Probabilistic Uncertainty Quantification for Structural Dynamics Under Limited Data

Felipe Igea

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2023

Table of Contents

Acknowledgements

I want to convey my sincerest gratitude to my DPhil supervisor, Professor Alice Cicirello, for her unconditional support, direction, and encouragement throughout my DPhil journey. The outcome of this thesis has been greatly influenced by our regular discussions and her insightful comments. I would also like to thank Professor Manolis Chatzis, for his help and guidance during the first years of this DPhil which made me understand and appreciate the topic of Structural Health Monitoring. I am also grateful to Professor Youssef Marzouk for providing me with the valuable experience of visiting his research group at MIT, as well as the opportunity to collaborate and exchange ideas with him.

I would like to acknowledge the support and funding provided by the EPSRC and Schlumberger for an Industrial Case postgraduate scholarship (Grant ref. EP/T517653/1) which has made this research possible.

For their unwavering support and encouragement during my DPhil journey, I would like to sincerely thank Andreea, Filippo, Luca, and all the other present and previous members of the Data, Vibration and Uncertainty laboratory. I want to express my gratitude for the group's supportive and intellectually stimulating environment, which has been crucial to my development, both personally and professionally. I also wish to express my gratitude to the many friends I made during my brief time at MIT, Alejandra, Cristina, David, Luis, Pilar, and Teresa, who were always there and helped to make this experience much more enjoyable. I would also like to thank Danny, Emily, Joanna, Kelvin and Loek for making me feel welcome during my time in the ACDL Laboratory. A special thanks to all of my Oxford friends, specially to Andreea and Subha, whose support was crucial, particularly during the pandemic's most difficult days. Last but not least, I would like to dedicate this DPhil to my family who have always believed in me and supported me every single day through this fulfilling but difficult journey.

Abstract

Two grand challenges under the context of limited data availability in structural dynamics are faced in this thesis: the design of complex engineering structures, and the updating of data on already existing critical structures by integration of physics-based models. The motivation to confront these challenges is driven by the need to develop techniques able to prevent unforeseen failures of structures that could result in significant losses.

To achieve this end, in this thesis, these two grand challenges were split into four subchallenges: a) quantifying uncertainties at the design stage. The presence of uncertainties should be considered during the design phase, so that the structure can be operated safely, even if it deviates from its nominal design parameters. In this manner, designs robust to variability can be built; b) where should measurements be taken? After the design stage, when the structure is built, velocity and acceleration measurements may be obtained from sensors located in the structure. However, these sensors should be placed at locations on the structure where the largest amount of information can be gained. This is not a trivial task, as different locations may not provide the same information as others, and therefore, some locations may not be optimal; c) updating the model: using physics-based models that represent the structure of interest, and the measurements collected in the structure by sensors, the knowledge on the condition of the structure can be updated; d) handling prior uncertainty under limited data availability, the posterior prediction may be significantly sensitive to prior uncertainty. Understanding how sensitive the posterior prediction might be under limited information may allow the practitioner to be more or less confident about the condition of the structure.

The main contributions of this thesis are: (i) obtaining distributions of the modal parameters at the design stage, when only data from a numerical model or prototype is available. This is achieved by developing a non-parametric method based on the combination of Random Matrix Theory and the Eigensystem Realization algorithm; (ii) the investigation of utility functions for optimal sensor placement, that highlights the challenges found when using a Bayesian Optimal Design approach;

(iii) the investigation of Variational Inference for fast inference of the latent parameters of a physics-based model; (iv) a new technique for the reduction of computational cost, a better capture of the possible multi-modality of the latent parameters, and efficient sampling of the parameter space, by using a cyclical schedule with the combination of Bayesian Quadrature and Variational Inference; (v) developing a new approach that can be used when the prior is uncertain, based on interacting Wasserstein Gradient Flows, able to compute the worst-case and optimal priors in the vicinity of a selected prior, and their associated posteriors.

Chapter 1 - Introduction

1.1 Research motivation

In the field of structural dynamics, the improvement and development of methods that allow a better understanding of the performance and condition of a system is currently a pressing issue. This fact becomes especially significant for the cases that involve critical components or structures where their unexpected failures may cause important immediate losses e.g., lives, repair costs; but also, consequential losses like the ones derived from the inability to continue to operate the system.

During the design stage, before structures are built, and data can be obtained from them, the presence of uncertainties should be taken into account, so the structure can be safely operated, even if it deviates from its nominal design specifications [1]. The design of the structure may include numerical and/or prototype phase/s, to ensure that the structure works safely, and it may be efficiently maintained.

The numerical phase/s may include the development of a physics-based model that is used to represent the engineering system, so its performance under different conditions may be understood [2]. If the parameters of the physics-based model that represent the structure of interest are chosen to be deterministic, (unrealistic) designs overfitted to those parameters may be obtained [2]. Therefore, uncertainty in the parameters should be considered, so designs that show sufficient flexibility to capture the variability in the system, and are robust to varying conditions of e.g., static and dynamic loads, changes of geometry and mechanical characteristics of the materials, are produced [3].

These physics-based models may be used to provide outputs such as the static and dynamic response of the structure and its components, that may be of great value to the practitioner. However, caution should be exerted before the outputs of these physics-based models are used, as some of its parameters, such as mechanical and dimensional properties of the elements, loads and boundary conditions, may exhibit variability or uncertainty. Furthermore, during the development of the physics-based model, implicit and explicit simplifying assumptions of the structural model are made,

mostly because of limited knowledge, and computational budget. These simplifying assumptions compromise the accuracy and precision of the model and its intended objectives [2].

Data collected from the structure, such as velocity and acceleration readings, and modal information derived from those readings, are usually employed in the field of structural dynamics to update the physics-based models. The model updating framework, using observations collected in the structure, is able to produce estimations of the parameters to be employed in the physics-based model [2]. Model updating is becoming an increasingly relevant topic in Structural Health Monitoring (SHM), as it may provide support for decision-making in processes and activities such as maintenance, detection of anomalies, and diagnosis of failure of engineering structures [4,5].

1.2 Research challenges

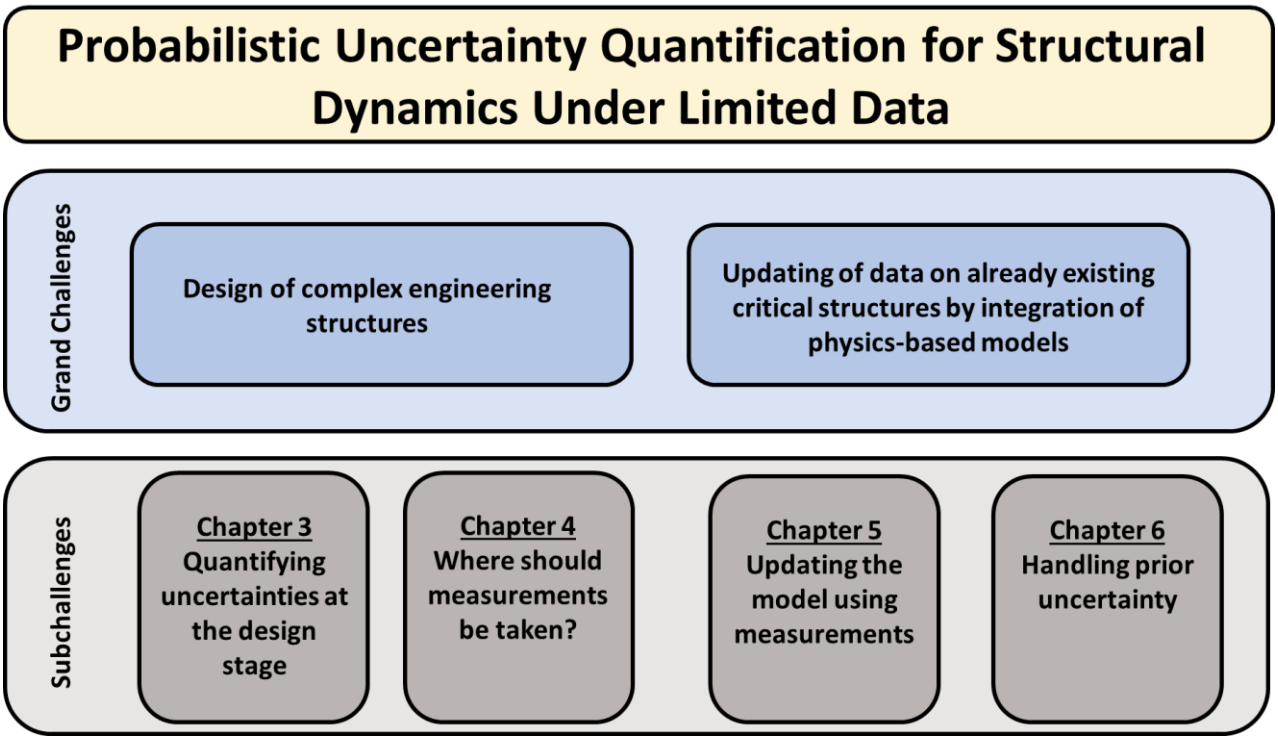


Fig. 1. Grand challenges and subchallenges.

In this DPhil thesis, two main challenges currently faced in the area of structural dynamics have been addressed. The first challenge is the design of complex engineering structures under limited data availability, and the second challenge consists of updating information relevant to the structure's

condition by using physics-based models and limited data. These research challenges are split into four subchallenges, as shown on fig. 1, and are discussed in the following subchapters.

1.2.1 Quantifying uncertainties at the design stage

The behaviour of engineering systems is frequently simulated using either physical or numerical models. The validity of the numerical models is based on their accuracy, precision and quality. Nevertheless, a major challenge during the development of a numerical model is that several properties (e.g., geometric or material properties, loading and boundary conditions) of the system, are either unknown, or show some uncertainty. It also has to be considered that due to limited computing power, lack of some information, and other reasons, simplifying assumptions about the model are explicitly or implicitly made.

One of the main challenges found during the design process of structures, is the quantification of the uncertainties on the modal parameters to be used for the assessment of the structures' dynamic performance [6,7]. This is due to uncertainties originated by the variability of the manufacturing processes of structural components, boundary conditions and assemblage [8,9]. An adequate identification of a prior distribution of the modal parameters across the ensemble of nominally identical structures would enable an effective selection of designs that are robust to these uncertainties, avoiding extensive modifications of the manufactured product [3].

At the design stages (before a prototype is being built), when a physics-based model is developed, uncertainties on the model parameters are usually described using probabilistic [2,10,11] or non-probabilistic uncertainty descriptions [12]. These uncertainties are then propagated through the equations of motion to yield the corresponding description of the response. This is the so-called parametric model of uncertainty. However, the choice of the uncertain parameters, and their description, would directly affect the distribution of the resulting modal parameters [6,7,11]. Even for cases where the probability density functions (pdfs) of the uncertain parameters of the structure modelled by expensive-to-evaluate physics-based models are known, the numerical assessment of

the performance of the structures becomes impractical, as the number of evaluations of the physics-based models required is very high. Therefore, a suitable approach that can avoid specifying the uncertainties' sources and the description of the model's parameters uncertainty would be of interest, as they are both often hard to determine.

1.2.2 Where should measurements be taken?

Sensorial data such as acceleration and velocity measurements are usually employed to infer structural characteristics, particularly, the modal parameters: mode shapes, damping ratios, and natural frequencies. The parameters obtained are subsequently employed to assess the state of the structures. It should also be noted that the inference of the structural parameters may be significantly impacted by the placement of the sensors.

From a general perspective, the approaches to sensor placement frameworks can be categorised in two groups: those falling under non-information techniques, and those rooted in information theory [13]. This DPhil primarily concentrates on information-based techniques, which to a great extent depend on the use and enhancement of the general Bayesian methodologies. Readers interested in non-information techniques are encouraged to refer to [13]. The general Bayesian framework, initially introduced in [14,15] for system identification purposes, has subsequently been expanded to address sensor placement problems [14]. The primary purpose of the sensor placement framework is to choose the optimal amount and position of sensors in a manner that the information required for estimating the uncertain parameters is maximised [16]. The selection of the most suitable optimisation approach, and the characterisation of the metric to be employed for the assessment of the suggested sensor layouts (positions and number of sensors), are the key issues tackled by the framework [17,18].

In this thesis, methods are developed for cases where the layout and working conditions of the structure introduce a limited availability of the number of observations that may be taken from them, their locations, and the time periods between those observations. Therefore, the standard approach of positioning the sensors where traditional optimal sensor placement or expert knowledge recommends

may not be used to tackle the problem, as the layout of the structure may restrict the access to some of its components. An example of this engineering problem, may be found in chapter 2 of this thesis.

1.2.3 Updating the model using measurements

In engineering, Bayesian model updating methods are commonly employed to compute the values of unknown latent parameters of some physics-based models, or to evaluate the implicit variability of uncertain latent parameters [19]. The Uncertainty Quantification (UQ) techniques allow the consideration of the uncertainties of the parameters of the updated model, and the assessment of the uncertainties of the system's response.

Bayesian model updating methods require the numerical calculation of the posterior distribution. As a result, for the cases where the computing power budget is limited, the extremely high number of forward model runs required may represent a problem. Therefore, the implementation of new approaches using advanced sampling algorithms is required [20]. These approaches for the problem of inference of uncertain parameters of expensive-to-run physics-based models are of significant interest, as frequently, due to budget or deadline constraints, only a limited number of model evaluations may be performed.

Another reason why efficient sampling algorithms are required is to obtain uncertain parameters in physics-based models which exhibit multi-modality. In this case, significant computational resources are needed to sample the parameter space. There are several reasons that may create the multi-modality of the uncertain parameters in structural dynamics. One frequent reason for multi-modality to occur may be environmental changes (e.g., temperature changes may imply stiffness changes [21]).

1.2.4 Handling prior uncertainty

In Bayesian Inference, for cases under limited data availability, it may be observed that the assumption of prior may influence significantly the posterior obtained. In some cases, e.g., the opinions of several consulted experts differ, therefore, it may not be feasible to characterise priors for the latent parameters. In those instances, and for a given prior, it is important to study how the choice

of priors in its neighbourhood may affect the predictions of their associated posteriors, as this might affect later computations. Therefore, being able to handle prior uncertainty in structural dynamics is of value, as this sensitivity to the choice of prior may impact decisions to be taken about the condition of the system, its need of maintenance, and remaining lifetime.

1.3 Research questions and strategy

The main research questions involved in the development of this general framework are as follows:

- How to efficiently characterise the uncertainties found in the design process for the assessment of the dynamic performance of a structure. An efficient method that can bypass the need to propagate the assumed uncertainties in expensive to evaluate models, such as large finite element (FE) models would be of interest. Similarly, a very interesting problem is the study of a methodology where from measurements performed on a prototype, a (reduced order) model is created. This reduced order model may be used to investigate uncertainties without defining the source of those uncertainties.
- How to define what sensor locations in the structure may bring the greatest amount of information. The choice of a specific metric, and the definition of parameters of interest, are two important factors to determine the sensor locations.
- How to reduce the computational cost involved in the process of updating the model using measurements obtained from experiments. The development of a fast inference methodology that efficiently samples the parameter space, especially when multi-modality is present, would reduce the overall computational cost when complex to evaluate models are used.
- How to produce the worst-case and optimal prior with respect to (w.r.t). a metric, and their associated posteriors. Those posteriors can be used as upper and lower bounds for robustness assessment computations.

1.4 Contributions of the thesis

The research work undertaken during this DPhil can be summarised as follows:

• the development of a novel approach that combines Random Matrix Theory and the Eigensystem Realization algorithm, to evaluate the modal parameters distributions using measurements obtained from a prototype, or numerical data from a model, described in paper I.

• the numerical investigation of utility functions for optimal sensor placement to highlight the current challenges in the Bayesian Optimal Design approach, shown in paper II.

• the investigation of Variational Inference (VI) techniques for fast inference (paper III) of the latent parameters in a physics-based model, and the development of a new technique (paper IV), for the reduction of computational cost, and efficient sampling of the parameter space, when multi-modality in the latent parameters is present, by using a cyclical schedule with the combination of Bayesian Quadrature and VI.

• the development of a new method (paper V), based on a Robust Bayesian Inference framework that employs an interacting Wasserstein Gradient Flow formulation. This new approach evaluates the optimal and worst-case prior in the vicinity of a chosen nominal prior, and their corresponding posteriors, which may be used as lower and upper bounds on subsequent calculations that require the estimation of the posterior.

1.5 Outline of thesis

The introduction of this thesis (chapter 1), initially presents its research motivation, discussing both the research challenges, and the research questions to be answered in this work. The key contributions of the thesis, and its outline, are then given.

Next, in chapter 2, an overview of the building blocks required for tackling the four subchallenges described in subchapter 1.2 is given. A review of the most relevant topics in UQ for structural dynamics is given first. Then the physics-based model of the coupled-beam system that is used as a

case study in a couple of papers in this thesis is introduced. The coupled-beam structure is first described, and both, the numerical FEA model used to simulate the dynamical behaviour of the structure, and the damping models used to simulate the damping phenomenon are shown. The Bayesian Inference framework is also presented. VI is then introduced as an alternative method to the sampling techniques that formulates the inference problem into an optimisation approach. Lastly, the concept of Gradient Flows and Wasserstein Gradient Flows for the optimisation of functionals in the Wasserstein-2 space is presented.

Chapter 3 introduces the problem of limited information at the design stage. This is an important topic, as it may allow to avoid costly modifications of the structure after it is built by ensuring the robustness of the structure's design to uncertainty. The attached paper I 'On the Combination of Random Matrix Theory with Measurements on a Single Structure' proposes a new method based on non-parametric uncertainty, that for an ensemble of nominally identical structural systems, allows the calculation of the pdfs of the modal parameters, taking as inputs the value of a so-called dispersion parameter, and observations taken from one member of the ensemble. The obtained pdfs of the modal parameters include the effect of the modelling errors, variability of assembly, boundary conditions, and manufacturing processes of the components. The importance of the knowledge of those distributions lies in the fact that they can be used before the manufacturing of the product, and therefore, a robust design can be chosen avoiding expensive alterations at later stages. The most important feature of the technique is its lack of requirement of the specification of the sources of uncertainty. If this approach is used, the numerical propagation of the uncertain parameters through the equations of motion of the structure is not needed. This becomes of greater advantage for cases where the model of the system is either unknown (that may result in significant modelling errors), or implies high complexity (that involves important numerical requirements).

Chapter 4 tackles the challenge of limited information in operating conditions by focusing on where measurements should be taken. In the paper II, 'On the Investigation of Utility Functions on Optimal Sensor Locations', attached at the end of the chapter 4, a simple case study is investigated. The subject

of investigation is the optimal location of sensors for the identification of two physical parameters of the structure. This is done using three different utility functions that each require the evaluation of the posterior pdfs of the latent parameters. It was found that for all three different utility functions, the optimal position of the sensors was identical. Most of the numerical cost of those calculations is due to the Bayesian Inference approach employed.

In chapter 5, the challenge of updating our physics-based model is introduced using a Bayesian Inference framework. VI is then presented as an alternative method to the sampling techniques that formulates the inference problem into an optimisation approach. The first paper relevant to chapter 5 (paper III) 'Structural Model Updating Using Variational Inference', attached in this chapter, shows that compared to Markov Chain Monte Carlo (MCMC), the number of model evaluations can be significantly reduced if a technique based on VI is employed. The paper III describes a technique that can be used for cases where the uncertain parameters may be approximated using a multivariate Gaussian distribution. The paper III proposes multivariate Gaussian pdfs with full covariance matrices, as posteriors of the latent parameters. However, it should be noted that the lower numerical cost of these optimization techniques is usually accompanied by a reduction of accuracy.

In the second paper relevant to chapter 5 (paper IV), 'Cyclical Variational Bayes Monte Carlo for Efficient Multi-Modal Posterior Distributions Evaluation', a second approach, also based on VI is developed. This technique may be used for the evaluation of multi-modal posterior pdfs of latent parameters, for cases where given the available data, the physics-based model is expensive to evaluate. The approach can achieve a better approximation of the posterior, as it uses a multivariate Gaussian mixture as its postulated posterior. The method is built on a combination of VI with active-sampling Bayesian Quadrature. The unnormalized posterior is cyclically annealed using an artificial temperature parameter that enhances the exploration potential and mode coverage of the algorithm. Under a limited number of model runs, the Cyclical Variational Bayesian Monte Carlo (Cyclical VBMC) technique is able to deal with the restrictions imposed by poor initialisations.

In chapter 6, the problem of prior uncertainty is introduced. This is of particular relevance, as when limited information is available, the assumed prior may significantly affect the posterior prediction of the uncertain parameters. The paper V attached in chapter 6, discusses a novel method for Robust Bayesian Inference. With the purpose of obtaining a robust estimation of the posterior pdfs of the latent parameters, and given some experimental data, an algorithm based on interacting Wasserstein Gradient Flows has been developed. The method produces estimations of the posteriors of the parameters using the worst-case and optimal prior distributions with respect to a functional of interest. To calculate the worst-case and optimal prior distribution, an ambiguity set is defined by a nominal prior, and the Wasserstein-2 distance. The paper V includes the derivation from first principles of the interacting Wasserstein Gradient Flow formulation, and the equations for the particle discretisation used to evaluate the optimal and worst-case prior. The method reduces the numerical cost, as the number of prior distributions to be evaluated is smaller than if all the possible prior distributions that lie inside the ambiguity set have to be computed.

The proposed technique gives valuable information about the effects of the selected prior on the predictions of the posteriors, and therefore, the effect on the structure's maintenance needs, particularly in the context of limited data availability. The approach first obtains the worst-case and optimal prior, and subsequently produces the associated posteriors. Those may be used to infer upper and lower bounds for the following computations that require estimations to the posterior.

In the seventh and last chapter of the thesis, a thorough summary of the research's conclusions, key takeaways, and recommendations for future research are given.

Chapter 2 - Overview of building blocks

In order to undertake the research challenges described in subchapter 1.2, the following building blocks (fig. 2) need to be investigated: UQ, physics-based models, Bayesian Inference and Gradient Flows. These building blocks are described in the following subchapters.

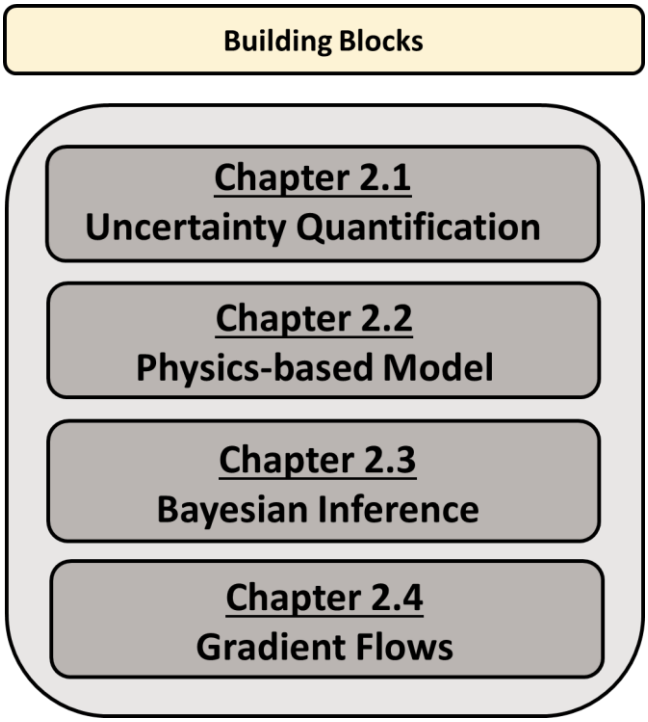


Fig. 2. Building blocks.

2.1 Uncertainty Quantification

2.1.1 Types of uncertainty

Depending on the particular research field, ‘uncertainty’ has different meanings [22]. For example, uncertainty in the data and the model, and convergence of the model are factors that could contribute to a system's response being uncertain [23].

The uncertainties of the inputs in a system are called data uncertainty [23]. Operational loads, geometrical and mechanical properties in a structure, are some examples of these unknown inputs.

