol. 275: 98–112.

Ni, Y., Ko, J. and Zheng, G. (2002). "Dynamic analysis of large-diameter sagged cables taking into account flexural rigidity". *Journal of Sound and Vibration*, vol. 257(2): 301–319.

Ni, Y., Lu, X. and Lu, W. (2017). "Operational modal analysis of a high-rise multi-function building with dampers by a Bayesian approach". *Mechanical Systems and Signal Processing*, vol. 86: 286–307.

Ni, Y.C., Zhang, F.L., Lam, H.F. and Au, S.K. (2016). "Fast Bayesian approach for modal identification using free vibration data, Part IIPosterior uncertainty and application". *Mechanical Systems and Signal Processing*, vol. 70: 221–244.

Ntotsios, E., Papadimitriou, C., Panetsos, P., Karaiskos, G., Perros, K. and Perdikaris, P.C. (2009). "Bridge health monitoring system based on vibration measurements". *Bulletin of Earthquake Engineering*, vol. 7(2)(2): 469–483.

Ostachowicz, W.M. (2008). "Damage detection of structures using spectral finite element method". *Computers & Structures*, vol. 86(3-5): 454–462.

Papadimitriou, C. (2004). "Optimal sensor placement methodology for parametric identification of structural systems". *Journal of Sound and Vibration*, vol. 278(4-5): 923–947.

Papadimitriou, C. (2016). "Bayesian Uncertainty Quantification and Propagation (UQ+P): State-of-the-Art Tools for Linear and Nonlinear Structural Dynamics Models". In E. Chatzi and C. Papadimitriou, editors, "Identification Methods for Structural Health Monitoring", chap. 6. Springer International Publishing, pp. 137–170.

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

Papadimitriou, C., Beck, J.L. and Katafygiotis, L.S. (2001). "Updating robust reliability using structural test data". *Probabilistic Engineering Mechanics*, vol. 16(2): 103–113.

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

Papadimitriou, C., Ntotsios, E., Giagopoulos, D. and Natsiavas, S. (2011). "Variability of updated finite element models and their predictions consistent with vibration measurements". *Structural Control and Health Monitoring*, vol. 19(5): 630–654.

Papadimitriou, C. and Papadioti, D.C. (2013). "Component mode synthesis techniques for finite element model updating". *Computers and Structures*, vol. 126: 15–28.

Park, K.S., Seong, T.R. and Noh, M.H. (2015). "Feasibility Study on Tension Estimation Technique for Hanger Cables Using the FE Model-Based System Identification Method". *Mathematical Problems in Engineering*, vol. 2015: 1–12.

Qureshi, Z.H., Ng, T.S. and Goodwin, G.C. (1980). "Optimum experimental design for identification of distributed parameter systems". *International Journal of Control*, vol. 31(1): 21–29.

Rabinovich, D., Givoli, D. and Vigdergauz, S. (2007). "XFEM-based crack detection scheme using a genetic algorithm". *International Journal for Numerical Methods in Engineering*, vol. 71(9): 1051–1080.

Rabinovich, D., Givoli, D. and Vigdergauz, S. (2009). "Crack identification by 'arrival time' using XFEM and a genetic algorithm". *International Journal for Numerical Methods in Engineering*, vol. 77(3): 337–359.

Ren, W.X., Chen, G. and Hu, W.H. (2005). "Empirical formulas to estimate cable tension by cable fundamental frequency". *Structural Engineering and Mechanics*, vol. 20(3): 363–380.

Rus, G., Lee, S.Y. and Gallego, R. (2005). "Defect identification in laminated composite structures by BEM from incomplete static data". *International Journal of Solids and Structures*, vol. 42(5): 1743–1758.

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

Sebastiani, P. and Wynn, H.P. (2000). "Maximum entropy sampling and optimal Bayesian experimental design". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62(1): 145–157.

Sehgal, S. and Kumar, H. (2015). "Structural Dynamic Model Updating Techniques: A State of the Art Review". *Archives of Computational Methods in Engineering*.

Shah, P.C. and Udwadia, F.E. (1978). "A Methodology for Optimal Sensor Locations for Identification of Dynamic Systems". *Journal of Applied Mechanics*, vol. 45(1): 188–196.

Shinke, T., Hironaka, K., Zui, H. and Nishimura, H. (1996). "Practical Formulas for Estimation of Cable Tension By Vibration Method". *Journal of Structural Engineering*, vol. 122(6): 651–656.

Simoen, E., Moaveni, B., Conte, J.L.P. and Lombaert, G. (2013a). "Uncertainty Quantification in the Assessment of Progressive Damage in a 7-Story Full-Scale Building Slice". *ASCE Journal of Engineering Mechanics*, vol. 139(12): 1818–1830.

Simoen, E., Papadimitriou, C. and Lombaert, G. (2013b). "On prediction error correlation in Bayesian model updating". *Journal of Sound and Vibration*, vol. 332(18): 4136–4152.

Stavroulakis, G.E. and Antes, H. (1998). "Flaw identification in elastomechanics: BEM simulation with local and genetic optimization". *Structural Optimization*, vol. 16(2-3): 162–175.

Stephan, C. (2012). "Sensor placement for modal identification". *Mechanical Systems and Signal Processing*, vol. 27(1): 461–470.

Sukumar, N., Chopp, D.L., Moës, N. and Belytschko, T. (2001). "Modeling holes and inclusions by level sets in the extended finite-element method". *Computer Methods in Applied Mechanics and Engineering*, vol. 190(46-47): 6183–6200.

Sun, H., Waisman, H. and Betti, R. (2013). "Nondestructive identification of multiple flaws using XFEM and a topologically adapting artificial bee colony algorithm". *International Journal for Numerical Methods in Engineering*, vol. 95(10): 871–900.