For a system, an algorithm that transforms input data into an associated output can be defined as a model. Inaccuracies or oversimplifications in the model, or errors in its tuning parameters can introduce uncertainty in its output [2]. If an oversimplified algorithm is used to represent the physical problem at hand, model uncertainty is introduced [2]. This may happen for example, when modelling the behaviour of a beam, the Euler-Bernoulli assumptions are used, instead of the more appropriate Timoshenko assumptions, or when representing non-linear phenomena, linear models are used. Model uncertainty is unavoidable, due to the impossibility of building models able to represent the actual physical behaviours of the structures with absolute accuracy.

Two types of uncertainties may be found when dealing with engineering applications: aleatory variability and epistemic uncertainty [24]. Aleatory variability is related to the system's natural random behaviour, while epistemic uncertainty is due to a lack of knowledge of the practitioner. The main difference between aleatory and epistemic uncertainty is that the aleatory is generally irreducible, and it cannot be decreased with additional knowledge. However, epistemic uncertainty can be reduced, for example, by taking additional measurements with a sensor on a structure, this additional data may reduce the uncertainty introduced by the noisy measurements.

In the next subchapters, uncertainty models for random engineering structures are categorised into probabilistic or non-probabilistic, and parametric or non-parametric [23].

2.1.2 Probabilistic and non-probabilistic uncertainty

Two different approaches may be used to perform UQ. These two methodologies have their own strengths and limitations that can be outlined as follows:

(i) According to the probabilistic approach, the uncertainty of a given parameter is studied presuming that the parameter follows a specific pdf. The numerical model propagates this pdf to determine the probability of obtaining a specific model output, or in some limited situations, this can be performed analytically [25].

Two different approaches are taken in statistical inference for the interpretation of probability: the Frequentist and the Bayesian, these differ in how the uncertainties are managed [26–28]. Frequentists use frequencies to describe probabilities, while Bayesians use probabilities to represent subjective ignorance.

Analysis of published works, especially in an industrial context, shows that trying to overcome this difficulty, authors have frequently assumed the parameters to follow Gaussian distributions. Nonetheless, the previous assumption may produce misleading results and poor estimations of the model's behaviour [25].

(ii) The non-probabilistic approaches define the uncertainties by different methods, such as: an interval defined by the range of the uncertain parameter [29], a convex region where parameters lie [30], a fuzzy number usually represented by a membership function $\mu(x)$, that associates a degree of membership to each possible value of the uncertain parameter [31]. One of the main challenges that the non-probabilistic approaches share with the probabilistic approaches, is the propagation of the uncertainties through the model.

2.1.3 Uncertainty propagation through Monte Carlo Simulation

A method of uncertainty propagation of general use is the Monte Carlo Simulation (MCS) [23, 24, 25]. This approach involves the generation of a number of random samples sufficient to reflect the level of uncertainty associated with a particular input quantity, the deterministic output for each of the generated random samples, and the assessment of the statistical characteristics of the output.

It has to be considered that for the cases when the problem requires higher moments of the output, or the use of an expensive-to-evaluate model, such as a FE model with a large number of degrees of freedom, then the numerical cost of the MCS may result unaffordable.

With the purpose of overcoming that prohibitive computational cost, numerous sampling methods have been created. These approaches have the aim of decreasing the number of evaluations required

to achieve a certain level of accuracy in the output statistics, or to enhance the sampling space coverage. Some methods designed to decrease the number of evaluations are: the Subset Simulations approach [32], used in cases that require the computation of small failure probabilities; Importance Sampling [33] can be used in cases where the sampling concentrates on the failure zone, allowing for faster convergence to the true failure probability.

Another method developed for the enhancement of the sampling space coverage is the Latin Hypercube Sampling [34], this sampling method is frequently used to run computer experiments or for Monte Carlo integration.

2.1.4 Parametric and non-parametric uncertainty

Uncertainty Quantification techniques under a parametric approach involve a three-step process: first, the uncertain parameters have to be identified (for example some mechanical characteristics and geometry of a structure); second, either a probabilistic or a non-probabilistic description should be chosen; and third, the uncertain parameters are propagated through the model to estimate the output of the system. If the parametric probabilistic approach is followed, assumptions of the pdfs of the parameters to be propagated are required. As a summary, the objective of a parametric study is the calculation of the uncertain output through the propagation of the uncertain inputs of the model.

Conversely, if a non-parametric representation of uncertainty is chosen, it is assumed that all the uncertainties of the system may be encompassed by some ‘global uncertainty’ representation, irrespectively of the specific characteristics of each of the uncertainties [23]. A well-known case in Random Matrix Theory (RMT) [35], is the Gaussian Orthogonal Ensemble (GOE), for which, in general terms, the specific statistics of the elements of the random matrix become less relevant, and the local data of both, eigenvalues and eigenvectors tend to a global pattern imposed by the GOE.

The Maximum Entropy Principle [36] and RMT [11,37] has been used in FE models where the mass, stiffness, and damping of structures are characterised using real positive-definite symmetric random matrices. In these matrices, the mean value of the elements are the values that would be used to

construct the system's matrices. Making use of the statistical properties of random matrices and MCS, a methodology to efficiently produce an ensemble of random models can be designed.

In RMT, the so-called dispersion parameter may be used to define the variability of a matrix. The maximum likelihood approach [9] may be used to assess the dispersion parameter from the response variability of experimental data.

2.1.5 Forward and inverse Uncertainty Quantification

As shown on fig. 4, when facing the problem of uncertainty quantification, two major kinds of problems can be found: the input to output, forward problem, studies the propagation through the model (physical or mathematical) of the certain and uncertain parameters of the inputs [38].

On fig. 3, the inverse problem, from output to input, focuses on the study of the discrepancies between the outputs of the experimental and the mathematical or physical model used to describe the system studied. The inverse problem is of great interest, as designers may use the methodology for the improvement and simplification of the model that characterises an actual engineering system [38].

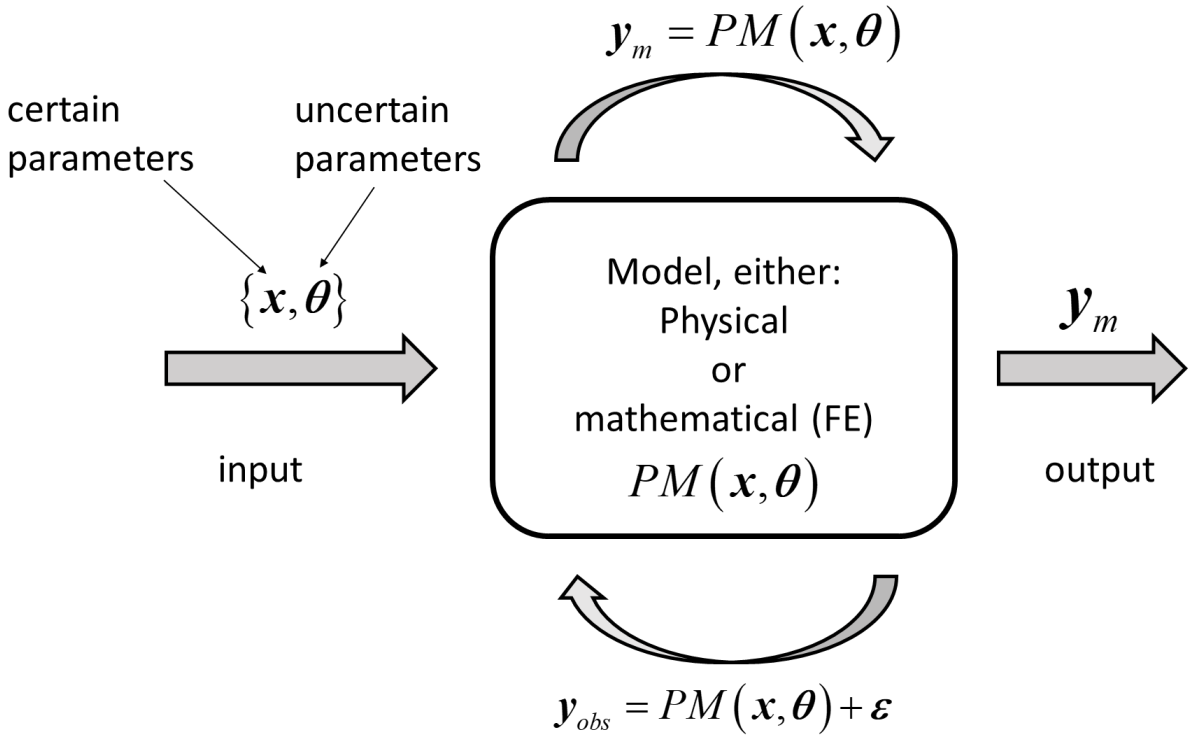


Fig. 3. Forward and inverse UQ.

The forward uncertainty propagation quantifies the uncertainties of the outputs $\boldsymbol{y}_m$ of the model (e.g., in several papers attached in this thesis $\boldsymbol{y}_m$ refers to the modal parameters) for a given set $\{\boldsymbol{x},\boldsymbol{\theta}\}$ of certain $\boldsymbol{x}$ and uncertain $\boldsymbol{\theta}$ parameters. The main focus of this problem is the study of how the outputs of the model are affected by the uncertainty of the input parameters introduced into the model [38]. The implementation of this technique requires the generation of a great quantity of input samples that are computationally run through the system to obtain the uncertainties of the response. As a result, a significant computational cost is incurred.

Inverse UQ is a much more complex problem than forward UQ, and it is usually performed during the process of model updating [38]. Some difficulties arise from the fact that experimental data is required to update the model, this data is normally assumed to be affected by Gaussian noise $\boldsymbol{\varepsilon}$ [38]. Experiments have to be run, and for these to produce useful data, the reading sensors have to be properly located (for example using sensor placement techniques). Also, in inverse uncertainty problems, multiple runs of the model have to be performed, and therefore, significant computational resources are required.

The process of quantification and characterisation of the uncertainties associated with data and computational models is called UQ. The UQ methodologies assess the uncertainties in the results obtained by numerical models, and analyse how those uncertainties affect the reliability of those outputs [2].

The UQ process identifies the sources of the uncertainties involved in a numerical simulation (input, model), and propagates those through the model using statistical methods to obtain an estimation of the output uncertainties. The process provides support to the designer to determine the inputs and model parameters that exert the most significant influence on the output uncertainties [2].

2.1.6 Uncertainty in statistical inference

A physics-based model is represented by $PM(\boldsymbol{x},\boldsymbol{\theta})$ that defines the relationship between an input vector of model parameters $\{\boldsymbol{x},\boldsymbol{\theta}\}$, and the output response vector $\boldsymbol{y}_{\mathrm{m}}$. The vector of model parameters is described by $\boldsymbol{x}$, representing a vector with fixed properties known in advance, and a vector of uncertain model parameters $\boldsymbol{\theta}$. In practice, the model response $\boldsymbol{y}_{\mathrm{m}}$ will be different to the true output response $\boldsymbol{y}$ of the corresponding real system because of three different types of uncertainties [2]: two of them related to the model (described below), and the last one related to the measurements. The uncertainties related to the measurements are mainly caused by the sensor noise, and will lead to a discrepancy $\boldsymbol{\varepsilon}_{\mathrm{obs}}$ between the real system response and the measurements $\boldsymbol{y}_{obs}$, so that [2]:

$$\boldsymbol{y}_{obs} - \boldsymbol{y} = \boldsymbol{\varepsilon}_{obs} \tag{1}$$

The uncertainties related to the model can be distinguished into modelling uncertainties and the uncertainties associated with the model parameters. Modelling uncertainty represents simplifications, modelling assumptions and numerical approximations that would lead to a discrepancy between the real system, and the model that represents it. These uncertainties would cause a discrepancy $\boldsymbol{\varepsilon}_{\mathrm{m}}$ between the real system response, and the response yielded by the model, such that [2]:

$$\boldsymbol{y}_{\mathrm{m}} - \boldsymbol{y} = \boldsymbol{\varepsilon}_{\mathrm{m}} \tag{2}$$

The uncertainties described in eq.(1) and eq.(2) are usually accounted for by using an additive error modelling, such that [2]:

$$\boldsymbol{y}_{\mathrm{obs}} = PM(\boldsymbol{x},\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{3}$$

where $\boldsymbol{\varepsilon}$ indicates the total prediction error caused by model uncertainties $\boldsymbol{\varepsilon}_{\mathrm{m}}$ (e.g., choosing Euler-Bernoulli beam theory to model beam behaviour) and measurement errors $\boldsymbol{\varepsilon}_{\mathrm{obs}}$ (e.g., uncertainty in

the modal properties due to noisy measurements), which are usually assumed to be independent and identically distributed [2].

The main task in the Bayesian model updating framework [14] is then to update knowledge on the uncertain model parameters $\boldsymbol{\theta}$ of the physics-based model by using measurements $\boldsymbol{y}_{obs}$ taken on the real system.

2.2 Physics-based model

To assess the general techniques developed in this thesis for structural dynamics under limited data, some simple physics-based models have been used. In this subchapter, a physics-based model, composed of simple structural elements, such as Euler-Bernoulli beams was chosen. Those elements are able to represent the physical behaviour of coupled structural elements without adding further complexity into the model. It consists of a coupled beam structure modelled by two beams connected using translational, rotational and shear springs as shown in fig. 4. A more detailed explanation on this physics-based model can be found in paper V.

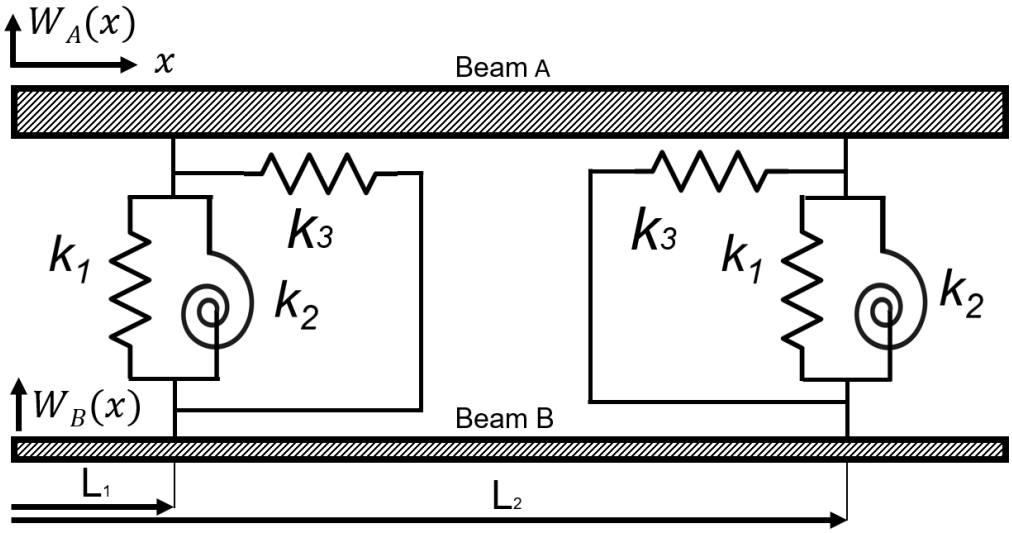


Fig. 4. Example of theoretical model of a coupled beam structure used in [39].

Models of greater complexity are not considered, as the emphasis of this project is the development of general techniques to be used in structural dynamics under limited data, and not the modelling of structures. The solution for more complex and larger structures may be obtained by generalising the solution obtained for the simplified problem. Other numerical examples have also been used, and are shown in the attached papers.

2.2.1 Motivation for using this physics-based model

In engineering, it is common to find structures that are composed of structural elements coupled together in such a manner that some of their elements cannot be directly observed. However, changes/damages on those inaccessible elements may compromise the overall performance of the structure, causing unexpected failure during operating conditions. The layout of these structures introduces a limitation on the number of observations that may be taken from them, their locations, and the time periods between those observations. As it frequently occurs in engineering that measurements can only be taken in certain parts of structures, this is a situation of practical interest.

A relevant example is found in the oil and gas industry, where drill strings (fig. 5) consisting of an arrangement of multiple elements connected at their edges forming a string are used for drilling. Each element is made of two concentric pipes of circular section connected at their ends. The internal part of the element is called chassis and the external one collar. Laboratory tests at remote locations, and the disassembly of the tool are necessary to access the chassis. This operation is costly and time consuming. Therefore, vibration-based measurements must only be taken on the collar.

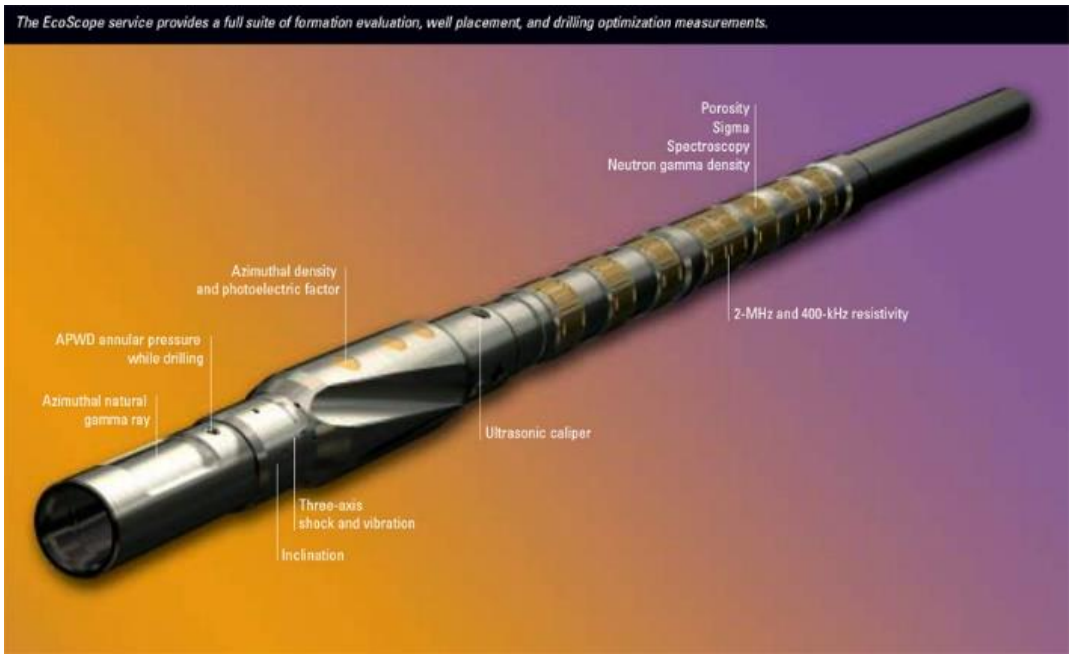


Fig. 5. Collar with EcoScope logging-while-drilling service [40].

Understanding if the techniques developed in this thesis can be utilised for coupled structures, in which some of the elements are not observable, would be of great interest. This would allow the validation of the methods developed on a simplified case that simulates a real-world engineering problem.

2.2.2 Finite Element Analysis (FEA): numerical model

In this subchapter, a FEA model is developed for the characterisation of the dynamical behaviour of the coupled beam structure. This baseline model may be used to produce codes where further complexity (different fault behaviours) may be introduced into the coupled beam structure.

The Finite Element Method (FEM) is a numerical method used for solving partial differential equations [41]. It divides a large system into finite elements, and subsequently it produces a large system of equations to model the entire problem. One dimensional Euler-Bernoulli finite elements (fig. 6), with one node at each end, are used to model each beam of the assembly. As shown on fig. 6, each node has two (translation (u_1, u_3) and rotation (u_2, u_4)) degrees of freedom (DOFs), therefore, each element has a total of four DOFs.

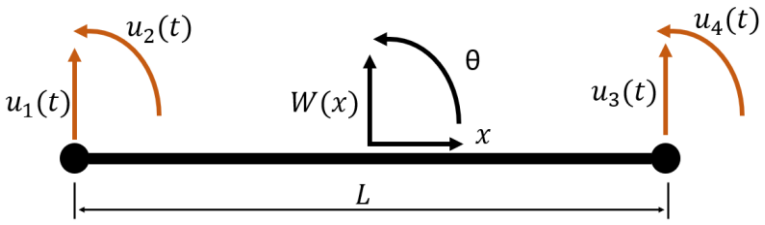
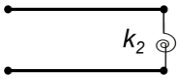
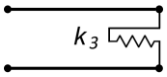
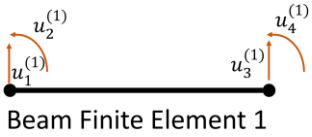


Fig. 6. Geometry of a finite beam element.

Using the Lagrange equations [42], the elemental matrices for the stiffnesses of the springs are derived. The stiffness and mass matrices for beam elements are calculated using third order shape functions as described in [41], and summarised in Table 1. The beam finite element 1 represents an element of the system.

Table 1: Elemental matrices. The length of the beam element is l

Matrices	Local Coordinates	Sketch
$[k]_{spring\_k_1} = \begin{pmatrix} k_1 & -k_1 \\ -k_1 & k_1 \end{pmatrix}$	$\begin{pmatrix} u_3^{(1)} \\ u_3^{(2)} \end{pmatrix}$	
$[k]_{spring\_k_2} = \begin{pmatrix} k_2 & -k_2 \\ -k_2 & k_2 \end{pmatrix}$	$\begin{pmatrix} u_4^{(1)} \\ u_4^{(2)} \end{pmatrix}$	
$[k]_{spring\_k_3} = \frac{k_3}{4}\begin{pmatrix} t_A^{\,2} & t_A t_B \\ t_A t_B & t_B^{\,2} \end{pmatrix}$	$\begin{pmatrix} u_4^{(1)} \\ u_4^{(2)} \end{pmatrix}$	
$[m]_{beam\_element} = \frac{\rho twl}{420}\begin{pmatrix} 156 & 22l & 54 & -13l \\ 22l & 4l^2 & 13l & -3l^2 \\ 54 & 13l & 156 & -22l \\ -13l & -3l^2 & -22l & 4l^2 \end{pmatrix}$	$\begin{pmatrix} u_1^{(1)} \\ u_2^{(1)} \\ u_3^{(1)} \\ u_4^{(1)} \end{pmatrix}$	 Beam Finite Element 1
$[k]_{beam\_element} = \frac{EI}{l^3}\begin{pmatrix} 12 & 6l & -12 & 6l \\ 6l & 4l^2 & -6l & 2l^2 \\ -12 & -6l & 12 & -6l \\ 6l & 2l^2 & -6l & 4l^2 \end{pmatrix}$	$\begin{pmatrix} u_1^{(1)} \\ u_2^{(1)} \\ u_3^{(1)} \\ u_4^{(1)} \end{pmatrix}$	

2.3 Bayesian Inference

2.3.1 From prior to posterior

Compared to previous assumptions or knowledge, the Bayesian Inference framework generates an updated statistical model able to combine prior assumptions with new knowledge.

Measurements from some observable quantities $\boldsymbol{y}_{obs}$, and a physics-based model, for which some latent parameters $\boldsymbol{\theta}$ are described by prior pdfs, are combined in the Bayesian model updating approaches to update our knowledge on these latent parameters [14,15]. More specifically, the information about the uncertain parameters $\boldsymbol{\theta}$, before observations $\boldsymbol{y}_{obs}$ are collected, is represented using a prior probability density function $p(\boldsymbol{\theta})$. The likelihood function $p(\boldsymbol{y}_{obs}|\boldsymbol{\theta})$ defines the ability to describe the degree of adequacy of the model, given the collected $\boldsymbol{y}_{obs}$ observations, and the vector of uncertain parameters $\boldsymbol{\theta}$. Information obtained from the observations $\boldsymbol{y}_{obs}$ (e.g., modal parameters), and from the physics-based model, is used to define the likelihood function $p(\boldsymbol{y}_{obs}|\boldsymbol{\theta})$.

Being the likelihood function $p(\boldsymbol{y}_{obs}|\boldsymbol{\theta})$ now available, using the Bayes theorem, the posterior probability density function $p(\boldsymbol{\theta}|\boldsymbol{y}_{obs})$ of the uncertain parameters, may be calculated [14,15]:

$$p(\boldsymbol{\theta}|\boldsymbol{y}_{\text{obs}}) = \frac{p(\boldsymbol{\theta})p(\boldsymbol{y}_{\text{obs}}|\boldsymbol{\theta})}{p(\boldsymbol{y}_{\text{obs}})} \tag{4}$$

The term $p(\boldsymbol{y}_{\text{obs}})$ is called the evidence, and it is used as a normalising constant of the posterior distribution. For the cases where likelihood and prior are not part of the conjugate family, numerical integration is required to obtain an approximation to the posterior [43].

2.3.2 Bayesian utility functions

In [44] the foundations for Bayesian experimental design are established. The Bayesian Optimal Design approach requires the definition of a utility function $U(\xi, \theta, y)$. This utility function represents the value of selecting from the design space Ξ, a design ξ, that produces data y in a model with parameters θ.

The object of the Bayesian Optimal Design is the search in the design space Ξ of the design ξ^* maximising the utility function $U(\xi, \theta, y)$ w.r.t. the model parameters θ and the data y. Therefore, for some data y the design ξ^*, is the one that maximises the expected utility posterior.

Monte Carlo approaches, are often necessary for the calculation of the Bayesian utility functions, as closed form solutions of the posterior pdf are only available for very particular cases. For the cases where no closed form expression of the posterior pdf is available, computational methods are used to get samples from the posterior or to estimate it. For each plausible data set y obtained from a sample of the prior, the posterior distribution has to be calculated, this means that a great number of posterior pdfs have to be computed [45]. Therefore, methods are currently being developed to reduce the computational cost incurred from the calculation of these posterior pdfs for any given prior and physics-based model.

For the cases where no closed expression of the posterior pdf is available, it may be approximated using several methods such as: Laplace [46], Numerical Quadrature [47] and Monte Carlo [48]. The traditional method of posterior distribution estimation for Bayesian utility function assessments is MCMC [49]. The disadvantage of the use of MCMC for calculating predictions of the posterior pdfs is the significant computational cost of the very many iterations needed during the Bayesian experimental design process.

Different experimental problems require different utility functions $U(\xi, \theta, y)$. These utility functions $U(\xi, \theta, y)$ should be defined considering the objectives of the experimental design. In other words, some designs may show effectiveness for prediction of future observations, but may produce poor estimations of the model parameter [45].

A commonly used Bayesian utility function is the expected information gain, also known as mutual information, that quantifies the mutual dependence of two random variables. This utility function is closely linked to the information theory concept of entropy, that for a random variable measures the amount of information that contains. Abundant literature about efficient prediction of model parameters using the concept of mutual information is available [46,50–52].

2.3.3 Variational Inference

Sampling based approaches, for an accurate estimation of the posterior distribution, depend upon a high number of samples, and therefore, imply a high computational cost. However, they usually imply higher flexibility and applicability to a greater variety of models than other Bayesian Inference techniques [19].

The applicability of sampling methods becomes more limited as engineering problems grow in complexity, requiring more intricate models to describe their physical behaviour [53,54]. Those high-dimensional and multi-modal engineering problems require the evaluation of multiple expensive likelihoods for model parameter inference [53,54]. In VI an alternative approach that allows to bypass the computation of the evidence term is taken [55]. The Kullback-Leibler (KL) divergence between the best member of the assumed family and the posterior density is minimised. As shown in greater detail below, in VI, with the objective of obtaining the posterior distribution, the statistical inference problem is transformed into an optimization problem.

In VI [55] a family of densities Q is postulated as an approximation to the posterior density $p(\theta \mid y_{\text{obs}})$, where the optimization scheme chooses the member of the family $q(\theta)$ that is 'closest'

to the posterior density. The choice of members of the family is performed in such a manner, that it is flexible enough to capture the posterior density, and simple enough to be optimised [55]. In VI, 'over-fitting' the posterior density by using highly flexible distributions is not a problem, as the highly flexible distributions allow better approximations of the true posterior distribution [56].

VI chooses the optimal member $q(\boldsymbol{\theta})$ of the family Q by solving the following optimisation problem [55]:

$$q^*(\boldsymbol{\theta}) = \underset{q(\boldsymbol{\theta})\in\mathcal{Q}}{\arg\min}\, KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) \tag{5}$$

where KL is the Kullback-Leibler divergence (also called relative entropy), defined as:

$$KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) \equiv \int_{\mathrm{R}^D} q(\boldsymbol{\theta}) \log\left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})}\right) d\boldsymbol{\theta} \tag{6}$$

and R^D is given by the real coordinate space of dimension D.

The KL divergence is a measure of how different a distribution is from a second reference distribution. The KL divergence is always non-negative $KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) \geq 0$ and is non-symmetric (i.e., $KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) \neq KL(p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}}) \| q(\boldsymbol{\theta}))$). When the distributions are the same i.e., $q(\boldsymbol{\theta}) \equiv p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})$, the KL divergence is minimised, and therefore, it returns a value of zero [57].

However, the calculation of the optimal member $q(\boldsymbol{\theta})$ of the family of densities Q that minimises the $KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}}))$ in eq.(5) cannot be analytically computed, as the evidence term $p(\boldsymbol{y}_{\text{obs}})$ should be known, as shown below [55]:

$$KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) = \mathbb{E}_q\left[\log q(\boldsymbol{\theta})\right] - \mathbb{E}_q\left[\log p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})\right] \tag{7}$$

$$KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) = \mathbb{E}_q\left[\log q(\boldsymbol{\theta})\right] - \mathbb{E}_q\left[\log p(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}})\right] + \log p(\boldsymbol{y}_{\text{obs}}) \tag{8}$$

where $\mathbb{E}_q$ indicates the expectation with respect to $q(\boldsymbol{\theta})$.

As the KL divergence between the posterior approximation $q(\boldsymbol{\theta})$ and the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})$ cannot be directly computed, the objective function is changed to the evidence lower bound (ELBO), which is equivalent to the negative KL divergence between the prior $p(\boldsymbol{\theta})$ and posterior approximation $q(\boldsymbol{\theta})$, plus a constant [55]:

$$ELBO(q) = \mathbb{E}_q\left[\log p(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}})\right] - \mathbb{E}_q\left[\log q(\boldsymbol{\theta})\right] = \mathbb{E}_q\left[\log p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta})\right] - KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta})) \quad (9)$$

Therefore, maximising the ELBO objective function in eq.(9), is equivalent to the minimisation of the KL divergence between the posterior approximation $q(\boldsymbol{\theta})$ and the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})$ [53]:

$$\arg\max_{q(\boldsymbol{\theta}) \in \mathcal{Q}} ELBO(q(\boldsymbol{\theta})) \equiv \arg\min_{q(\boldsymbol{\theta}) \in \mathcal{Q}} KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) \quad (10)$$

In engineering problems, due to the complexity of the physics-based models, the term $\mathbb{E}_q\left[\log p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})\right]$ lacks a suitable form to permit a closed-form solution of eq.(10).

Typically, factorised posterior distributions with the form $q(\boldsymbol{\theta}) = \prod_{i=1}^{D} q_i(\theta_i)$, are chosen, such as in the mean-field VI [58,59]. In that expression, $q(\boldsymbol{\theta})$, is the specific parametric form of a member of the family Q, that most effectively approximates the posterior, and the number of parameters to be inferred is D. The optimisation process in eq.(5) gets streamlined when those factorised posteriors are used, but this occurs at the expense of the accuracy of the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})$ evaluation. However, in practice, calculating the likelihood requires the execution of a costly physics-based model, leading to a substantial rise in computational expenses, to the extent that it becomes impractical, even when employing the mean-field algorithm [55].

In the field of computer science, various methods [53,54,60–64] have been devised to decrease the number of runs of the physics-based model to be performed, and to eliminate the necessity of a closed-form solution for the ELBO.

2.4 Gradient Flows

With the purpose of facilitating the reader's understanding of the paper V in chapter 6, a brief introduction to the concepts of Gradient Flows and Wasserstein Gradient Flow is given in this subchapter.

The mathematical tools called Gradient Flows (also known as steepest descents) are curves $\boldsymbol{\theta}(t)$ that follow the direction of steepest descent of a functional $E:\mathbb{R}^D \to \mathbb{R}$ [65,66]. In mathematical terms, $\mathbb{R}^D$ is a vector field, and the Gradient Flow is the flow of that vector field. The problem requires to find the $\boldsymbol{\theta}(t)$ that produces the fastest minimisation of E starting from $\boldsymbol{\theta}_0$. This is, the Gradient Flow of $E(\boldsymbol{\theta})$, is the solution $\boldsymbol{\theta}(t)$ of the following partial differential equation:

$\boldsymbol{\theta}'(t) = -\nabla E(\boldsymbol{\theta}(t))$ for $t \in \mathbb{R}^+$, for a given initial condition $\boldsymbol{\theta}_0 \in \mathbb{R}^D$, where the operator ∇ is defined as the gradient operator. A comprehensive review of Gradient Flows may be found in [65,66].

For the more specific case where $\mathbb{R}^D$ represents a Euclidian space of dimension D, the Gradient Flow can be expressed by an Ordinary Differential Equation (ODE). A gradient descent minimization algorithm can be defined through discretization of the previously mentioned ODE.

For example, if the ODE is discretized using the Euler method, the following equation is obtained:

$$\frac{\boldsymbol{\theta}^{n+1} - \boldsymbol{\theta}^n}{\tau} = -\nabla E\left(\boldsymbol{\theta}^{n+1}\right) \tag{11}$$

that can be written as:

$$\nabla\left(\frac{\left\|\boldsymbol{\theta} - \boldsymbol{\theta}^n\right\|^2}{2\tau} + E(\boldsymbol{\theta})\right)\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{n+1}} = 0 \tag{12}$$

where $\boldsymbol{\theta}^{n+1}$ can be written as the solution to a minimisation problem (Jordan Kinderleher Otto (JKO) scheme [67]):

$$\boldsymbol{\theta}^{n+1} = \arg\min\left(\frac{\left\| \boldsymbol{\theta} - \boldsymbol{\theta}^{n} \right\|^{2}}{2\tau} + E\left(\boldsymbol{\theta}\right) \right) \tag{13}$$

At this stage, the discrete scheme does not include gradients, and it can be generalised to a more general metric space $\left(\mathcal{X}, d\right)$ where the space $\mathcal{X}$ is endowed with the metric $\boldsymbol{d}$.

$$\boldsymbol{\theta}_{\tau}^{n+1} \in \arg\min\left(\frac{d\left(\boldsymbol{\theta}, \boldsymbol{\theta}_{\tau}^{n}\right)^{2}}{2\tau} + E\left(\boldsymbol{\theta}\right) \right) \tag{14}$$

Therefore, the concept of Gradient Flow can also be applied in other types of metric spaces [65,66].

A possible time discretisation is given by:

$$\boldsymbol{\theta}_{\tau}\left(t\right) = \boldsymbol{\theta}_{\tau}^{n} \text{ where } t \in \left[\left(n-1\right)\tau, n\tau\right] \tag{15}$$

The discretisation scheme used to approximate the Gradient Flow converges to the continuous flow in the limit $\tau \to 0$.

2.4.1 Wasserstein Gradient Flows

In paper V, it is of particular interest the study of Wasserstein Gradient Flows. The Wasserstein Gradient Flow (WGF) is a specific type of Gradient Flow used in the optimisation of functionals of probability measures [68]. The WGF applies on a probability measures space where a 2-Wasserstein metric has been defined.

The Table 2 below provides a comparison of metrics, formulation of the minimisation problem, formulation of the Gradient Flows, and discrete formulations for the Euclidean and Wasserstein-2 space also called the Probability Space.

Table 2: Euclidean versus Probability Space minimisation comparison

	Euclidean Space	Probability Space
Minimisation problem	$x^* = \arg\min_{x \in \mathbb{R}^n} E(x)$	$\rho^* = \arg\min_{\rho \in \mathbb{P}(\Omega)} E(\rho)$
Gradient Flow (strong formulation)	$\begin{cases} x'(t) = -\nabla E(x(t)) \\ x(0) = x_0 \end{cases}$	$\begin{cases} \partial \rho_t = -\nabla E(\rho_t) \\ \rho_0 \text{ is given} \end{cases}$
JKO scheme (weak formulation)	$x_\tau^{n+1} = \arg\min_x \left\{ E(x) + \frac{d(x, x_\tau^n)^2}{2\tau} \right\}$	$\rho_\tau^{n+1} = \arg\min_\rho \left\{ E(\rho) + \frac{\mathcal{W}_2^2(\rho, \rho_\tau^n)}{2\tau} \right\}$
Metric	$d(x,y)^2 = \|x-y\|^2$	$\mathcal{W}_2(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int \|x-y\|^2 \pi(dx,dy)$

Chapter 3 - Limited information at the design stage

The UQ techniques are frequently used for the assessment of the pdfs of the natural frequencies, damping ratios, and modal shapes of engineering structures. To avoid costly modifications of the structure after it is built, at the design stage, the engineer should perform assessments to ensure the robustness of the structure's design. These calculations require the evaluation of the pdfs of the modal properties of the structure so that its dynamic properties can be appropriately modeled.

However, existing methods of UQ require to specify the different sources of uncertainties [69], and the description of the model's parameters uncertainty, which are often hard to determine. Using the existing UQ parametric methods, the uncertainty of the parameters has to be propagated through the equations of motion, becoming this propagation a significant drawback when the model of the system is highly complex or unknown, a situation that may introduce modelling errors that affect the overall results.

In the following subchapter, paper I develops an approach that employs a non-parametric model of uncertainty to calculate the pdfs of the modal parameters by exploiting Random Matrix Theory (RMT) results [6,69]. Using a non-parametric approach, it avoids the need to specify the different sources of uncertainties in advance.

3.1 Paper I – 'On the Combination of Random Matrix Theory with Measurements on a Single Structure'

This paper is reproduced as originally published with minor revisions:

'Igea F., Chatzis M.N.., Cicirello A., On the Combination of Random Matrix Theory with Measurements on a Single Structure. ASME J. Risk Uncertainty Part B, 2022'.

https://doi.org/10.1115/1.4054172

On the Combination of Random Matrix Theory with Measurements on a Single Structure

Felipe Igea[1]*, Manolis N. Chatzis[1] and Alice Cicirello[1,2]

[1] Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

[2] Department of Engineering, Faculty of Civil Engineering and Geosciences, Department of Engineering Structures, Delft University of Technology, Stevinweg 1, Delft 2628, NL

*Corresponding author: felipe.igea@eng.ox.ac.uk

Abstract

An approach is proposed for the evaluation of the probability density functions (pdfs) of the modal parameters for an ensemble of nominally identical structures when there is only access to a single structure and the dispersion parameter is known.

The approach combines the Eigensystem Realization Algorithm on sets of dynamic data, with an explicit non-parametric probabilistic method. A single structure, either a mathematical model or a prototype, are respectively used to obtain simulated data or measurements that are employed to build a discrete time state-space model description. The dispersion parameter is used to describe the uncertainty due to different sources such as the variability found in the population and the identification errors found in the noisy measurements from the experiments. With this approach, instead of propagating the uncertainties through the governing equations of the system, the distribution of the modal parameters of the whole ensemble is obtained by randomising the matrices in the state-space model with an efficient procedure.

The applicability of the approach is shown through the analysis of a 2-DoF mass-spring-damper system and a cantilever system. The results show that if the source of uncertainty is unknown and it

is possible to specify an overall level of uncertainty, by having access to a single system's measurements it is possible to evaluate the resulting pdfs on the modal parameters. It was also found that high values of the dispersion parameter may lead to non-physical results such as negative damping ratios values.

1 Introduction

The quantification of the uncertainties on the modal parameters during the design process of structures is of great interest for the assessment of their dynamic performance [1,2]. The measurement of the dynamic response across nominally identical structures can reveal largely different modal parameter estimates due to uncertainties originated by the variability of the manufacturing processes of structural components, boundary conditions and assemblage [3,4]. Identifying the distribution of the modal parameters across the ensemble of nominally identical structures would enable the selection of designs that are robust to these uncertainties, avoiding extensive modifications of the manufactured product [5]. This paper proposes an approach for the evaluation of the probability density functions (pdfs) of the natural frequencies, damping ratios and modal shapes of an ensemble of nominally identical linear time-invariant systems, for cases where there is only access to a single structure and additional information is known in the form of the so-called dispersion parameter. This single structure, represented either as a mathematical model or a prototype, is used to obtain simulated data or measurements, respectively, which in turn are employed to build a discrete time state-space model description. This model is then used to efficiently assess the effects on the modal properties of different levels of uncertainties, represented through suitably chosen dispersion parameters. This method is also of interest within the Bayesian Inference framework, as it can be used to build a prior distribution consistent with the available information, for the cases where there would be insufficient information in the data so that resulting likelihood cannot overrule the prior assumptions [6].

At early design stages (before a prototype is being built), when a physics-based model is developed, uncertainties on the model parameters are usually described using probabilistic [7–9] or non-probabilistic uncertainty descriptions [10]. These uncertainties are then propagated through the

equation of motions to yield the corresponding description of the response. This is the so-called parametric model of uncertainty. However, the choice of the uncertain parameters and their description would directly affect the resulting distribution of the resulting modal parameters [1,2,8]. An alternative approach is to employ a non-parametric model of uncertainty by exploiting Random Matrix Theory (RMT) results [1,11]. This is useful as non-parametric methods [11] avoid the need to specify the uncertainties' sources and the description of the parameters model's uncertainty, which are often hard to determine. Moreover, using these methods modelling errors may also be accounted for [1]. A review of different random matrices and their properties is given in [12]. The application of these matrices to engineering problems, and in particular, for uncertainty quantification has been the subject of much recent research [13–16]. Broadly speaking, random matrices can be applied using an implicit approach that is based on the derivation of analytical results given a set of assumptions, e.g., by assuming that the physical properties of a structural component are sufficiently random so that the statistical distribution of the natural frequencies and mode shapes tends to a universal distribution associated with the Gaussian Orthogonal Ensemble (GOE) of random matrices [17,18], or also by using an explicit approach that is based on the use of Monte Carlo simulations to propagate uncertainty [1,11].

Vishwajeet et al [19] have recently used results from RMT in combination with the system identification (SI) method ERA for the calculation of the analytical expression of the pdfs of the singular values of the Hankel Matrix. This was done by assuming that the elements of the Hankel Matrix were Gaussian random variables, and therefore, the Hankel Matrix times its own conjugate transpose conforms with the non-central Wishart distribution. This assumption might not be always valid for linear-time-invariant systems, as noise is non-Gaussian in nature, and often only approximated as Gaussian. Moreover, the prediction of the uncertainty of the properties of the modal parameters associated to a priori knowledge of uncertainties related to manufacturing variability, boundary conditions, and assemblage of other nominally identical structures, using measurements taken from a single structure during the design stage, has not been addressed with current methods.

Quantifying the effect of those uncertainties on the untested members of the population becomes paramount if reliable performance assessments of the structures must be produced. This is the focus of the present paper, where no assumption is made on the type of the distribution followed by the elements in the Hankel Matrix, and a non-parametric explicit implementation of RMT is used to build the ensemble employed to evaluate the pdfs of the modal parameters, using the dispersion parameter to control the level of uncertainty without specifying its sources. Relevant uncertainty propagation approaches based on this type of explicit implementation of RMT are briefly reviewed in what follows.

Soize [1,11] pioneered an RMT explicit approach that models the system's matrices (mass, stiffness and damping) as random matrices. Using the maximum entropy method and the information available (system's matrices are symmetric positive-definite, the second order moment of their inverse and their mean matrices exist) an ensemble of random matrices can be used to model the uncertainty of the system's matrices in a non-parametric manner. This explicit implementation of RMT has been successfully applied for building stochastic models to quantify uncertainty in nominally identical systems by defining a single parameter, the so-called dispersion parameter [1,11]. The dispersion parameter is used to control the overall level of uncertainty in the random matrix without the need to specify its origin. This parameter can be used at the design stage to build an ensemble of nominally identical models/structures, and it can be updated once experimental observations become available [1,11]. The dispersion parameter is calculated by assuming that the nominal model/structure is available (or approximated), and that the experimental observations are different realizations of the ensemble of structures [1,20]. The dispersion parameter can be varied to create different ensembles and evaluate the effect on the produced results [21]. An example, where the dispersion parameter that controls the level of the uncertainty in the stiffness matrix, was identified from experimentally obtained frequency response functions (FRF) of six nominally designed aircraft T-tails, used for the calculation of the experimental modal parameters, is shown in paper [16]. Paper [14] illustrates another recent application using experimental data from booster pumps thermal units to identify the

dispersion parameter. This dispersion parameter is used to build a stochastic computational model that considers mode crossing and veering phenomena through the introduction of an adapted transformation for the calculated modal quantities. Legault et al [21] applied this non-parametric approach to simple numerical examples and studied its meaning and consequences for the output space (e.g., average modal density, dispersion relation, etc.), especially for large values of the dispersion parameters.

This current paper follows the explicit approach developed by Soize and comprehensively reviewed in [1,11], where the general applicability and proofs of the approach are given. In particular, with the proposed approach, starting from a set of measurements or simulated data obtained from either a mathematical model or a prototype, the Hankel Matrix is calculated. The matrix resulting from the multiplication of the Hankel Matrix times its own conjugate transpose is randomised using the normalized positive definite ensemble defined in [1,11]. For each realization of the ensemble, the Eigensystem Realization Algorithm (ERA) [22] is applied to identify the modal parameters (natural frequencies, damping ratios and modal shapes). ERA has been chosen for its simplicity, although its application is limited to the analysis of structures under some assumptions [22], such as impulsive excitations or free vibration. If the appropriate changes to the proposed method are considered, other loading conditions may be investigated, however, these changes are not explored in this paper. The results of each realization are then used to build the pdfs of the modal parameters of the ensemble. The proposed approach enables the assessment of the effect on the modal parameters' uncertainties of different dispersion parameter values and size of Hankel matrices.

It is worth noting that the pdfs of the modal parameters of an ensemble of nominally identical structures obtained with the proposed approach correspond to those that would be obtained from an ensemble of structures from a production line, such as cars, or by considering manufactured structures which are realizable in principle, but for which only one sample may be built, such as a bridge. This virtual ensemble accounts for a collection of linear structures (not all of them are necessarily existent but are physically feasible) in which the geometric, mechanical properties and boundary conditions

of all the elements of each structure are uncertain and conform to unknown distributions. The applicability of the proposed approach is shown using numerical applications that represent situations where the system is under free vibration.

This paper is structured as follows. The ERA method is reviewed in section 2. The steps involved on how to combine RMT and ERA for the quantification of the uncertainty of the modal parameters are found in section 3. The numerical results obtained are found in section 4. The physical consequences of the application of RMT are then discussed on section 5. The appendix provides a short summary of the ERA method.

2 Proposed approach

The proposed approach combines a SI technique with a non-parametric uncertainty description to quantify the effect of different levels of uncertainty on the modal properties of an ensemble of nominally identical structures, using probability density functions of these modal properties. In particular, by using measurements or simulated data obtained from a single structure, an ensemble able to encompass different sources of uncertainty is built as a product of the Hermitian matrices, which is the Hankel Matrix times its own conjugate transpose. The random matrix is specifically constructed to reflect a given level of uncertainty (expressed in terms of the so-called dispersion parameter) and to retain some properties of the Hermitian matrix (i.e., positive definiteness and the existence of the second-order moment and of its inverse). In particular, the normalised positive-definite random matrix (random matrix ensemble $SG+$) is chosen [11].

A schematic representation of the approach is shown in fig. 1. Each of the blocks of the approach are discussed in more detail in the following subsections.

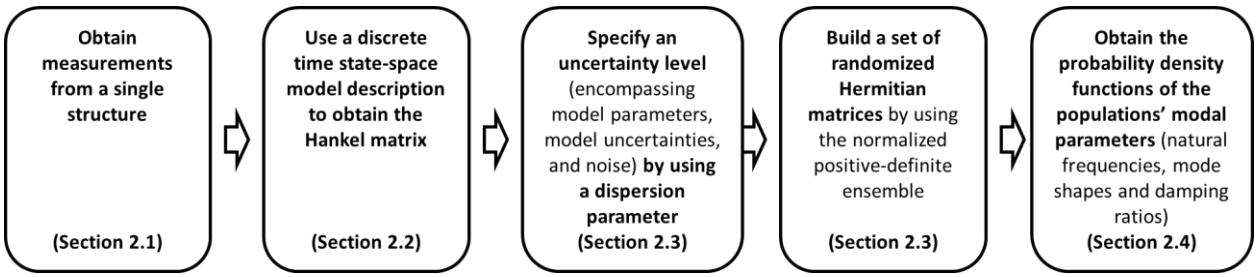


Fig. 1. Schematic representation of the approach.