Sun, H., Waisman, H. and Betti, R. (2014). "A multiscale flaw detection algorithm based on XFEM". *International Journal for Numerical Methods in Engineering*, vol. 100(7): 477–503.

Teughels, A. and Roeck, G. (2005). "Damage detection and parameter identification by finite element model updating". *Archives of Computational Methods in Engineering*, vol. 12(2): 123–164.

Tierney, L. and Kadane, J.B. (1986). "Accurate approximations for posterior moments and marginal densities". *Journal of the american statistical association*, vol. 81(393): 82–86.

Tsopelas, P., Okamoto, S., Constantinou, M.C., Ozaki, D. and Fujii, S. (1994). "Experimental and Analytical Study of Systems Consisting of Sliding Bearings, Rubber Restoring Force Devices and Fluid Dampers". Tech. rep., Buffalo.

Ucinski, D. (2000). "Optimal sensor location for parameter estimation of distributed processes". *International Journal of Control*, vol. 73(13): 1235–1248.

Udwadia, F.E. (1994). "Methodology for Optimum Sensor Locations for Parameter Identification in Dynamic Systems". *Journal of Engineering Mechanics*, vol. 120(2): 368–390.

Vanik, M.W. and Beck, J.L. (1998). "A Bayesian probabilistic approach to structural health monitoring". *Structural Health Monitoring*, vol. 3243: 140–151.

Vanik, M.W., Beck, J.L. and Au, S.K. (2000). "Bayesian probabilistic approach to structural health monitoring". *ASCE Journal of Engineering Mechanics*, vol. 126(7)(7): 738–745.

Waisman, H., Chatzi, E. and Smyth, A. (2010). "Detection and quantification of flaws in structures by the extended finite element method and genetic algorithms". *International Journal for Numerical Methods in Engineering*, vol. 82(3): 303–328.

William, T.T. (1996). *Theory of vibration with applications*. CRC Press, 4 ed.

Xu, Y. and Liu, G. (2002). "Detection of flaws in composites from scattered elastic-wave field using an improved μGA and a local optimizer". *Computer Methods in Applied Mechanics and Engineering*, vol. 191(36): 3929–3946.

Yan, W.J. and Katafygiotis, L.S. (2015). "A novel Bayesian approach for structural model updating utilizing statistical modal information from multiple setups". *Structural Safety*, vol. 52(PB): 260–271.

Yan, G., Sun, H. and Waisman, H. (2015). "A guided Bayesian inference approach for detection of multiple flaws in structures using the extended finite element method". *Computers & Structures*, vol. 152: 27–44.

Ye, S.Q. and Ni, Y.Q. (2012). "Information entropy based algorithm of sensor placement optimization for structural damage detection". *Smart Structures and Systems*, vol. 10(4): 443–458.

Yin, T., Lam, H.F. and Chow, H.M. (2010). "A Bayesian Probabilistic Approach for Crack Characterization in Plate Structures". *Computer-Aided Civil and Infrastructure Engineering*, vol. 25(5): 375–386.

Yuen, K.V. (2010). *Bayesian Methods for Structural Dynamics and Civil Engineering*. John Wiley & Sons, Singapore.

Yuen, K.V. (2012). "Updating large models for mechanical systems using incomplete modal measurement". *Mechanical Systems and Signal Processing*, vol. 28: 297–308.

Yuen, K.V., Beck, J.L. and Katafygiotis, L.S. (2002). "Probabilistic approach for modal identification using non-stationary noisy response measurements only". *Earthquake Engineering and Structural Dynamics*, vol. 31(4): 1007–1023.

Yuen, K.V., Beck, J.L. and Katafygiotis, L.S. (2006a). "Efficient model updating and health monitoring methodology using incomplete modal data without mode matching". *Structural Control and Health Monitoring*, vol. 13(1): 91–107.

Yuen, K.V., Beck, J.L. and Katafygiotis, L.S. (2006b). "Unified probabilistic approach for model updating and damage detection". *Journal of Applied Mechanics-Transactions of the ASME*, vol. 73(4): 555–564.

Yuen, K.V., Katafygiotis, L.S., Papadimitriou, C. and Mickleborough, N.C. (2001). "Optimal Sensor Placement Methodology for Identification with Unmeasured Excitation". *Journal of Dynamic Systems, Measurement, and Control*, vol. 123(December): 677.

Yuen, K.V. and Kuok, S.C. (2015). "Efficient Bayesian sensor placement algorithm for structural identification: A general approach for multi-type sensory systems". *Earthquake Engineering and Structural Dynamics*, vol. 44(5): 757–774.

Yuen, K.V. and Mu, H.Q. (2010). "Peak Ground Acceleration Estimation by Linear and Nonlinear Models with Reduced Order Monte Carlo Simulation". *Computer-Aided Civil and Infrastructure Engineering*, vol. 26(1): 30–47.

Yuen, K.V. and Mu, H.Q. (2015). "Real-Time System Identification: An Algorithm for Simultaneous Model Class Selection and Parametric Identification". *Computer-Aided Civil and Infrastructure Engineering*, vol. 30(10): 785–801.

Zhang, F.L. and Au, S.K. (2016). "Fundamental two-stage formulation for Bayesian system identification, Part II: Application to ambient vibration data". *Mechanical Systems and Signal Processing*, vol. 66: 43–61.

Zhang, F.L., Ni, Y.C., Au, S.K. and Lam, H.F. (2016). "Fast Bayesian approach for modal identification using free vibration data, Part I Most probable value". *Mechanical Systems and Signal Processing*, vol. 70: 209–220.