2.1 Measurements on a single structure to construct a virtual ensemble

When the physics-based model of a structure is available, and the structure is not yet built, the model can be used for simulating its dynamic response, i.e., displacement, velocity or acceleration signals, using its nominal properties [11]. The outputs signals can then be contaminated with noise. Therefore, a set of response signals, ideal measurements, can be obtained by using the physics-based model. Alternatively, real measurements can be carried out on a prototype of the real structure or on a structure that is a member of an ensemble of structures that would be produced. These simulated data or measurements on a single structure can then be used to create the nominal model, that in conjunction with the dispersion parameters is used to construct the virtual ensemble. In this work, the virtual ensemble consists of a set of linear structures, described by a set of common geometric properties, mechanical properties, and boundary conditions, and where each realization corresponds to a different structure across the ensemble. Therefore, this ensemble is used to account for structure-to-structure response variability caused by different sources of uncertainty that may occur in the modelling and manufacturing process. In addition, it has to be considered that the variability that the virtual ensemble encompasses includes several sources of uncertainty, such as the noise of the experimental measurements and the uncertainties introduced by the modelling process.

Starting from the simulated or real set of measurements, a discrete time state-space model description can then be readily built or identified, respectively. The number of measurements (in terms of sensors) needed depends on the model order, number of modes and the complexity of the structure. The measurements used are those obtained from vibration-based experiments such as displacement,

velocity and acceleration readings.

In this paper, the ensemble is constructed starting from the physics-based model in order to ensure reproducibility of the results shown.

2.2 Evaluation of Hankel Matrix

Modal identification methods are used to obtain the modal properties (natural frequencies, mode shapes and damping ratios) of a system. The approach proposed in this paper requires the evaluation of the Hankel Matrix.

The Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$ of size $r \times p$ by s, contains time series data from measurement data [23], and it can be built as shown in eq.(1). The time series data available are arranged in r block rows and s block columns that are used to build the Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$. The values of r and s, are called the observability and controllability indices respectively, and are chosen according to the criteria explained in [22]. The output vector sequences $y_{\{k,p\}}$ contains the measurements read in p channels at different times t_k. Assuming free vibration conditions, the matrices $\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \ldots, \boldsymbol{y}_k$ are obtained.

$$\boldsymbol{H}_{rs}(k-1) = \begin{bmatrix} \boldsymbol{y}_k & \boldsymbol{y}_{k+1} & \cdots & \boldsymbol{y}_{k+s-1} \\ \boldsymbol{y}_{k+1} & \boldsymbol{y}_{k+2} & \cdots & \boldsymbol{y}_{k+s} \\ . & . & & . \\ . & . & & . \\ \boldsymbol{y}_{r+k-1} & \boldsymbol{y}_{r+k} & \cdots & \boldsymbol{y}_{r+k+s-2} \end{bmatrix} \quad (1)$$

In the present analysis, the Eigensystem realization algorithm (ERA) is chosen because of its simplicity when dealing with systems excited using an impulse force (e.g., hammer strike) or systems in free vibration [22].

However, other loading conditions may be considered if the appropriate changes to the proposed approach are implemented. For example, the combination of the Natural Excitation Technique with

ERA [24] allows to deal with situations where ambient excitation is observed. These methods are beyond the scope of the present work, and the reader is referred to [25].

2.3 Defining uncertainty by using a dispersion parameter δ_S and Matrix randomization

In parametric probabilistic approaches for uncertainty quantification [7,8], the different types of uncertainties (model noise, measurement noise, population uncertainty, etc.) are described by using some assumed pdfs. The uncertainties can then be propagated through the relevant equations describing the behaviour of the system, and the effects of the uncertainties on the parameters of interest are assessed.

In this paper, an alternative approach that accounts for the effects of uncertainties without the need to make explicit assumption on the pdf of specific model parameters is considered. In particular, a non-parametric technique based on RMT is used by following the explicit approach proposed by Soize [1,11]. A dispersion parameter δ_S is used to build the random matrix $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ that will be used to create the virtual ensemble.

The dispersion parameter value encompasses the overall level of uncertainty caused by different sources of uncertainty that may be present, such as modelling errors, manufacturing variability, identification errors, errors in the recorded signals, and variability of the boundary conditions among the members of the ensemble.

The prior estimation of the dispersion parameter δ_S, can be classified into three cases:

i. The first case occurs when no prior data are available, and two approaches can be considered. The nominal value and the mean matrix of the stochastic model are the same. In this case, the dispersion parameter δ_S is a variable that is used for a sensitivity analysis that relates the level of uncertainties to the stochastic solution [1]. The sensitivity analysis is performed by varying the uncertainty levels controlled by the dispersion parameter value δ_S in a predefined range

and observing how the pdfs of the modal parameters (natural frequencies, mode shapes and damping ratios) change [1].

Alternatively, Legault et al. [21], define a dispersion parameter δ_S that is set a priori based on an expected level of uncertainty. For example, low, mid and high values of the dispersion parameter were assigned in [21] to evaluate the effects on their chosen properties. When high uncertainty is present, higher dispersion parameter values are used.

ii. The second case occurs when prior data (either measurements taken on the ensemble of structures or an existent computational model with specified parametric uncertainties) are available [1]. For this case, from the observed data, it is possible to update the mean matrix, and using the least-square method or the maximum likelihood method to obtain the value of the dispersion parameter δ_S that optimizes the employed objective function (e.g., minimizing the difference on the resulting coefficient of variation on the first natural frequency). A detailed explanation about both methods may be found in [1].

iii. The third case occurs when in addition to the parametric prior information (case (ii)) there is also a statistically independent non-parametric prior information (case (i)). For example, the uncertainty in the measurements noise can be specified as a zero mean Gaussian with a given variance, and the modelling error is specified with a dispersion parameter δ_A. The goal is to obtain an overall dispersion parameter $\delta_s = \delta_D + \delta_a$, where δ_D is calculated as shown in case (ii) without accounting for the non-parametric uncertainty.

The three cases above are investigated in the first numerical application (section 4.1). The dispersion parameter quantifies the quadratic distance from the random matrix $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ to the identity matrix I_n as shown by eq.(2) below [1,11].

$$\delta_S = \left(\frac{1}{n} E\left\{ \left\| \hat{\boldsymbol{G}}_{\boldsymbol{S}} - \boldsymbol{I}_n \right\|_F^2 \right\} \right)^{1/2} \qquad (2)$$

where $\| \ \|_F$ denotes the Frobenious norm and n is the number of rows or columns of the matrix $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$. The dispersion parameter takes values within the inequality shown by the eq.(3) below [11]:

$$0 < \delta_S < \sqrt{(n+1)(n+5)^{-1}} \qquad (3)$$

In general, as it is not possible to have multiple elements of the virtual ensemble or prior measurements in the design stage, in this paper, the strategy (i) described in [21] is followed, and different a priori values of the dispersion parameter δ_S are used to evaluate the effect of the dispersion value on the resulting pdfs of the modal properties of the virtual ensemble. In particular, low and high values of the dispersion parameters are considered.

For a given full row rank Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$, a Hermitian matrix $\boldsymbol{S}(k-1)$ is calculated [19] using:

$$\boldsymbol{S}(k-1) = \boldsymbol{H}_{rs}(k-1)\boldsymbol{H}^{*T}{}_{rs}(k-1) \quad (4)$$

A Hermitian matrix $\boldsymbol{S}(k-1)$ that is symmetric positive-definite may be Cholesky factorized [1,11] as shown in the following equation:

$$\boldsymbol{S}(k-1) = \boldsymbol{L}_S^{*T}\boldsymbol{L}_S \qquad (5)$$

where $\boldsymbol{L}_S$ is an upper triangular matrix.

In order to generate the ensemble of matrices $\hat{\boldsymbol{S}}(k-1)$, the randomization of the matrix $\boldsymbol{S}(k-1)$ is performed according to [1,11]:

$$\hat{\boldsymbol{S}}(k-1) = \boldsymbol{L}_S^{*T}\hat{\boldsymbol{G}}_{\boldsymbol{S}}\boldsymbol{L}_S \qquad (6)$$

where $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ is a normalized positive-definite matrix [1,11]. The expected value of the $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ matrix is equal to the identity matrix I_n. As shown in eq.(2), the value of the dispersion parameter depends on the size number n of the matrix $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$.

The randomizing matrix $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ is built using [1,11]:

$$\hat{\boldsymbol{G}}_{\boldsymbol{S}} = \hat{\boldsymbol{L}}_{\boldsymbol{GS}}^{*T} \hat{\boldsymbol{L}}_{\boldsymbol{GS}} \qquad (7)$$

Where the elements ik (row, column) of the random upper triangular matrix $\hat{\boldsymbol{L}}_{\boldsymbol{GS}}$ are defined through the equations below [11]:

$$if\ i<k, \quad \hat{\boldsymbol{L}}_{\boldsymbol{GS,ik}} = \sigma_{\boldsymbol{S}} U_{ik} \qquad (8)$$

$$if\ i=k, \quad \hat{\boldsymbol{L}}_{\boldsymbol{GS,ik}} = \sigma_{\boldsymbol{S}} (2V_i)^{1/2} \qquad (9)$$

where

$$\sigma_{\boldsymbol{S}} = \frac{\delta_{\boldsymbol{S}}}{(n+1)^{1/2}} \qquad (10)$$

U_{ik} is a Gaussian random variable of unit variance and zero mean, and V_i is a Gamma random variable with parameters α , β [1,11] as:

$$\alpha = \frac{n+1}{2\delta_{\boldsymbol{S}}^2} + \frac{1-i}{2} \qquad (11),$$

$$\beta = 1 \qquad (12)$$

The choice of the normalized positive-definite matrix $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ is motivated by the use of the maximum entropy principle given the information available: there is a mean matrix $\boldsymbol{S}(k-1)$, it is positive definite, and the second-order moment of its inverse exists [1,11]. The theoretical statistical properties of the random matrix $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ can be found in [1,11].

2.4 Probability density functions of the modal parameters for an ensemble of nominally identical structures

To calculate a new matrix $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ in each realization of the ensemble, a Monte Carlo simulation in which independent random samples are drawn for the Gaussian random variable U_{ik} and the Gamma random variable V_i, is performed. The number of samples needed to generate the ensemble has to ensure that the pdfs of the uncertain parameters obtained, converge to their true pdf. This is achieved by checking the KL-divergence value of the pdfs of each identified parameter with respect to the pdfs for incrementally increasing number of simulations. The modal parameters (natural frequencies, damping ratios and modal shapes) estimated using ERA from each random realization of the system are used to approximate the pdfs of the virtual ensemble using a kernel density estimate [26].

3 Advantages and summary of the steps for the proposed method

The main advantages of the proposed approach when compared to forward uncertainty quantification methods with physics-based models, are:

(i) The different sources of uncertainties are not required to be specified explicitly.

(ii) The uncertainty in the parameters does not have to be propagated through the equations of motion.

(iii) In some cases, it may be difficult or impractical to model a structure/product based on physics. Consequently, the results obtained would have been significantly affected by the ensuing modelling errors. In these occasions, building the ensemble from the measurements of a prototype structure, i.e., a data-driven model, can be advantageous.

(iv) Compared to other SI methods that account for uncertainties, this methodology does not require to make explicit assumptions on the type of the pdfs of specific model parameters. Additionally, the pdfs of the modal parameters may be produced without specifying the origin of such uncertainties, and therefore, no assumptions on the sources of the uncertainties are required.

The results of this method may also be applied to anomaly detection techniques, where the evaluated ensemble's distributions of the modal parameters can be directly used.

The steps proposed in this methodology are:

Step 1: A Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$ of size r by s is built by arranging the outputs $\boldsymbol{y}_k$ (obtained from the measurement signals) as shown on eq.(1).

Step 2: For a given full row rank Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$, a Hermitian matrix $\boldsymbol{S}(k-1)$ is calculated using eq.(4).

Step 3: The resulting matrix $\boldsymbol{S}(k-1)$ is symmetric positive-definite and is Cholesky factorized using eq.(5), where $\boldsymbol{L_S}$ is an upper triangular matrix.

Step 4: The dispersion parameter value is defined. A realization $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ of the random matrix $\boldsymbol{G}_{\boldsymbol{S}}$ is calculated using eq.(7), the theoretical statistical properties of the random matrix $\boldsymbol{G}_{\boldsymbol{S}}$ can be found in [1,11]. The randomization of the matrix $\boldsymbol{S}(k-1)$ is performed according to eq.(6).

Step 5: The eigenvalues and eigenvectors of the matrix $\hat{\boldsymbol{S}}(k-1)$ are calculated. The eigenvalues and the eigenvectors of the matrix $\hat{\boldsymbol{S}}(k-1)$ are respectively the squared singular values and the left singular vectors of the Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$ [23]. The right singular vectors of the Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$, are given by the product matrix $\boldsymbol{H}_{rs}^{*T}(k-1)\boldsymbol{U}\boldsymbol{\Sigma}^{-1}$ [23].

Step 6: During step 5, the singular values of $\boldsymbol{H}_{rs}(k-1)$ are calculated. Physical singular values should be separated from spurious (mathematical) singular values related to the noise in this process that defines the model order of the system [23]. From a practical point of view the elimination of spurious modes may be performed using stabilization diagrams [27]. In the numerical simulations shown in sections 4 and 5 the number of singular values n kept are known.

Step 7: The procedures defined from Step 1 to Step 4 are also applied to the shifted Hankel Matrix $\boldsymbol{H}_{rs}(k)$. For every realization, the Gaussian random variable U_{ik} and Gamma random variable V_i are independently resampled to produce different realizations of the random matrices $\hat{\boldsymbol{S}}(k-1)$ and $\hat{\boldsymbol{S}}(k)$.

The randomization of $\hat{\boldsymbol{S}}(k)$ is carried out with the same dispersion parameter value δ_S but a different realization $\hat{\boldsymbol{G}}_{\boldsymbol{S}}$ is used. This is because although the measurements used to construct $\boldsymbol{S}(k-1)$ and $\boldsymbol{S}(k)$ are the same, they are obtained using the Hankel and the shifted Hankel matrices, respectively.

Step 8: Modal characteristics of the system (natural frequencies, modal shapes and damping defined in the Appendix where further equations and explanations about the referred quantities are provided) are calculated using eq.(13), eq.(14) and eq.(15) respectively [22].

$$\omega_i = \left|\lambda_{c(i)}\right| \qquad (13)$$

$$\boldsymbol{\phi}_i = \tilde{\boldsymbol{C}}\boldsymbol{v}_i \qquad (14)$$

$$\zeta_i = -\frac{\mathrm{Real}(\lambda_{c(i)})}{\left|\lambda_{c(i)}\right|} \qquad (15)$$

Each realization of the modal characteristics is stored.

Steps from 4 to 8 are repeated for a prescribed number of times for each realization of the random matrix $\hat{\boldsymbol{S}}(k-1)$ and $\hat{\boldsymbol{S}}(k)$. The modal parameters estimated from each random realization of the system (via a Monte Carlo simulation) are used to approximate the pdfs of the virtual ensemble using a kernel density estimate [26].

It is worth mentioning that the nominal damping matrix of the studied structure is proportional. For the calculation of the damping ratios in eq.(15) only the real part is taken, this assumes that Rayleigh or proportional damping is present. However, the perturbations introduced by the randomization process introduced by the RMT do not necessarily introduce proportional damping in the system.

Therefore, results obtained from eq.(15) for high values of the dispersion parameter may produce non-physical values, as shown in section 4.

4 Numerical results

This section shows the applicability of the proposed approach by considering two different numerical simulations: the first a two-degree-of-freedom (2-DoF) mass-spring-damper (subsection 4.1), and the second a uniform mass cantilever system (subsection 4.2). For each simulation, two different dispersion parameter values that translate to a low and high level of uncertainty are considered. In the cantilever system the physical model that generated the signal, and the identified model have a different state order. The dispersion parameter is used to calculate the realizations of the random matrices $\hat{\boldsymbol{S}}(k-1)$ and $\hat{\boldsymbol{S}}(k)$. The values of the modal parameters obtained in each realization are used to calculate the probability distributions of the modal parameters of the virtual ensemble of these systems. The effects of the value of the dispersion parameter, the state order and their probability distributions in this approach are also shown.

4.1 Two degree of freedom system

Let us consider the 2-DoF mass-spring-damper system shown in fig. 2. This is a toy problem that shows the applicability of the algorithm, how it works, and the general trends observed when uncertainty is introduced in the form of a dispersion parameter value.

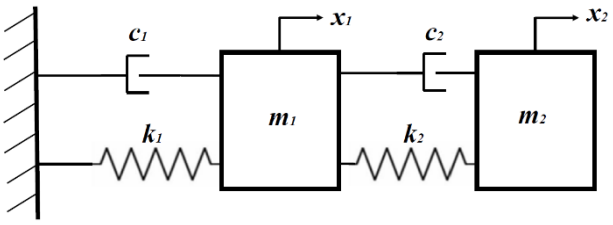


Fig. 2. 2-DoF mass-spring-damper system.

The 2-DoF shear mass-spring-damper system has equal spring stiffnesses $k_1 = k_2 = 1 Nm^{-1}$; equal masses $m_1 = m_2 = 0.05\,kg$ and equal viscous dampers with coefficients of damping

$c_1 = c_2 = 0.1 Nsm^{-1}$. The system is experiencing a free-vibrations response with non-zero initial conditions: $x_1(0) = x_2(0) = 0;\ \dot{x}_1(0) = 1 m/s;\ \dot{x}_2(0) = 0$, where $(\dot{\cdot})$ represents the derivative with respect to time.

The undamped natural frequencies and damping ratios that correspond to the specified properties of the two-degree-of freedom system are given by: $f_1 = 0.4399\ Hz$, $f_2 = 1.1517\ Hz$, $\zeta_1 = 0.1382$ and $\zeta_2 = 0.3618$.

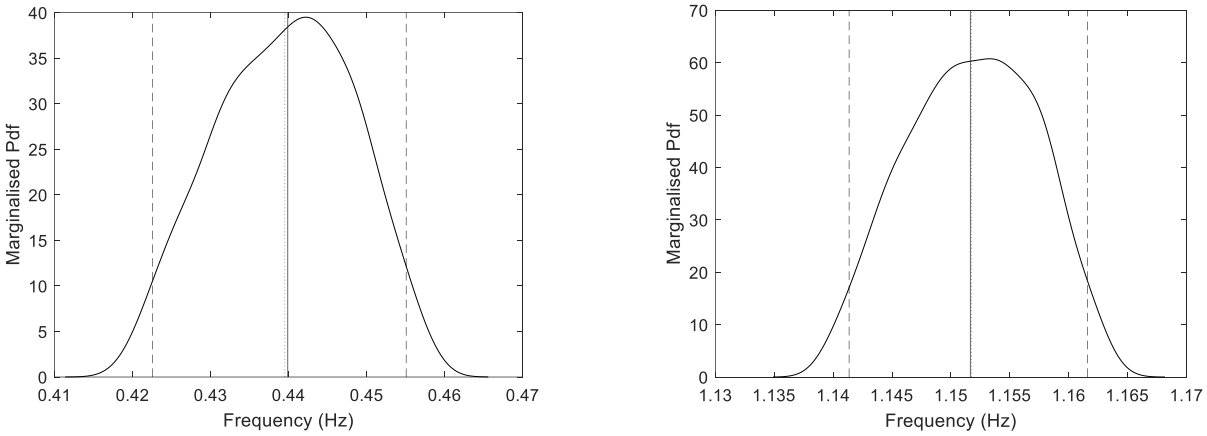
For this example, no measurement noise is considered. Parametric uncertainties are assumed in subsection 4.1.1. Non-parametric uncertainties are assumed in subsection 4.1.2. These uncertainties in the system correspond to uncertain modal parameters which are investigated.

This simple example is used to: (i) directly compare the difference between a parametric and non-parametric uncertainty description; (ii) assess the effect of the dispersion parameter value on the resulting pdfs of the modal parameters.

4.1.1 Pdfs using a parametric probabilistic approach

An example of model parametric uncertainty is shown to illustrate how this affects the uncertainty in the natural frequencies of the 2-DoF mass-spring-damper system.

The parametric probabilistic approach requires the assumption of the pdfs of the uncertain input parameters and the propagation of these uncertainties through the equations of motion. Given the unavailability of information, uniform priors were used for both the stiffness, with $k_1 \sim U(0.95, 1.05)$ [N/m] and the mass $m_2 \sim U(0.0475, 0.0525)$ [kg]. A Monte Carlo simulation with 10,000 (k_1, m_2) pairs of independent samples was used to obtain the pdfs of the natural frequencies of the 2-DoF mass-spring-damper system, which are shown in fig. 3. In this specific example the computational complexity of the model is low, however if a more complex model had been used, the computational cost required to obtain estimates of the modal properties would have been substantially higher.



a) b)

Fig. 3. Probability density functions of modal parameters subject to parametric uncertainty. ····· mean value; ------ 95th percentile; —— deterministic value.

a) first natural frequency. b) second natural frequency.

It is well-known that if different parametric distributions would have been used (e.g., Lognormal), the resulting pdfs of the natural frequencies would have been different. Therefore, this approach would require the identification of the uncertainty parameters, the knowledge of the correct parametric uncertainty affecting the system and the availability of the physics-based model used to propagate the uncertainty in each run of the model to obtain the modal parameters of the system.

4.1.2 Pdfs using the proposed approach

With the proposed approach, no previous knowledge on the uncertainty of specific parameters is needed as described in case (i) of subsection 2.3. The statistics of the modal parameters (natural frequencies, damping ratios and modal shapes) are investigated for two cases: low $(\delta_S = 0.001)$ and high dispersion parameter $(\delta_S = 0.3)$. These two values correspond to a low and a high level of uncertainty, without explicitly identifying the source of this uncertainty. For this numerical case, the state order of the identified system is equal to four.

The free decay of the system is simulated for thirty seconds and the values of $\dot{x}_1$ and $\dot{x}_2$ (that are the velocity of mass 1 and mass 2, respectively) are recorded using a sampling frequency of 10 Hz. The

physics-based model is not needed in any of the following steps. These velocity signals shown in fig. 4 are used to construct the Hankel Matrix $\boldsymbol{H}_{rs}$ described in eq.(1) for $\boldsymbol{H}_{rs}(k-1)$ and $\boldsymbol{H}_{rs}(k)$. The values of r and s in eq.(1) have been set to be equal to 50.

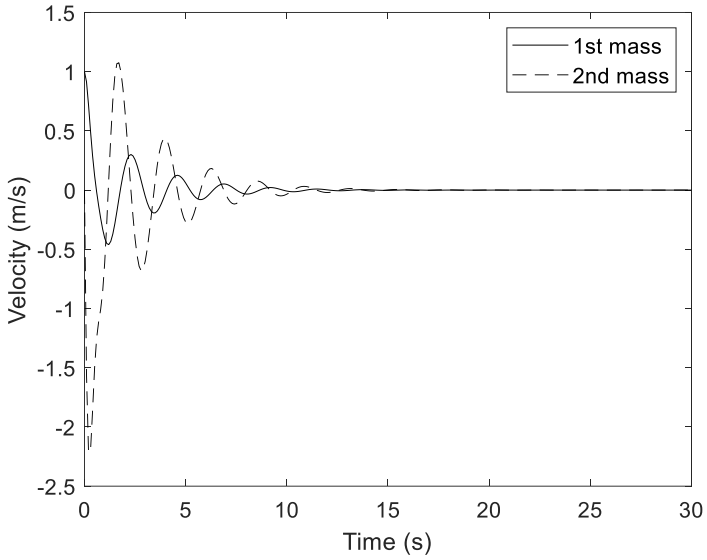


Fig. 4. Velocity of the 1st and 2nd mass subject to specified initial conditions.

The algorithm's steps 3 to 7 in section 3 are run 10,000 times to produce each time a realization of the random matrices $\hat{\boldsymbol{S}}(k-1)$ and $\hat{\boldsymbol{S}}(k)$. For the modal shapes the mean value, the deterministic value and the 95th percentile bounds were calculated. To ensure that convergence was achieved after those 10,000 simulations, the KL-divergence value of the pdfs of each identified parameter with respect to the pdfs for incrementally increasing number of simulations was calculated. It was found that when comparing the distributions obtained from 5,000 vs. 10,000 simulations the KL-divergence value was approximately equal to zero for each identified parameter.

The pdfs in fig. 6 and fig. 7 are obtained using a 200 points kernel smoothing function with the `ksdensity` function in MATLAB [28] on the samples (f,ζ,ϕ) produced by the combined ERA & RMT method.

In this section, the general qualitative trends on these results are stated. In the following section 5, the physical significance and limitations of these results obtained are explained by using new

simulations where the dispersion parameter value and the Hankel Matrix size is varied.

The modal parameter f, ζ pdfs shown in fig. 6 and the modal shapes ϕ in fig. 8 are calculated for a given dispersion parameter value $\delta_S = 0.001$. The dispersion parameter determines the randomization level of the matrices $\hat{\boldsymbol{S}}(k-1)$ and $\hat{\boldsymbol{S}}(k)$, for this case a low level of uncertainty is introduced. As it was expected, for this level of randomization, the deterministic and mean values of the uncertain modal parameters are found to be approximately equal. The pdfs are also found to be fairly symmetrical.

For the case with higher dispersion value $(\delta_S = 0.3)$ which corresponds to higher uncertainties for the identified parameters (fig. 7 and fig. 9), the pdfs of the identified modal parameters are found to be slightly asymmetrical. Negative values in the support of the pdf are also found for the first damping ratio. The mean and deterministic values for the modal shapes are shown to be quite similar with a slightly higher deviation between them for the second mode. The 95% percentile bounds increase as the mode number increases. These effects are also shown in fig. 17 and fig. 18 and are explained in detail in section 5.

As expected, higher levels of uncertainty are seen for the case with higher dispersion value. In this case there is a higher deviation between the mean and deterministic values, and the range of the 95th percentile bounds increases compared to the case with low dispersion value.

It should be noted that slight discrepancies between the mean identified values and the deterministic values of the system appear. These discrepancies are observed in the form of a shift in the mean identified values. This shift is shown in fig. 17 where the mean identified values are affected by the dispersion parameter and the size of the Hankel Matrix. The shapes of pdfs of the identified parameters change significantly, and it has to be noted that although the mean is usually a good descriptor, it may not be the best for cases with high dispersion values.

It has been found, that as expected, when a small dispersion parameter δ_S is considered, results are

close to the deterministic values.

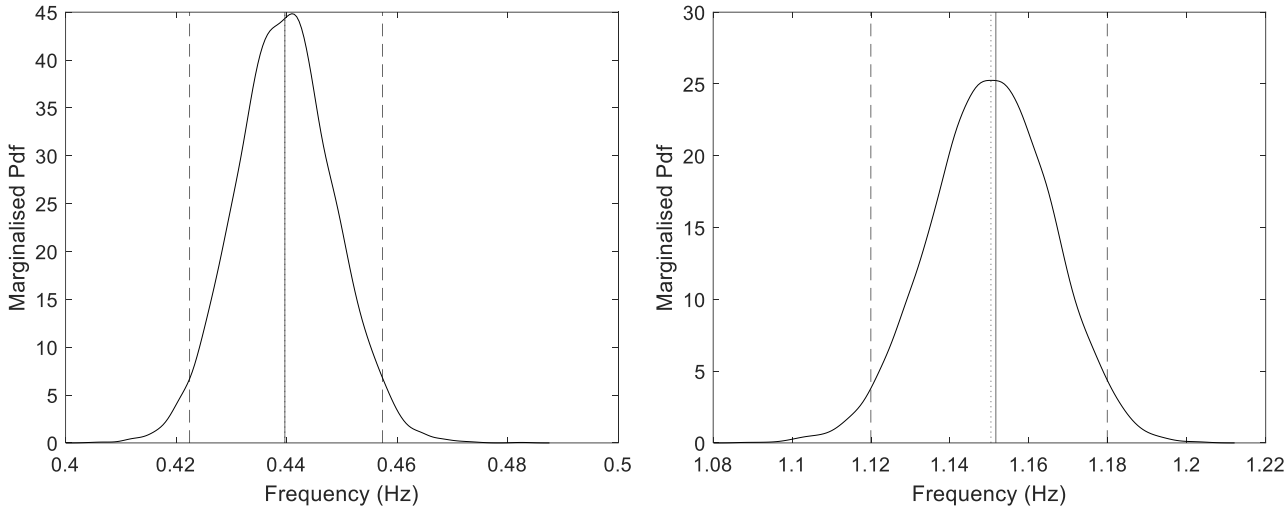
4.1.3 Comparison of results

For the parametric uncertainty example in subsection 4.1.1, the resulting densities of the modal parameters are obtained from the prior knowledge of the uncertain physical parameters k_1 and m_2, by propagation of uncertainties through the physical model. Therefore, for this parametric approach, access to the model is required, and the computational cost involved in the propagation of that parametric uncertainty is dependent on the complexity of that model.

For the non-parametric uncertainty method in subsection 4.1.2, a physical model is not required for calculating the pdfs of the modal parameters. Calculations are based on the knowledge of the dispersion parameter value, and measurements obtained from either a prototype or numerical 'measurements' from the nominal model.

The main difference between the parametric and non-parametric method is how the source of uncertainty is defined [1]. In the parametric method each source of uncertainty needs to be specified, however for the non-parametric approach it is not specified. As shown in example 4.1.2, the overall uncertainty of the system is encompassed in the dispersion parameter.

The dispersion parameter value that yields the same coefficient of variation (2%) of the first natural frequency for the parametric case, is calculated using the maximum likelihood method [1] as described in case (ii) in subsection 2.3.



a) b)

Fig. 5. Probability density functions of modal parameters given equivalent dispersion parameter value $\delta_S = 0.098$. ····· mean value; ------ 95th percentile; ——— deterministic value.

a) first natural frequency. b) second natural frequency.

This equivalent dispersion parameter is $\delta_S = 0.098$. By comparing fig.3 and fig.5, it is possible to observe that the pdf of the first natural frequency is similar to the one obtained using the parametric probabilistic approach. However, the second natural frequency obtained from the proposed approach has a higher coefficient of variation compared to that obtained from the parametric probabilistic approach. This is expected, since the dispersion parameter would inherently account for additional model-parameter uncertainties. The parametric probabilistic approach is not capable of considering those uncertainties. Even though the above is only shown for the parametric case of model parameters (stiffness and mass), the same methodology can be applied for other cases such as the case of known measurement noise. However, the user may believe that substantial modelling errors are still present and would be dominating the overall uncertainty. Therefore, the user may represent the additional level of uncertainty with $\delta_A = 0.202$ to yield an overall dispersion parameter of $\delta_S = 0.3$ as shown in the example in subsection 4.1.2.

As a result, the sources of uncertainty for the parametric and the non-parametric methods are different. Therefore, the overall results obtained from these two methods are not directly comparable.

It should be noted that the method to be chosen is dependent on the information available on the system and the sources of uncertainty.

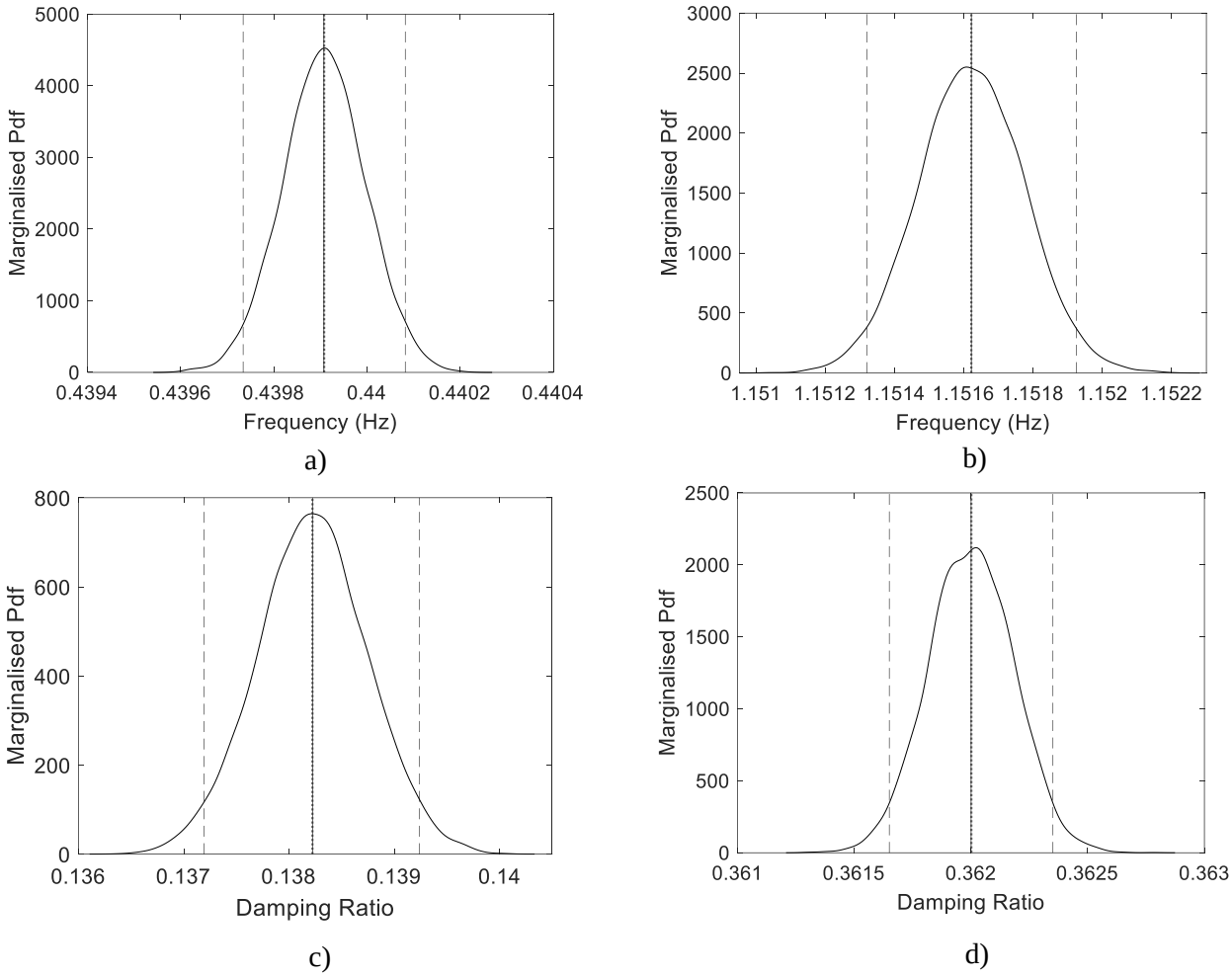


Fig. 6. Probability density functions of modal parameters for a dispersion value: $\delta_s = 0.001$. ····· mean value; ------ 95th percentile; —— deterministic value.

a) first natural frequency. b) second natural frequency. c) first damping ratio. d) second damping ratio.

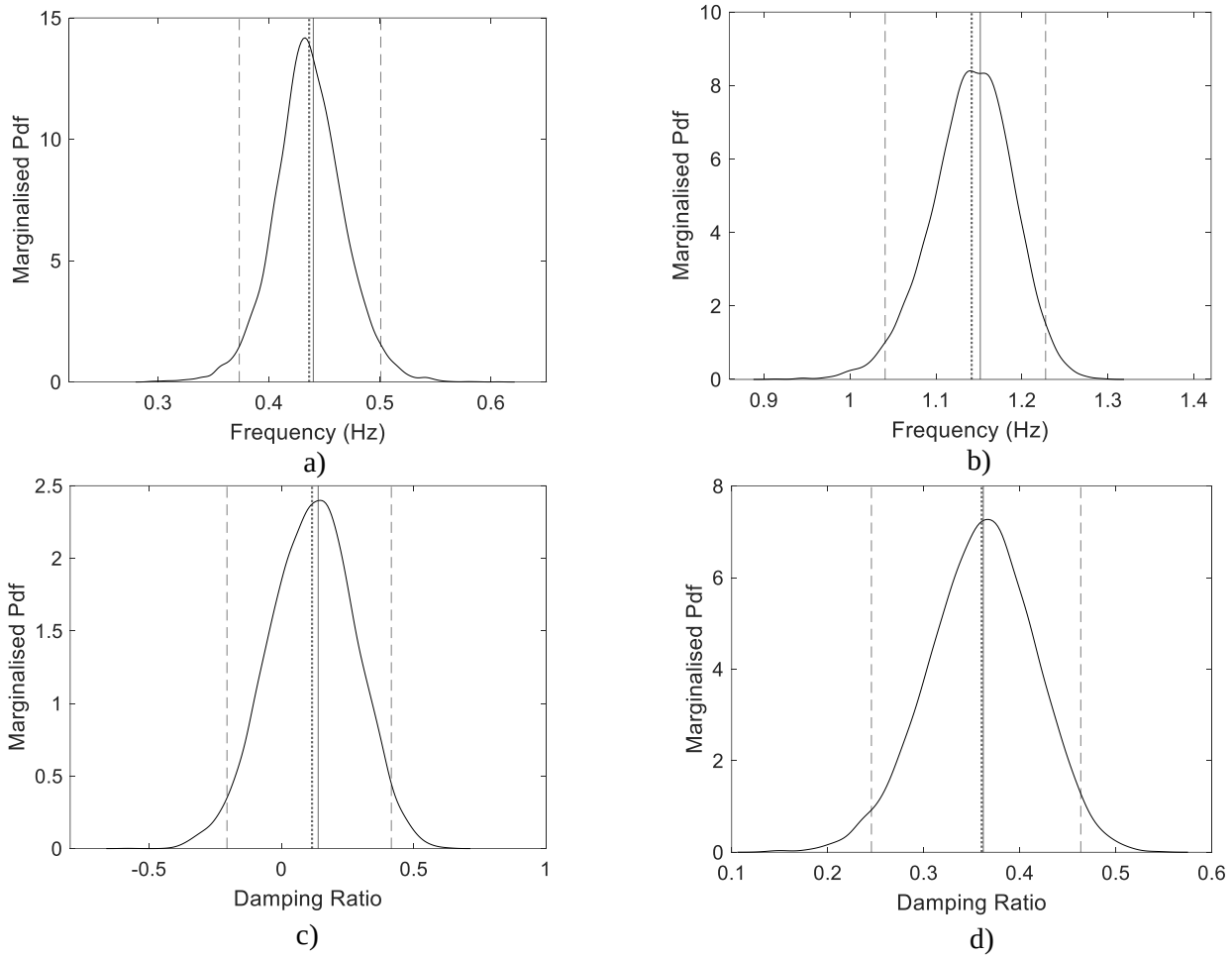
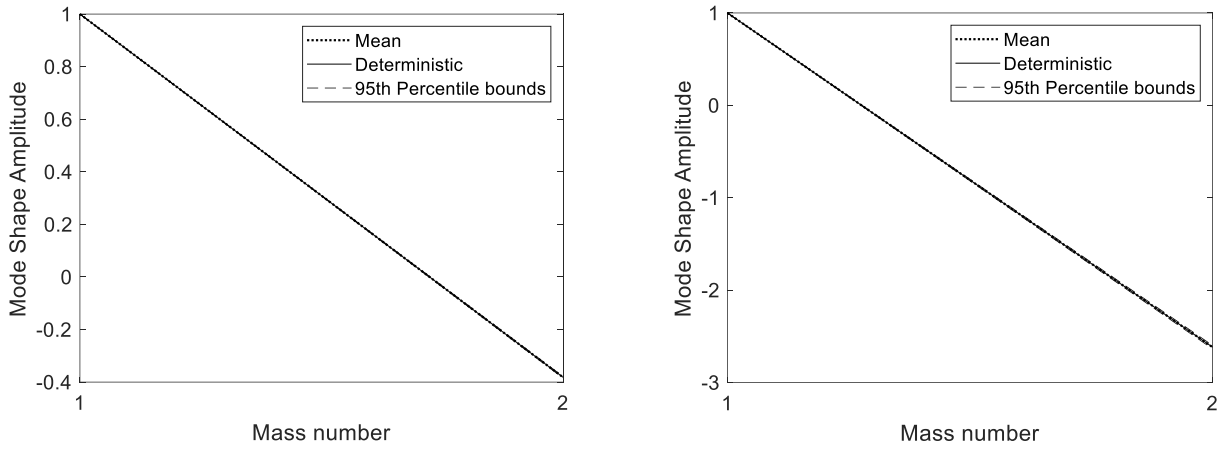


Fig. 7. Probability density functions of modal parameters for a dispersion value: $\delta_S = 0.3$.

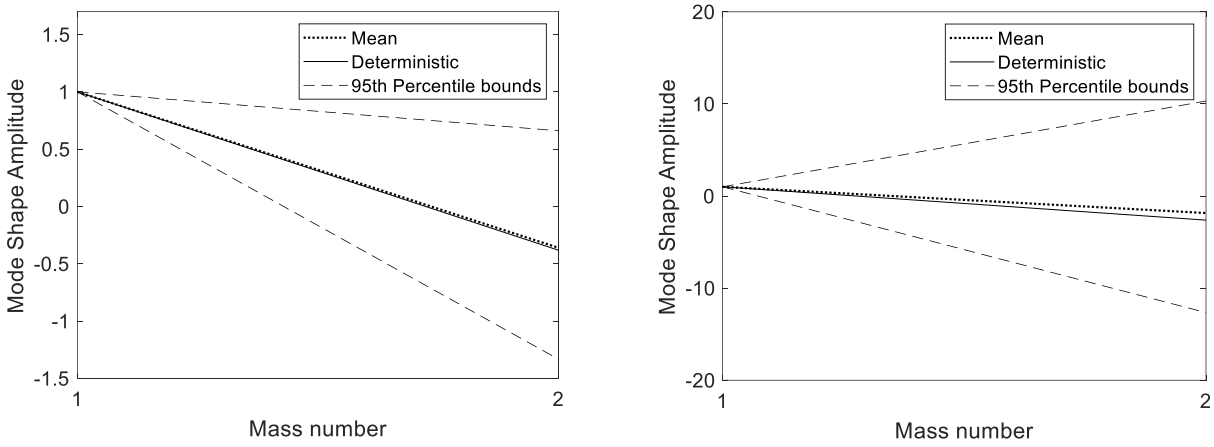
mean value; ------ 95th percentile; —— deterministic value.

a) first natural frequency. b) second natural frequency. c) first damping ratio. d) second damping ratio.



a) b)

Fig. 8. Modal shapes for a dispersion value: $\delta_S = 0.001$. a) first modal shape. b) second modal shape.



a) b)

Fig. 9. Modal shapes for a dispersion value: $\delta_S = 0.3$. a) first modal shape. b) second modal shape.

4.2 Cantilever system

The cantilever system shown on fig. 10 is the continuous system under investigation in this second example.

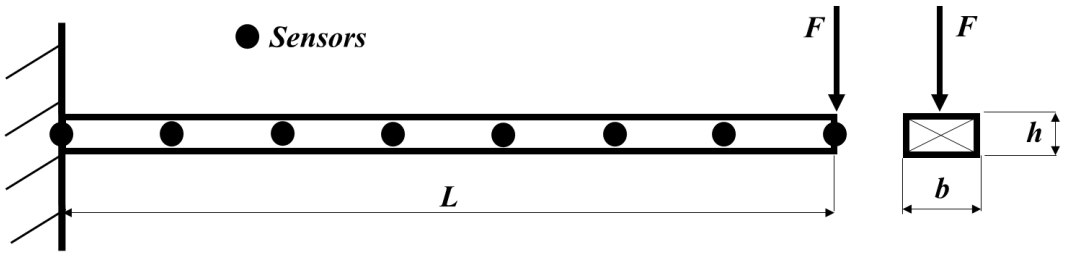


Fig. 10. Schematic representation of the cantilever system.

A force $\boldsymbol{F}$ (triangular pulse of length 4 ms and maximum amplitude of 2000 N) is used to excite the tip of a cantilever beam modelled as a rectangular Euler-Bernoulli beam with uniform density. The following geometric and material properties were used: L (length)=1.5 m; b (base)=0.05 m; h (height)=0.03 m; ρ (density) =7850 kg/m3; E (Young's modulus) =70 GPa. The cantilever has eight velocity sensors that are at equally spaced intervals (fig. 10). The velocity values are recorded over time with a sampling frequency of 1000 Hz in order to capture the frequency content well-excited by the triangular pulse. As a rule-of-thumb, the modes of the system that are below a maximum frequency given by 1.5 times the inverse of the pulse duration are used in the simulation. Therefore, for this system, the number of modes that are used are those with frequencies below 375 Hz.

The response of the continuous system to the force $\boldsymbol{F}$ is numerically simulated using only the first five modes. These free vibration responses are the signals used to construct the Hankel Matrices $\boldsymbol{H}_{\text{rs}}$ described in eq.(1) for $\boldsymbol{H}_{\text{rs}}$ (k-1) and $\boldsymbol{H}_{\text{rs}}$ (k). The values of r and s on the Hankel Matrices have been set to be equal to 100.

The natural frequencies and damping ratios that correspond to the specified properties of the cantilever system are given by: f_1=6.4 Hz, f_2=40.3 Hz, f_3=112.9 Hz, f_4=221.2 Hz, f_5=365.6 Hz, $\zeta_1 = \zeta_2 = \zeta_3 = \zeta_4 = \zeta_5$=0.1.

For this numerical case the order of the identified system is assumed to be equal to eight. This corresponds to the assumption that the measured signals are given by the contribution of a model that is described entirely by four modes. A lower order model, than the one that generated the data, has been chosen to observe if the identified modal parameters distributions are affected by this choice.

The statistics of the modal parameters (natural frequencies, damping ratios and modal shapes) are obtained for two cases: low, $\delta_S = 0.001$, and high dispersion parameter, $\delta_S = 0.075$. The same KL-divergence checks and pdf estimations performed in subsection 4.1 were also carried out for this system.

The pdfs of the modal parameters shown in fig. 11, fig. 12 and fig. 13 are calculated using the method described in section 3 for a dispersion parameter value $\delta_S = 0.001$. With this small dispersion value, it is shown that for the identified modal parameters the deterministic value is approximately equal to the mean value and the pdfs are also fairly symmetrical. The range of the 95% percentile bounds also appears to be small but it is increased as the mode number increases. It was found that these results follow the same general trend described for the previous case in subsection 4.1.

For the case with higher dispersion value $(\delta_S = 0.075)$, which corresponds to a case with higher uncertainties for the identified parameters (fig. 14, fig. 15 and fig. 16), the natural frequencies pdfs are asymmetrical. The first and second damping ratios show negative values in the support of the pdf. As explained in section 3, this phenomenon occurs as a consequence of the randomization process of the method. The appearance of negative values may occur because the realizations of the ensemble may produce damping matrices that are non-proportional. The mean and deterministic values of the natural frequencies show minor discrepancies. However, for the modal shapes larger discrepancies are found between the mean and deterministic values. Assuming that classical damping/proportional damping is observed, the modal shapes obtained are normalised by defining a unit modal amplitude for the 8th node of the system. In some rare occasions, damping may be potentially non-proportional and therefore, considering real mode shapes may not be optimal [29]. The 95% percentile bounds

increase as the mode number increases. The results of this case are also found to follow the general trend found in subsection 4.1, and their significance is also discussed in section 5.

Also, in this numerical application, higher deviation between the mean and deterministic values is seen for the case with higher dispersion value, and the range of the 95th percentile bounds increases compared to the case with low dispersion value. The pdf shapes of the identified parameters change significantly with respect to the case with lower dispersion values.

In this example, the physical model that generated the signal, and the identified model have a different state order. The results show that for the chosen assumption of the state order of the model identified, the obtained results are not affected by the truncation order.

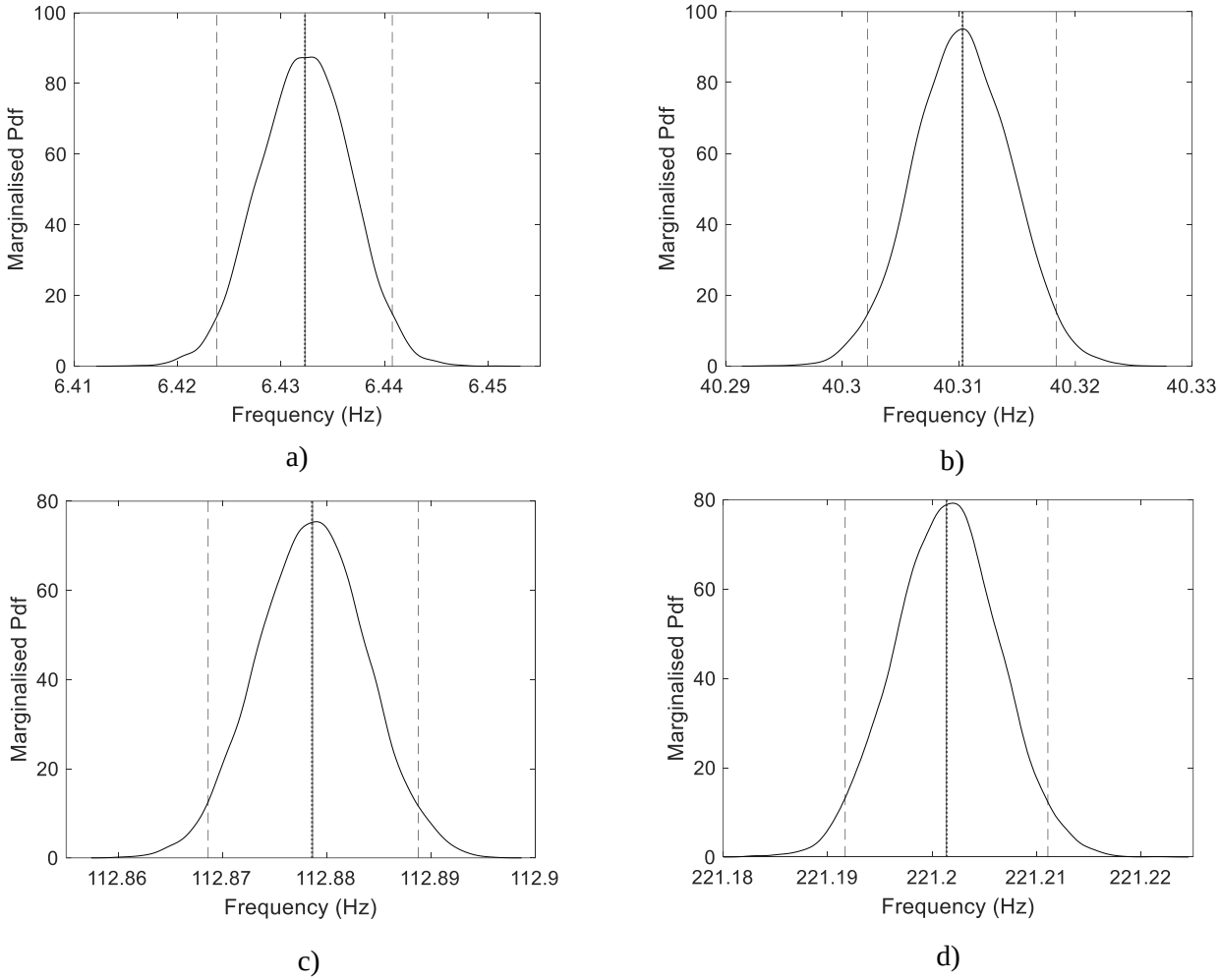


Fig. 11. Probability density functions of natural frequencies for a dispersion value: $\delta_S = 0.001$.

mean value; ------ 95th percentile; ——— deterministic value.

a) first natural frequency. b) second natural frequency. c) third natural frequency. d) fourth natural frequency.

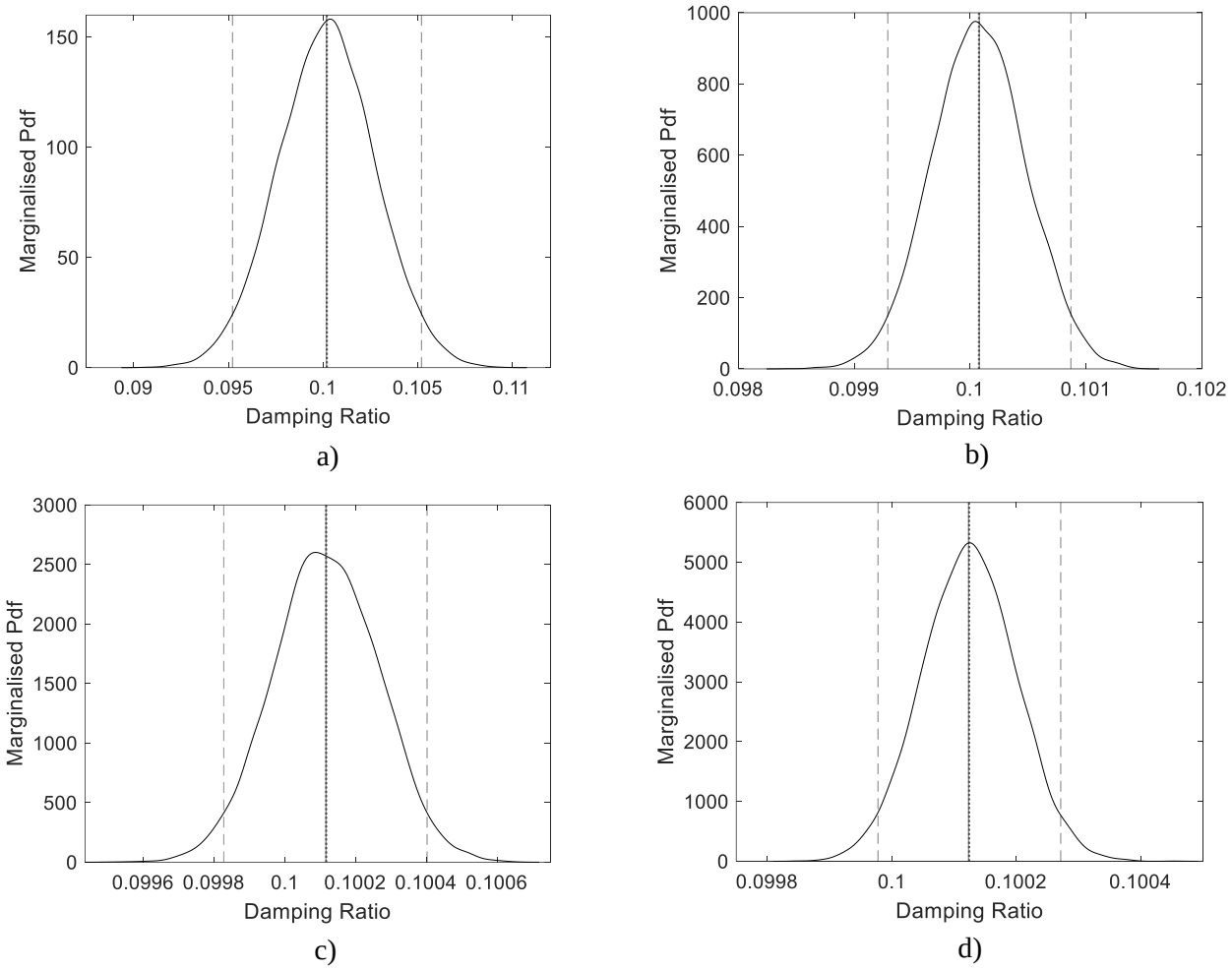


Fig. 12. Probability density functions of damping ratios for a dispersion value: $\delta_s = 0.001$.

mean value; ------ 95th percentile; —— deterministic value.

a) first damping ratio. b) second damping ratio. c) third damping ratio. d) fourth damping ratio.

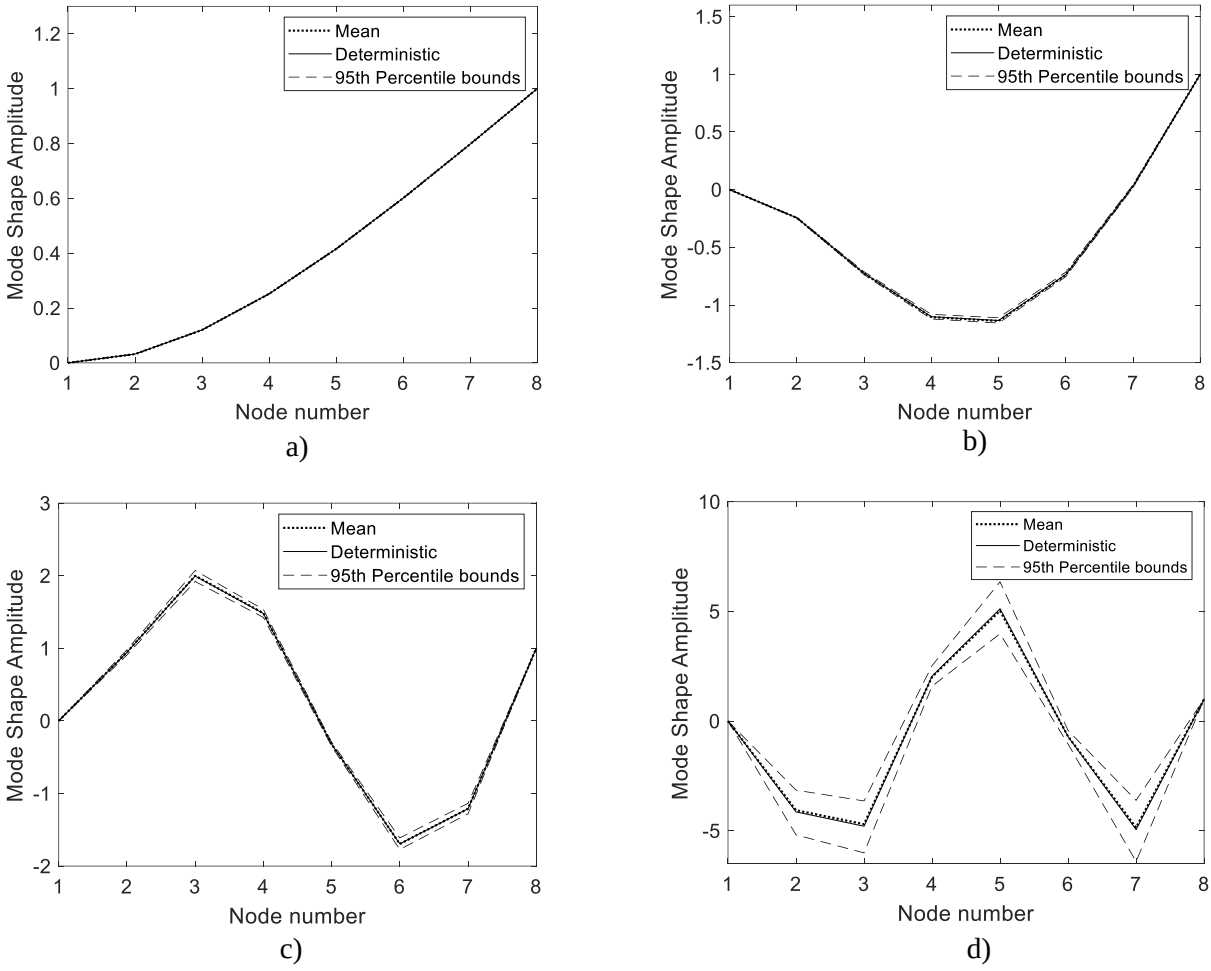


Fig. 13. Mode shape uncertainties for a dispersion value: $\delta_s = 0.001$.

a) mode 1. b) mode 2. c) mode 3. d) mode 4.

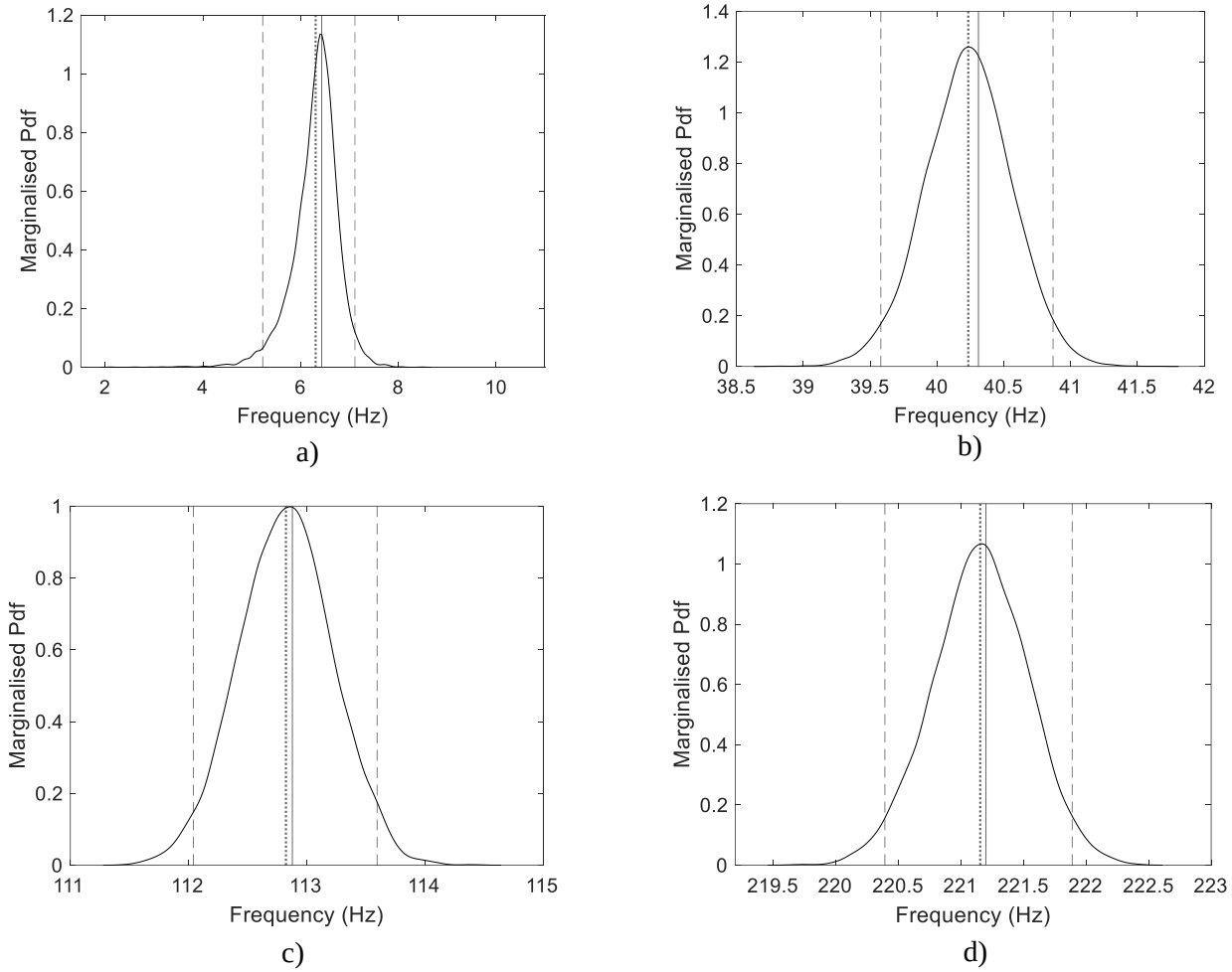


Fig. 14. Probability density functions of natural frequencies for a dispersion value: $\delta_s = 0.075$.

mean value; ------ 95th percentile; —— deterministic value.

a) first natural frequency. b) second natural frequency. c) third natural frequency. d) fourth natural frequency.

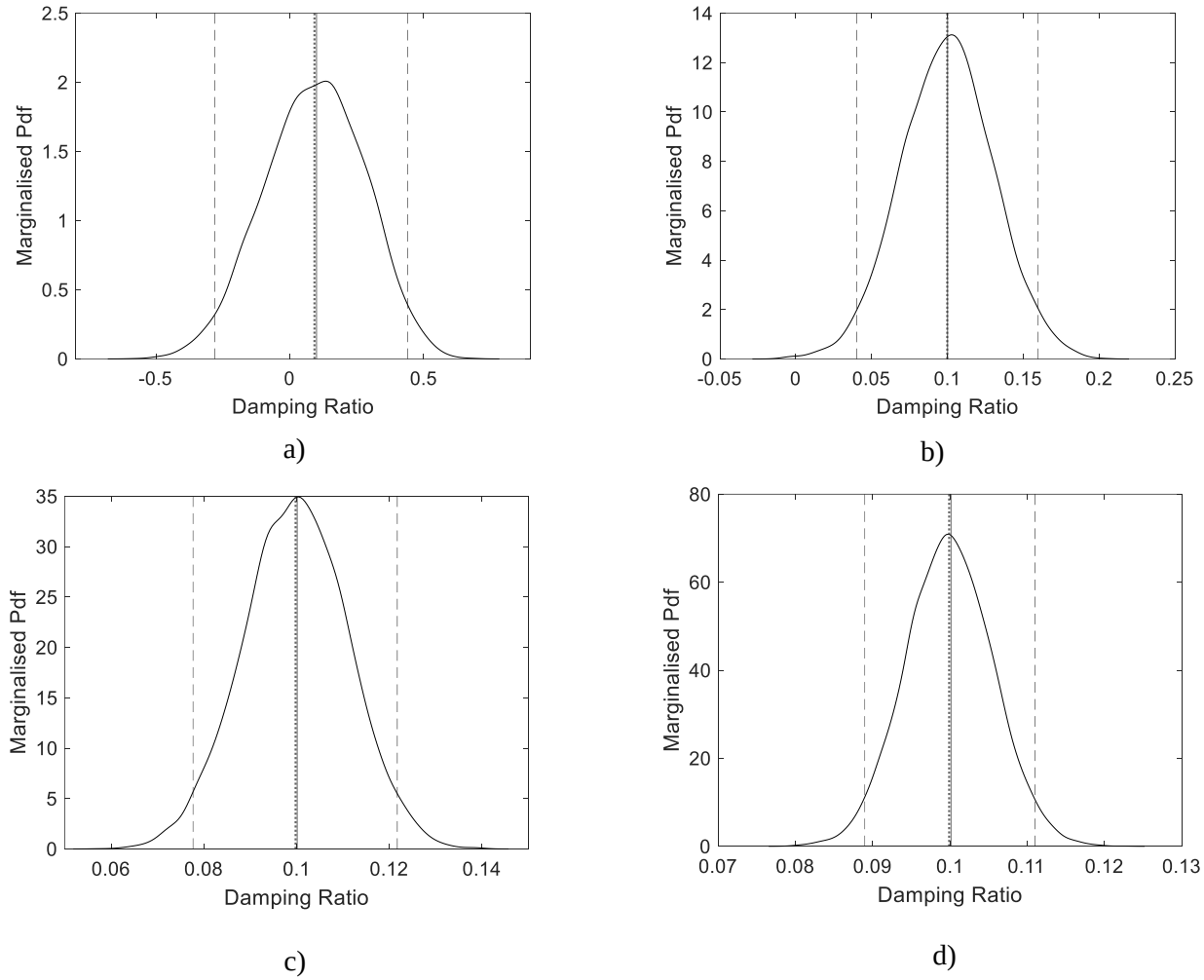


Fig. 15. Probability density functions of damping ratio for a dispersion value $\delta_s = 0.075$.

mean value; ------ 95th percentile; —— deterministic value.

a) damping ratio of the first resonant peak. b) damping ratio of the second resonant peak. c) damping ratio of third resonant peak. d) damping ratio of the fourth resonant peak.

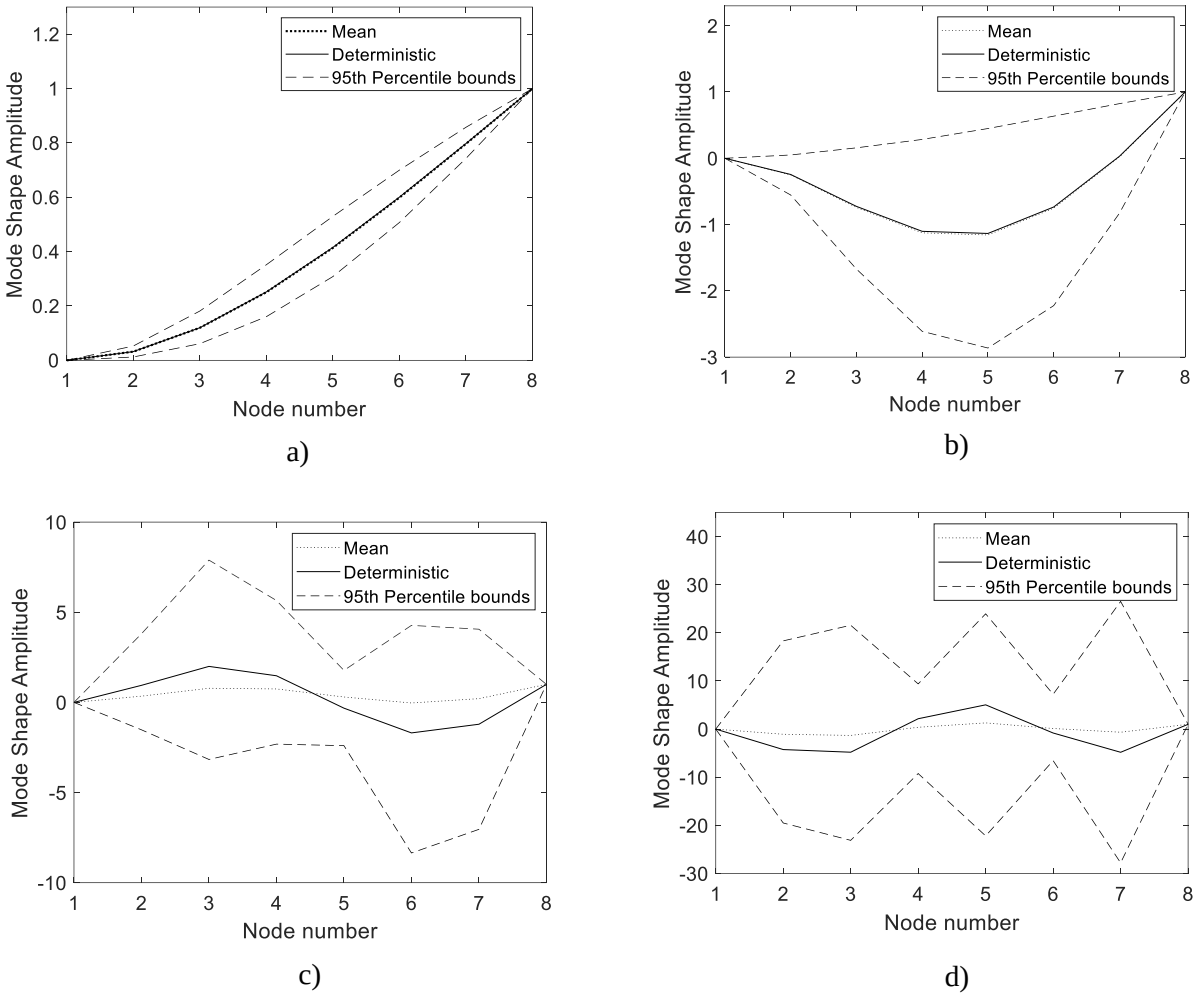


Fig. 16. Mode shape uncertainties for a dispersion value: $\delta_s = 0.075$. a) mode 1. b) mode 2. c) mode 3. d) mode 4.

5 Analysis of the effect of the dispersion value and the Hankel Matrix size on the modal parameters

This section further investigates the two-degree-of-freedom system (subsection 4.1) and the cantilever system (subsection 4.2) models.

The examples were chosen in such a manner that the similarities and common trends between them, when the method described in section 3 is applied, can be observed. For the two-degree-of-freedom system the state order chosen for identification purposes corresponds to the same order of the system that generated the data. However, for the cantilever system the state order of the identified model (eight) is lower than the order of the physical model that generated the signal. The numerical results obtained when the state order is different to the state order that generated the signal in subsection 4.2 show that the identified modal parameters are still of significance and have not been largely affected either by the truncation order or the randomisation process in this example. This is because the loading was such that primarily the first four modes were excited.

For both simulations it is observed that for low levels of dispersion parameter values the variability introduced in the modal parameters is low. The mean identified modal parameters, the identified parameters from the unperturbed Hankel Matrix and the modal parameters used to simulate the velocity signal are approximately equal. However, both examples demonstrate that high levels of dispersion value may lead to a shift between the mean of the occurring distribution of natural frequencies and damping ratios and the underlying nominal values. The uncertainty introduced on the modal parameters is higher than the one introduced by lower levels of dispersion value.

Therefore, further simulations of the two-degree-of-freedom system are used to investigate the physical significance and limitations of the results obtained from the application of the method described in section 3. Contour plots for the two-degree-of-freedom system have been created for varying values of dispersion parameter δ_s and numbers of columns/rows used to construct the Hankel Matrix and are presented in the fig. 17. In fig. 17 the same number of rows and columns was used, and it corresponds to the number value in the horizontal axis of the figure and the data used is taken

from the beginning of the signal. For all contour plots in this paper, the colour scale limits were selected as the maximum and the minimum of displayed data. The mean contour plots (fig. 17) show that there are regions where the mean values of the uncertain parameters identified are the same as the deterministic value of the system. However, if the dispersion value is significant and the number of columns/rows used in the Hankel is low, a bias on the average value of the parameters identified can be observed.

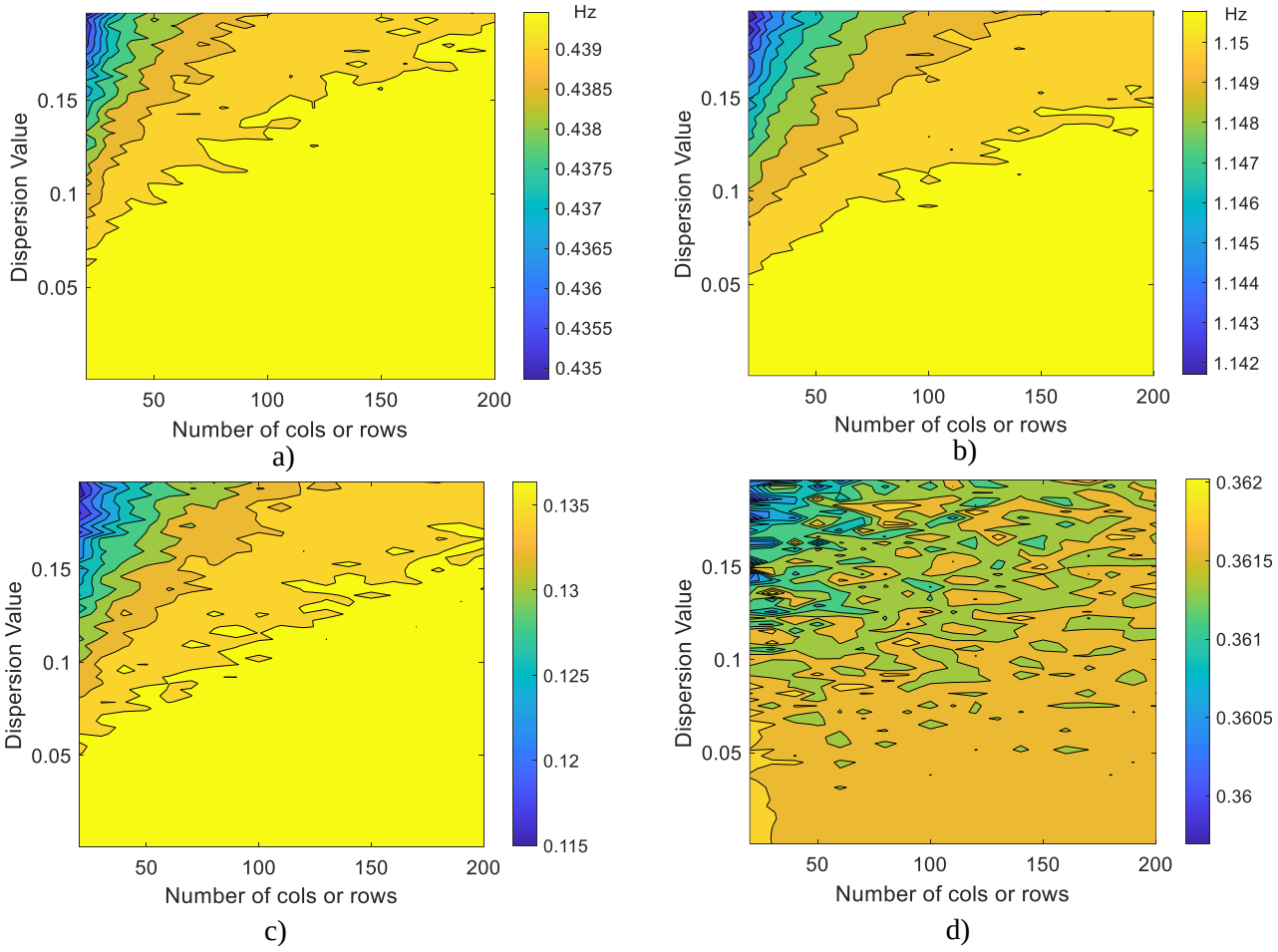


Fig. 17. Contour plots of natural frequencies and damping ratios of their mean values. a) first natural frequency. b) second natural frequency. c): first damping ratio. d) second damping ratio.

For the particular case in fig. 17 the mean values of the identified parameters decrease as the dispersion value increases and as the number of columns/rows decreases. This is because the RMT non-parametric method introduces a non-physical coupling when constructing the random matrices used in steps 3 and 6 of section 3. In this case, the non-physical coupling may be due to the coupling

of the measured signals at the different observed positions. This non-physical coupling was observed before in [21] when a non-parametric approach was used for the modelling of uncertainties of mechanical systems. RMT was applied in [21] to the mass and stiffness matrix, resulting in remote coupling of nonadjacent degrees of freedom as a result of the randomization process.

Therefore, it is found that high values of dispersion parameter affect the underlying physical system to be identified.

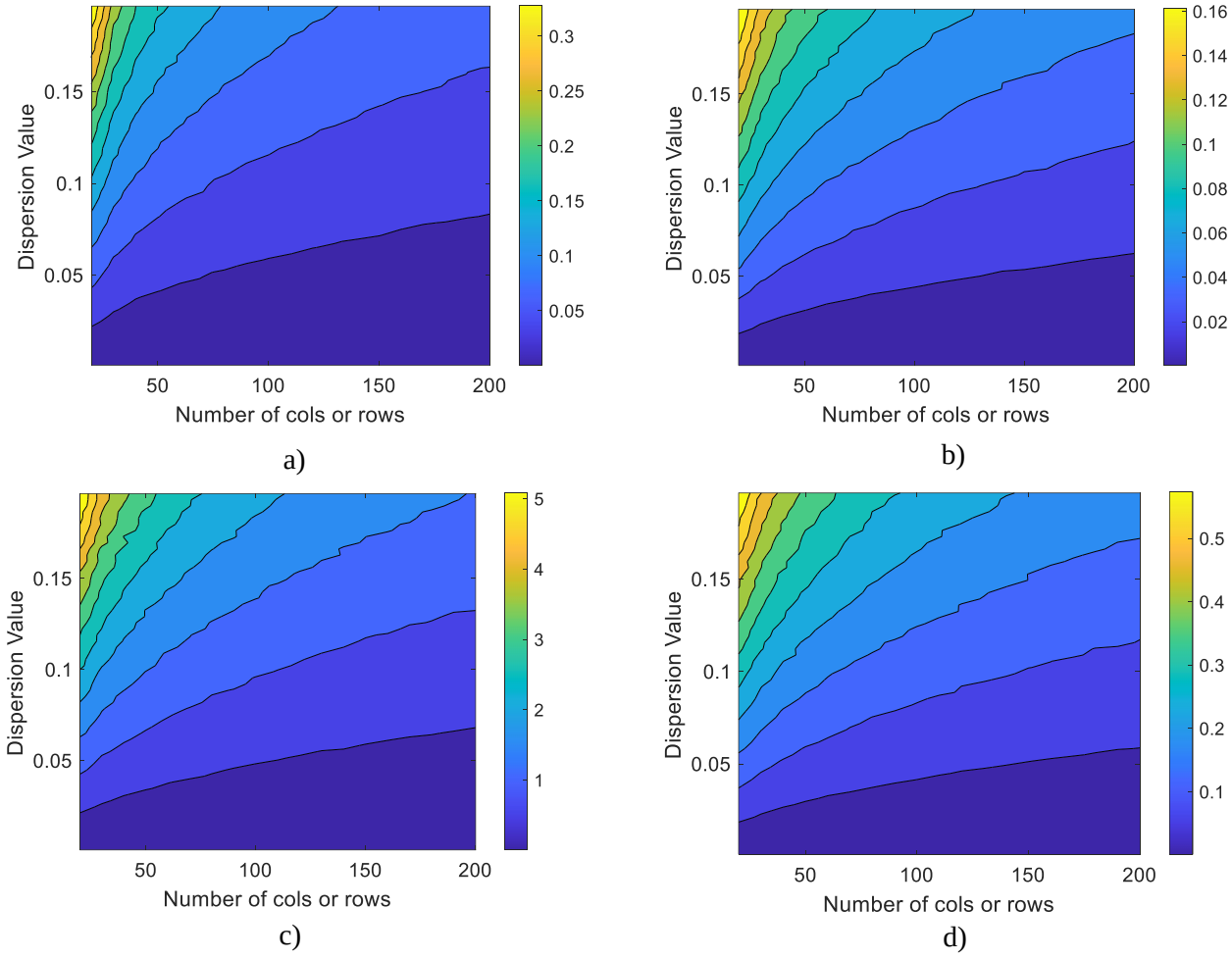


Fig. 18. Contour plots of natural frequencies and damping ratios of their 95% Confidence Interval divided by their Mean values. a) first natural frequency. b) second natural frequency. c) first damping ratio. d) second damping ratio.

It can be noted that in principle, the dispersion parameter should be specified as a function of the Hankel Matrix size and cannot be defined independently. That is, the dispersion parameter for a given Hankel Matrix size should be defined. To explain the shape of the 95% Confidence Interval/Mean contour plots shown in fig. 18, two different effects should be considered. Firstly, as the number of points used to construct the Hankel Matrix increases (i.e., more time steps of the measured signals are used), and for a fixed level of noise, the identified values will tend toward the deterministic value up to a limit where the number of data points will not increase the accuracy of the identified value. Secondly, the effect of keeping the dispersion parameter fixed and increasing the number of columns/rows rs of the Hankel Matrix, is that the uncertainty of the system is reduced, as shown in fig. 18. This double effect is more significant for an identified system with greater dispersion value and a smaller number of columns/rows rs of the Hankel Matrix than for a system with higher number of columns/rows rs of the Hankel Matrix and the same level of dispersion value. This shows that the 95% Confidence Interval/Mean values are less sensitive to changes in the dispersion value when Hankel Matrices of large size are used. This is in fact the reason that the 95% Confidence Interval/Mean contour plots have this specific shape and both effects are observed.

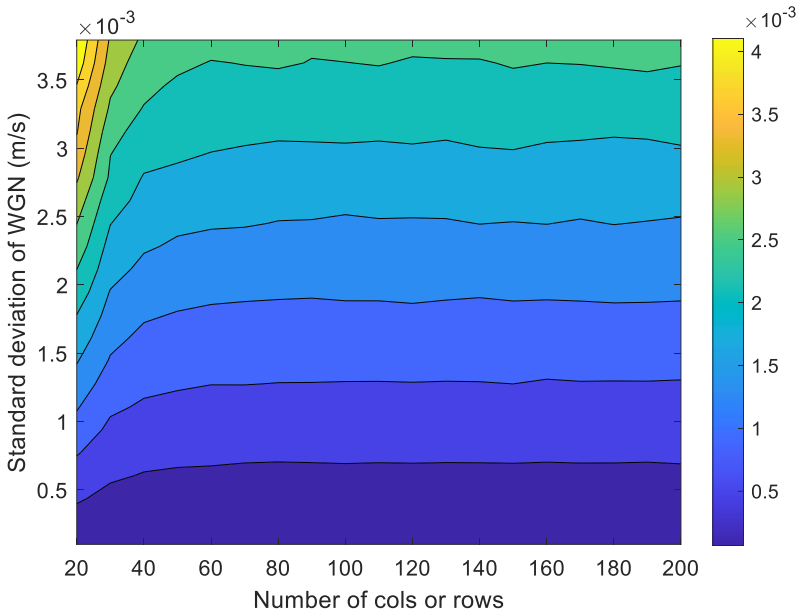


Fig. 19. Contour plot of the 95% Confidence Interval divided by the mean value for the first natural frequency as a function of the number of the number of columns or rows and the standard deviation of the White Gaussian Noise (WGN).

A complementary investigation is performed as shown in fig. 19, considering that the signal used to construct the Hankel Matrix is contaminated by white gaussian noise. A Monte Carlo simulation with 10,000 samples is used to build the contour plot on fig. 19, that has as its y axis varying levels of standard deviation of a white gaussian noise contaminated signal and as the x axis the number of columns/rows used to construct the Hankel Matrix. For a small number of columns/rows the contour lines have a positive gradient. It is also seen that in fig. 19, after the number of rows/columns increase to a specified value that depends on the standard deviation of the white gaussian noise, the contour lines become horizontal. In fig. 19 the first effect described previously on the value of the 95% Confidence Interval/Mean in fig. 18 is also clearly shown.

The uncertainty introduced by the dispersion parameter affects the modal parameters' uncertainty differently. As the dispersion value is increased, the uncertainty for both the natural frequency and the damping ratio estimation is shown to decrease if the mode number is also increased, as shown in fig. 6 in subsection 4.1 and also in subsection 4.2 for both fig. 14 and fig. 15. Therefore, the highest levels of uncertainty are found on the lower mode number natural frequencies and damping ratios as

shown in fig. 6, fig. 7, fig. 11, fig. 12, fig. 14 and fig. 15. For the cases studied, the modal shape uncertainty is observed to increase as the mode number is increased as shown in fig. 9, fig. 13 and fig. 16.

6 Conclusions

This paper presents a procedure for the evaluation of the pdfs of the modal parameters for an ensemble of nominally identical structures when there is only access to a single member of the ensemble, and additional knowledge on the statistical properties of the population expressed through the value of a known dispersion parameter is also available. Given a set of measurements or simulated data, acquired from either a prototype or mathematical model respectively, a discrete time state-space model description is built. The normalized positive definite ensemble [1,11] is used to randomise the matrix calculated by the multiplication of the Hankel Matrix by its own conjugate transpose. ERA is applied to identify the modal parameters (natural frequencies, damping ratios and modal shapes), for each realization of the random matrix. The results of each realization are used to build the pdfs of the modal parameters of the ensemble.

This method is developed to evaluate the modal parameter distributions using the dispersion parameter to account for the uncertain dynamic response across nominally identical structures originated by uncertainties in the manufacturing processes of structural components, modelling errors, boundary conditions and assemblage. Different a priori values of the dispersion parameter are used to evaluate the effect of dispersion value on the resulting pdfs. These distributions are important for the performance assessment of the designed structures, as this would enable the selection of designs that are robust to these uncertainties, avoiding extensive modifications of the manufactured product. The modal parameter distributions may also be applied in anomaly detection cases.

The main advantage of this approach is that it does not require to specify the different sources of uncertainties in advance. The uncertainty of the parameters does not have to be propagated through the equations of motion, becoming this of great advantage when the model of the system is highly

complex or unknown, a circumstance that may introduce modelling errors that affect the overall results. In these circumstances, it may be more convenient to build the ensemble from the measurements of a prototype structure.

Numerical studies were conducted on a two mass-spring-damper and a cantilever system to evaluate the pdfs of the natural frequencies and the damping ratios. The mean and 95th percentile bounds of the modal shapes are also calculated. The study illustrated the performance of the methodology for a wide range of dispersion parameter values. The results obtained from this method generalize well independently of the chosen system (discrete vs. continuous system). The physical consequences of the introduction of high values of dispersion were highlighted. It was found that the use of RMT introduces in the calculation a non-physical behaviour which affects the pdfs of the modal parameters identified. An example of the non-physical behavior introduced by the method is the appearance of negative values in the support of the pdf of the first damping ratio in the mass-spring-damper system. It was also shown that the state order of the identified system does not affect the identified modal parameters in the cantilever system.

Contour plots were produced to show the effect in the identified parameters' uncertainties of the number of rows/columns used in the Hankel Matrices and the dispersion parameter values. When the dispersion value is significant, and the number of columns/rows used in the Hankel is relatively low, the contour plots of the identified mean values show a systematic shift. This systematic bias increases when either, the dispersion value is increased, and/or the number of columns/rows is decreased. Two different effects can be observed on the 95% Confidence Interval/Mean contour plots. The first one is that increasing number of points used in the Hankel Matrix, the identified values will tend toward the deterministic value, and the second is that when fixing the dispersion parameter value and using larger values of columns/rows, the uncertainty of the system is being reduced. Therefore, the uncertainty introduced in the modal parameters is both dependent on the size of the Hankel Matrix and on the dispersion parameter value.

It is also shown that the uncertainty associated with the natural frequencies and damping ratios is

decreased as the mode number increases and/or the dispersion parameter value is increased.

For the case of parametric uncertainty in subsection 4.1.1, the pdfs of the modal parameters are calculated propagating the prior knowledge of those modal parameters through a physical model. For the non-parametric case in subsection 4.1.2, only measurements (numerically obtained) from the structure and the value of a dispersion parameter are available. The main difference between the methods is that they depend on how the source of uncertainty is defined.

The prior estimation of the dispersion parameter δ_S is illustrated for three cases. The first case assumes that no prior data are available, then, the dispersion parameter δ_S is used as variable in a sensitivity analysis to assess the stochastic response of the model. The second and third cases assume that prior data are available. Given the observed data, the value of the dispersion parameter value can be calculated. However, for the third case there is also an additional level of uncertainty present that is represented by and additional term that may be added to the dispersion parameter value calculated from the observed data.

Future work will focus on the estimation of the dispersion parameter value for cases when experimental measurements of the ensemble are available. Another potential direction of interest is the use of different SI methods to determine the compatibility of different experimental conditions of the system tested with the approach proposed in this paper.

Acknowledgments

Felipe Igea and Alice Cicirello thanks the EPSRC and Schlumberger for an Industrial Case postgraduate scholarship (Grant ref. EP/T517653/1). Alice Cicirello gratefully acknowledges the EPSRC for the Overseas Travel Grants (Grant Ref. EP/R008949/1).

References

[1] Soize, C., 2017, *Uncertainty Quantification -An Accelerated Course with Advanced Applications in Computational Engineering*, Springer International Publishing, Cham.

[2] Wright, M., and Weave, R., 2010, *New Directions in Linear Acoustics and Vibration: Quantum Chaos, Random Matrix Theory, and Complexity*, Cambridge University Press.

[3] Kompella, M. S., and Bernhard, R. J., 1993, "Measurement of the Statistical Variation of Structural-Acoustic Characteristics of Automotive Vehicles," *SAE Technical Papers*, SAE International.

[4] Durand, J.-F., Soize, C., and Gagliardini, L., 2008, "Structural-Acoustic Modeling of Automotive Vehicles in Presence of Uncertainties and Experimental Identification and Validation," The Journal of the Acoustical Society of America, **124**(3), pp. 1513–1525.

[5] Zhang, J., Shen, G., Du, Y., and Hu, P., 2013, "Modal Analysis of a Lightweight Engine Hood Design Considering Stamping Effects," *Applied Mechanics and Materials*, Trans Tech Publications Ltd, pp. 364–369.

[6] Reich, S., and Cotter, C., 2015, "Probabilistic Forecasting and Bayesian Data Assimilation," Probabilistic Forecasting and Bayesian Data Assimilation, pp. 1–297.

[7] Langley, R. S., 2000, "Unified Approach to Probabilistic and Possibilistic Analysis of Uncertain Systems," Journal of Engineering Mechanics, **126**(11), pp. 1163–1172.

[8] Soize, C., 2013, "Stochastic Modeling of Uncertainties in Computational Structural Dynamics - Recent Theoretical Advances," Journal of Sound and Vibration, **332**(10), pp. 2379–2395.

[9] Simoen, E., de Roeck, G., and Lombaert, G., 2015, "Dealing with Uncertainty in Model Updating for Damage Assessment: A Review," Mechanical Systems and Signal Processing, **56**, pp. 123–149.

[10] Faes, M., and Moens, D., 2020, “Recent Trends in the Modeling and Quantification of Non-Probabilistic Uncertainty,” Archives of Computational Methods in Engineering, **27**(3), pp. 633–671.

[11] Soize, C., 2005, “A Comprehensive Overview of a Non-Parametric Probabilistic Approach of Model Uncertainties for Predictive Models in Structural Dynamics,” *Journal of Sound and Vibration*, Academic Press, pp. 623–652.

[12] Mehta, M., 2004, *Random Matrices*.

[13] Arnst, M., and Soize, C., 2019, “Identification and Sampling of Bayesian Posteriors of High-Dimensional Symmetric Positive-Definite Matrices for Data-Driven Updating of Computational Models,” Computer Methods in Applied Mechanics and Engineering, **352**, pp. 300–323.

[14] Batou, A., Soize, C., and Audebert, S., 2015, “Model Identification in Computational Stochastic Dynamics Using Experimental Modal Data,” Mechanical Systems and Signal Processing, **50–51**, pp. 307–322.

[15] Ezvan, O., Batou, A., Soize, C., and Gagliardini, L., 2017, “Multilevel Model Reduction for Uncertainty Quantification in Computational Structural Dynamics,” Computational Mechanics, **59**(2), pp. 219–246.

[16] Vishwanathan, A., and Vio, G. A., 2019, “Numerical and Experimental Assessment of Random Matrix Theory to Quantify Uncertainty in Aerospace Structures,” Mechanical Systems and Signal Processing, **118**, pp. 408–422.

[17] Langley, R. S., and Cotoni, V., 2008, “Response Variance Prediction for Uncertain Vibro-Acoustic Systems Using a Hybrid Deterministic-Statistical Method,” The Journal of the Acoustical Society of America, **122**(6), p. 3445.

[18] Langley, R. S., Cicirello, A., and Deckers, E., 2018, "The Effect of Generalised Force Correlations on the Response Statistics of a Harmonically Driven Random System," Journal of Sound and Vibration, **413**, pp. 456–466.

[19] Vishwajeet, K., Singla, P., and Majji, M., 2017, "Random Matrix-Based Approach for Uncertainty Analysis of the Eigensystem Realization Algorithm," Journal of Guidance, Control, and Dynamics, **40**(8), pp. 1877–1891.

[20] Soize, C., 2005, "Random Matrix Theory for Modeling Uncertainties in Computational Mechanics," Computer Methods in Applied Mechanics and Engineering, **194**(12–16), pp. 1333–1366.

[21] Legault, J., Langley, R. S., and Woodhouse, J., 2012, "Physical Consequences of a Nonparametric Uncertainty Model in Structural Dynamics," Journal of Sound and Vibration, **331**(25), pp. 5469–5487.

[22] Juang, J. N., and Pappa, R. S., 1985, "An Eigensystem Realization Algorithm for Modal Parameter Identification and Model Reduction," Journal of Guidance, Control, and Dynamics, **8**(5), pp. 620–627.

[23] Brunton, S. L., and Kutz, J. N., 2019, *Data-Driven Science and Engineering*, Cambridge University Press.

[24] Caicedo, J. M., Dyke, S. J., and Johnson, E. A., 2004, "Natural Excitation Technique and Eigensystem Realization Algorithm for Phase I of the IASC-ASCE Benchmark Problem: Simulated Data," Journal of Engineering Mechanics, **130**(1), pp. 49–60.

[25] Reynders, E., 2012, "System Identification Methods for (Operational) Modal Analysis: Review and Comparison," Archives of Computational Methods in Engineering 2012 19:1, **19**(1), pp. 51–124.

[26] Silverman, B., 1986, *Density Estimation for Statistics and Data Analysis*.

[27] Wu, C., Liu, H., Qin, X., and Wang, J., 2017, “Stabilization Diagrams to Distinguish Physical Modes and Spurious Modes for Structural Parameter Identification,” Journal of Vibroengineering, **19**(4), pp. 2777–2794.

[28] 2020, “MATLAB.”

[29] Sinha, S., and Allemang, R. J., 2006, “Techniques for Real Normalization of Complex Modal Vectors for Updating and Correlation with FEM Models,” The Shock and Vibration Digest, **38**(4), pp. 373–374.

Appendix – Eigensystem Realization Algorithm

The time evolution of a linear, time invariant, dynamic system can be described using a discrete-time state-space representation:

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k \qquad \text{(A.1)}$$

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{D}\boldsymbol{u}_k \qquad \text{(A.2)}$$

Where $\boldsymbol{x}_k \in \mathbb{R}^n$, $\boldsymbol{y}_k \in \mathbb{R}^p$ and $\boldsymbol{u}_k \in \mathbb{R}^m$ are respectively the state vector, collected output and the value of the control input which remains constant between times t_k and t_{k+1}. Matrices $\boldsymbol{A} \in \mathbb{R}^{nxn}$, $\boldsymbol{B} \in \mathbb{R}^{nxm}$, $\boldsymbol{C} \in \mathbb{R}^{pxn}$, and $\boldsymbol{D} \in \mathbb{R}^{nxm}$ are the system, input, output and feedthrough matrices respectively.

The Eigensystem Realization Algorithm (ERA) [22] is a system identification method that is used to estimate the modal parameters of a system by evaluating the so-called transformed matrices to the modal-space. To evaluate the matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$, ERA uses an output vector $\boldsymbol{y}_k$ that contains the measurements read in p sensors at the different times t_k for a system excited by an impulse at time zero and assuming zero initial conditions. ERA obtains a transformation of eq.(A.1) and eq.(A.2) that has as its matrices $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{C}}$ which correspond to the state vector $\tilde{\boldsymbol{x}}_k$. The eigenvalues of the matrix $\tilde{\boldsymbol{A}}$ and the sensor-based modal shapes in eq.(A.9) are transformation invariant and can therefore be used to obtain the eigenvalues and the sensor-based modal shapes of the original system that generated the data. Subsequently, the eigenvalue decomposition of the matrix $\tilde{\boldsymbol{A}}$ (eq.(A.3)) is used to determine the discrete time eigenvalues λ_i and the corresponding set of eigenvectors $\boldsymbol{v}_i$ from [22]:

$$\tilde{\boldsymbol{A}}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \qquad \text{(A.3)}$$

And the following matrix $\boldsymbol{\Lambda}$ and vector $\boldsymbol{V}$ can then be obtained [22]:

$$\boldsymbol{\Lambda} = diag\left(\lambda_1, \quad \lambda_2, \quad ... \quad \lambda_n\right) \quad \text{(A.4)}$$

$$\boldsymbol{V} = \left[\boldsymbol{v}_1, \quad \boldsymbol{v}_2, \quad ... \quad \boldsymbol{v}_n\right] \quad \text{(A.5)}$$

The eigenvalues of the matrix $\tilde{\boldsymbol{A}}$ can then be used to calculate the eigenvalues $\lambda_{c(i)}$, natural frequencies and damping factors of the equivalent continuous in time system using eq.(A.6), eq.(13) and eq.(15) respectively [22].

$$\lambda_{c(i)} = \frac{ln(\lambda_i)}{\Delta t} \quad \text{(A.6)}$$

where Δt is the sampling period and $\Re$ denotes the real part. The sensor-based modal shapes are then obtained using eq.(14) [22].

The steps required to obtain the $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{C}}$ matrices are:

Step 1: Build the Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$ of size $r \times p$ by s is built that contains time series data from measurement data using eq.(1) [23].

Step 2: Build the matrix $\boldsymbol{H}_{rs}(k)$, also called the shifted Hankel Matrix which is one-time step into the future of the matrix $\boldsymbol{H}_{rs}(k-1)$.

$$\boldsymbol{H}_{rs}(k) = \begin{bmatrix} \boldsymbol{y}_{k+1} & \boldsymbol{y}_{k+2} & \cdots & \boldsymbol{y}_{k+s} \\ \boldsymbol{y}_{k+2} & \boldsymbol{y}_{k+3} & \cdots & \boldsymbol{y}_{k+s+1} \\ . & . & & . \\ . & . & & . \\ \boldsymbol{y}_{r+k} & \boldsymbol{y}_{r+k+1} & \cdots & \boldsymbol{y}_{r+k+s-1} \end{bmatrix} \quad \text{(A.7)}$$

Step 3: Obtain the approximate expression of the $\boldsymbol{H}_{rs}(k-1)$. By using the singular value decomposition, the Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$ can be expressed as [23]:

$$\boldsymbol{H}_{rs}(k-1) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{*T} \quad \text{(A.8)}$$

In eq.(A.8), ^{*T} denotes complex conjugate transpose, matrices $\boldsymbol{U} \in \mathbb{C}^{rxr}$ and $\boldsymbol{V} \in \mathbb{C}^{sxs}$ are unitary, have orthonormal columns, and matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{rxs}$ has at most s nonzero elements in its diagonal. The exact expression of eq.(A.8) can also be written as [23]:

$$\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{*T} = \begin{bmatrix} \tilde{\boldsymbol{U}} & \boldsymbol{U}_t \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & 0 \\ 0 & \boldsymbol{\Sigma}_t \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{V}}^{*T} \\ \boldsymbol{V}_t^{*T} \end{bmatrix} \quad \text{(A.9)}$$

Where $\tilde{\boldsymbol{U}}$, $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{V}}^{*T}$ denote truncated matrices containing the dominant singular values and vectors, and where $\boldsymbol{U}_t$, $\boldsymbol{\Sigma}_t$ and $\boldsymbol{V}_t^{*T}$ contain the non-dominant singular values and vectors that are discarded in the approximation [23]:

$$\boldsymbol{H}_{rs}(k-1) \approx \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^{*T} \quad \text{(A.10)}$$

If the Hankel Matrix $\boldsymbol{H}_{rs}(k-1)$ does not have full rank, then $\tilde{\boldsymbol{\Sigma}}$ may contain some zero singular values and the eq.(A.10) will be exact. However, if $\tilde{\boldsymbol{\Sigma}}$ contains a number of non-zero singular values smaller than the rank of the Hankel Matrix, then the expression of eq.(A.10) is only approximate.

Step 4: In this step the realization matrices of the system are estimated using the $\tilde{\boldsymbol{U}}$, $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{V}}^{*T}$ matrices as [23]:

$$\tilde{\boldsymbol{A}} \approx \tilde{\boldsymbol{\Sigma}}^{-1/2}\tilde{\boldsymbol{U}}^{*T}\boldsymbol{H}_{rs}(k)\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Sigma}}^{-1/2} \quad \text{(A.11)}$$

$$\tilde{\boldsymbol{B}} \approx \tilde{\boldsymbol{\Sigma}}^{1/2}\tilde{\boldsymbol{V}}^{*T}\begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{0}_m \\ \boldsymbol{0}_m & \boldsymbol{0}_m \end{bmatrix} \quad \text{(A.12)}$$

$$\tilde{\boldsymbol{C}} \approx \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0}_p \\ \boldsymbol{0}_p & \boldsymbol{0}_p \end{bmatrix}\tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}^{1/2} \quad \text{(A.13)}$$

Where $\boldsymbol{0}_p$ and $\boldsymbol{I}_p$ are respectively matrices zero and identity of size $p \times p$.

When the matrices $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{C}}$ are used in a system of equations similar to eq.(A.1) and eq.(A.2) become the discrete time modal-space representation of the dynamic system [23].

Using ERA, and the matrices $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{C}}$, deterministic values of the modal parameters of a system

are obtained for a given set of measurements. However, these measurements might be affected by different sources of uncertainties such as sensor noise. The existence of those uncertainties is considered in ERA by averaging out the noise of the signals using several measurements to obtain the deterministic values of the modal parameters.

3.2 Conclusions

At early design stages (before a prototype is being built), when a physics-based model is developed, the most common methods used to describe the uncertainty on the model parameters, are to employ a probabilistic or non-probabilistic approach. However, when only limited information is available at the design stage, specifying the uncertainties' sources and the description of the parameters model's uncertainty, is a difficult task.

In paper I, a new non-parametric probabilistic approach applicable to the ensemble of nominally identical structures that may arise during the modelling processes, due to the variabilities introduced by modelling errors, boundary conditions, and the assembly process is developed. This non-parametric approach can be used when only data from a single structure (either numerical model or prototype) is available, and it uses as inputs the mentioned data, and a so-called dispersion parameter. During the design process, the approach can be used to produce an appraisal of the dynamic performance of the structure through the quantification of the uncertainties of the modal parameters. This is important, as the modal parameters of nominally identical structures may exhibit important differences among them. These differences are found as a consequence of the variability of the boundary conditions, assemblage, and manufacturing processes among the members of the ensemble. The approach can be used as a tool to avoid major modifications of the structure after it is completed, as allows the selection of designs robust to the uncertainties of the distribution of the modal parameters across the ensemble of nominally identical structures.

In the presence of uncertainty, the challenge of locating the positions in the structure where observations should be taken arises. The information obtained varies based on where the observations are taken. This challenge is investigated in the next chapter.

Statement of Authorship for joint/multi-authored papers for PGR thesis

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	On the Combination of Random Matrix Theory with Measurements on a Single Structure
Publication Status	☒Published □ Accepted for Publication □Submitted for Publication □Unpublished and unsubmitted work written in a manuscript style
Publication Details	Igea F., Chatzis M.N.., Cicirello A., On the Combination of Random Matrix Theory with Measurements On a Single Structure. ASME J. Risk Uncertainty Part B, 2022

Student Confirmation

Student Name:	Felipe Igea		
Contribution to the Paper	Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing.		
Signature	Felipe Igea	Date	16/09/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title:	Dr Alice Cicirello		
Supervisor comments	The student made a substantial contribution to the publication, and the description above is accurate		
Signature		Date	18/09/23

Chapter 4 - Limited information in operating conditions – sensor placement

Statistical experimental design is a methodology for the optimization of the collection of data. It may be applied before the experiment is run in cases where uncertainty is present in the experimental conditions, and also in cases where the values of the parameters of interest are uncertain. This methodology defines which parameters will be monitored, and the recurrence, schedule, and number of observations used to update the knowledge on each parameter. As limited data gathering resources are available, their effective allocation becomes crucial.

In most cases, the estimation of model parameters is the main purpose of the experimental design. As the experiment is designed before the model parameters are precisely known, the designer needs to estimate their values, so that the experimental design process can be performed.

If the model parameters are poorly estimated, inferior designs may be produced. The assessment of the initial parameters, such that a robust experimental design can be performed, is the object of numerous research works [50,70,71]. In these studies, the so-called prior distributions for the model parameters are estimated by using prior available information in the specific area of expertise. The choice of a prior distribution is a fundamental part of the Bayesian design and Bayesian analysis. The robustness of the experimental design to the selection of a prior pdf should be investigated, as it is crucial for an efficient allocation of experimental resources [72].

The uncertainties in the model parameters may be better described using Bayesian Inference methods. These methods combine the experimental data and the forward model used to simulate the system. The Bayesian methods produce posterior distributions from which predictions of the unknown model parameters can be inferred. The posterior pdfs are obtained using as inputs the prior knowledge of the inputs (expertise in the area), and the likelihood deduced from the observational data (experimental data obtained from sensors). The posterior pdfs may be used to support maintenance decision-making such as repairment, further observation, etc.

It is crucial to design the methodology of the experimental data collection in such a manner that the amount of information gain is maximised, and therefore, the uncertainties of the parameters and posterior pdfs are minimised. The design of the experiment should define the experimental conditions, what variables, with which frequency, and where they are measured. This problem is commonly referred to as Optimal Experimental Design (OED) [45]. Within the context of Bayesian Inference the subject is referred to as Bayesian OED. The OED problems may be very complex, and have multiple ramifications, as the underlying models are usually governed by a system of partial differential equations (PDEs), and complications such as nonlinearities, couplings, etc., may be present.

Solving the PDEs systems required to characterise the physical model implies a significant computational effort. The high dimensionality of the problem after it is discretized, combined with dependency of the uncertain parameters to be estimated with respect to (w.r.t.) time and space, and the requirement of complex numerical integrations w.r.t. the uncertain parameters, make the solution of the OED computational problem highly complex [72]. The numerous times the evaluation of the OED objective function has to be performed, further increases the vast requirement of numerical resources.

4.1 Paper II – 'On The Investigation of Utility Functions on Optimal Sensor Locations'

This paper is reproduced as originally published with minor revisions:

'Igea F., Chatzis M.N.., Cicirello A., On the Investigation of Utility Functions on Optimal Sensor Location. Proceedings of the 4th ECCOMAS thematic conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2021), 2021'.

https://doi.org/10.7712/120221.8030.19059

On the Investigation of Utility Functions on Optimal Sensor Locations

Felipe Igea[1]*, Manolis N. Chatzis[1] and Alice Cicirello[1,2]

[1] Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

[2] Department of Engineering, Faculty of Civil Engineering and Geosciences, Department of Engineering Structures, Delft University of Technology, Stevinweg 1, Delft 2628, NL

*Corresponding author: felipe.igea@eng.ox.ac.uk

Abstract

Structural Health Monitoring uses data collected from sensors placed on structures to determine their operating condition and whether maintenance is required. Often, optimal sensor placement strategies are used to find the optimal locations for the identification of their modal properties, structural parameters and/or abnormal behaviours under the influence of model and measurement uncertainty. An approach that has been frequently used to solve the problem of sensor placement is the Bayesian experimental design. This approach chooses the locations using the data measured by the sensors to reduce the prior uncertainty of the parameters that are being inferred. The Bayesian experimental design minimizes the uncertainty of the parameters to be inferred through the use of metrics called utility functions. Most of these metrics are based on functions of the posterior distribution. In this paper, the use of three utility functions (Bayesian D-posterior precision, Bayesian A-posterior precision, and Expected Information Gain) is investigated for the problem of sensor placement.

The case study chosen consists of a beam with translational and rotational springs connected to the ground subject to an impulsive load. The goal of the analysis is to select the most informative position of a sensor in order to update the distribution of two uncertain physical parameters of the beam based on natural frequencies extracted using the Eigensystem Realization Algorithm. It is shown that for the case investigated, the three utility functions yield the same optimal sensor location.

Keywords: Optimal Sensor Placement, Uncertainty Quantification, Structural Health Monitoring.

1 Introduction

Structural Health Monitoring (SHM) often focuses on non-intrusive structure damage detection [1]. It can be used to provide early warnings on the health status of engineering systems. The equipment required for implementing SHM, includes sensors and data acquisition systems. The real time information obtained from the sensors has to be post-processed and statistical procedures are implemented to detect anomalies and suggest preventive actions [1,2]. Technological advances in sensor monitoring allow the development of optimal sensor strategies that have made SHM cost effective and easier to implement.

Structural parameters are usually inferred from the sensorial data (such as velocity or acceleration measurements), especially the modal parameters (natural frequency, damping ratio and mode shape). The inferred parameters are then used to assess the acceptability of the models of structures and to evaluate the structures' condition. For some cases, local forms of damage may be identified by a shift in the modal properties [3]. The position of the sensor can strongly influence the inference on the structural parameters. This has led to the widespread development of optimal sensor strategy techniques [4]. Broadly speaking, the sensor placement framework methods can be split into methods based on information theory or non-information methods [4]. The non-information-based methods are not discussed in this paper, however, more information on these techniques can be found in [4]. Work based on information theory heavily relies on the application and development of the general Bayesian framework [5]. This framework was proposed for system identification in [5,6] and it has been consequently extended to the problem of sensor placement [7–10]. The main objective is the selection of the location and number of sensors that maximises the information needed to estimate the uncertain parameters [8]. The research challenges linked to these approaches are the definition of the metric to be used to assess the different configurations of sensors (number and location of the sensors) and the choice of the most adequate optimisation technique [11,12].

In this paper a Bayesian experimental design framework [13] is used to solve the sensor placement problem. Within this framework, the number and locations of the sensors are chosen by using the

data obtained from the sensors to reduce the prior uncertainty on the parameters to be inferred. Therefore, the framework's focus is the minimization of the uncertainty of certain physical parameters of interest to the practitioner by comparing different metrics, the so-called utility functions. Two physical parameters of a beam attached to ground using translational and rotational springs subject to an impulsive load have to be inferred by using a single sensor. The beam is investigated by building a Finite Element model. Numerical simulations of the dynamic response (velocity signal) at different locations are used to obtain numerical 'measurements' of possible sensor locations. An intermediate step requires the post-processing of the numerical 'measurements' to obtain the modal parameters that are subsequently used as the data used to reduce the model parametric uncertainty via Bayesian model updating. This is achieved by using the Eigensystem Realization Algorithm. Model updating is then used to obtain the posterior probability density function of the parameters to be inferred, having assigned a uniform prior distribution and applying Monte Carlo sampling-based strategies. The obtained posterior is used to evaluate a utility function that is then used to select the optimal sensor location. Three utility functions are investigated.

2 Bayesian Optimal Design framework

Bayesian Optimal Design [13] allows the designation of resources required to obtain information for reduction of systematic error, inference of unknown parameters (i.e., reduction of prior uncertainty), obtaining future predictions and the comparison of models chosen to represent a system [13]. The framework's objective is the maximisation of the information obtained from a set of measurements for the inference of the unknown parameters of the model used to describe the physical system [13]. The choice of the optimal design improves parameter inference and reduces the experimental costs.

Lindley proposed a unifying theory of Bayesian Optimal Design in [14]. The definition of the best possible design given a set of objectives and restrictions is described by a utility function. The maximization of this function is used to choose the possible design that measures how well the set of objectives and restrictions are obeyed [15].

One of the major challenges in Bayesian Optimal Design methods is the reduction of their high computational cost incurred in the calculation of their utility functions [13]. This is because the utility functions require the knowledge of the posterior distribution of the parameters to be inferred. These distributions are dependent on the set of measurements available and therefore are different when different designs are considered.

2.1 Bayesian framework

Probability density functions are used to model the uncertain model parameters in the Bayesian Inference framework [5,16]. The prior knowledge on the uncertain parameters before any measurements or data is obtained, is described by the prior density function $p(\boldsymbol{\theta})$. The likelihood function $p(\boldsymbol{y}|\boldsymbol{\theta})$, is normally assumed to follow a specific distribution (e.g., Gaussian). The $p(\boldsymbol{y}|\boldsymbol{\theta})$ measures the degree of suitability of the model to justify the obtained measurements. The denominator $p(\boldsymbol{y})$ of eq.(1) below is the evidence pdf and normalizes the pdf of the posterior. If the above described pdfs are known, the eq.(1) can be used to calculate the so-called posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\theta})\,p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})} \tag{1}$$

The posterior distribution obtained can then be used to determine the utility function. Hence, it is important to obtain accurate estimations of both location (median or mean) and scale (interquartile range or standard deviation) of the posterior [13]. Most frequently, it is not possible to express the posterior distributions with a closed form, so computational methodologies are used to obtain samples from the posterior or to approximate it [16–21]. In this work, the sampling-based model updating techniques are used. Specifically, the Sequential Monte Carlo (SMC) [16] sampling and the Transitional Markov Chain Monte Carlo (TMCMC) [20] are chosen to infer the two physical parameters of the case study investigated.

2.2 Bayesian utility functions

Many different utility functions have been developed for inferring parameters of a model [13]. Metrics that quantify the performance of experiments are obtained by using a set of utility functions that are maximized (or minimized) with the objective of identifying the optimal experiment [13]. Three utility functions are reviewed in what follows.

A well-known utility function is expressed as the inverse of the determinant of the posterior covariance matrix [13]. This utility function also known as the Bayesian D-posterior precision maximises the posterior precision of the model parameters to be inferred and it is given by [13]:

$$U_D(\boldsymbol{d},\boldsymbol{y}) = \frac{1}{\det(\mathrm{cov}(\boldsymbol{\theta}\,|\,\boldsymbol{d},\boldsymbol{y}))} \tag{2}$$

Where $\boldsymbol{d}$ is the vector that represents the experimental design to be optimized (e.g. the sensor positions for a given number of sensors).

Another useful utility function similar to the Bayesian D-posterior precision is given by the inverse of the trace of the posterior covariance matrix [15]. This utility function also known as the 'Bayesian A-posterior precision' maximises the marginal posterior precision of the model parameters to be inferred and it is given by [15]:

$$U_A(\boldsymbol{d},\boldsymbol{y}) = \frac{1}{trace(\mathrm{cov}(\boldsymbol{\theta}\,|\,\boldsymbol{d},\boldsymbol{y}))} \tag{3}$$

Alternatively, the utility function may be expressed as the expected Kullback–Leibler (KL) divergence from the posterior distribution to the prior distribution [14]. The expected KL divergence utility function is also known as Expected Information Gain (EIG) over the parameters to be inferred [22], and it is expressed as:

$$U_{EIG}(\boldsymbol{d}) = E_y\left[D_{KL}\left(p_{\boldsymbol{\theta}|y}\,\|\,p_{\boldsymbol{\theta}}\right)\right] = \int_Y\int_Q p(\boldsymbol{\theta}\,|\,\boldsymbol{y})\log\left(\frac{p(\boldsymbol{\theta}\,|\,\boldsymbol{y})}{p(\boldsymbol{\theta})}\right)p(\boldsymbol{y})\,d\boldsymbol{y}d\boldsymbol{\theta} \tag{4}$$

Where $\mathbb{E}_{\boldsymbol{y}}$ is the expectation with respect to the measurements $\boldsymbol{y}$, $D_{KL}\left(p_{\boldsymbol{\theta}|\boldsymbol{y}} \| p_{\boldsymbol{\theta}}\right)$ is the KL-divergence from the posterior distribution to the prior distribution, $\mathcal{Y}$ and $\mathcal{Q}$ are the support of the measurements $\boldsymbol{y}$ and the parameters to be inferred $\boldsymbol{\theta}$ respectively.

The EIG can be interpreted as a non-linear generalization of the Bayesian D-optimal utility function [23]. It has been found [24] that this metric can be approximated using a Monte Carlo approach:

$$\tilde{U}(\boldsymbol{d}) = \frac{1}{N_{out}} \sum_{i=1}^{N_{out}} \left\{ \ln\left[p\left(\boldsymbol{y}^i \mid \boldsymbol{\theta}^i, \boldsymbol{d}\right)\right] - \ln\left[p\left(\boldsymbol{y}^i \mid \boldsymbol{d}\right)\right] \right\} \tag{5}$$

$$p\left(\boldsymbol{y}^i \mid \boldsymbol{d}\right) \approx \frac{1}{N_{in}} \sum_{j=1}^{N_{in}} p\left(\boldsymbol{y}^i \mid \boldsymbol{\theta}^j, \boldsymbol{d}\right) \tag{6}$$

Where N_{out} is the number of samples used in the outer loop and N_{in} is the number of samples used in the inner loop of the Monte Carlo approximations. The samples are obtained from the prior distribution and the likelihood is evaluated for these samples.

The number of likelihood function evaluations can be reduced if $N = N_{in} = N_{out}$, so that [24,25]:

$$\hat{U}(\boldsymbol{d}) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \ln\left[p\left(\boldsymbol{y}^i \mid \boldsymbol{\theta}^i, \boldsymbol{d}\right)\right] - \ln\left(\frac{1}{N} \sum_{j=1}^{N} p\left(\boldsymbol{y}^i \mid \boldsymbol{\theta}^j, \boldsymbol{d}\right) \right) \right\} \tag{7}$$

However, this result is a biased estimate of the EIG [24,25]. A large number of samples may be required if the prior assumed has a large support at regions of low probability density, as this results in arithmetic underflow [26].

Another way to calculate the EIG is by calculating the difference between the differential entropy of the prior $h(\boldsymbol{\theta})$ and the differential entropy of the posterior $h(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{d})$ [15]:

$$U_{EIG,2}(\boldsymbol{d}) = I(\boldsymbol{\theta}, \boldsymbol{y}) = h(\boldsymbol{\theta}) - h(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{d}) \tag{8}$$

The calculation of the differential entropy using samples from the posterior distribution can be

approximated using the recursive copula splitting approach given in [27].

3 Numerical results

A beam connected to the ground via two sets of translational and rotational spring positioned along the length of the beam, as shown on fig. 1, is investigated in this paper. This simple case study has been chosen as it can represent a variety of practical situations where a component is attached to some fixtures, but there are uncertainties that may due to its assembly, boundary conditions and/or manufacture. In particular, in this case, the location of the first set of springs and the magnitude of the rotational spring are investigated. A prior distribution is assigned to each of these two parameters. The goal of the analysis is to select the most informative position of a sensor in order to update the distribution of these two uncertain parameters. The utility functions defined in section 2 are used to assess the optimal position of the sensor.

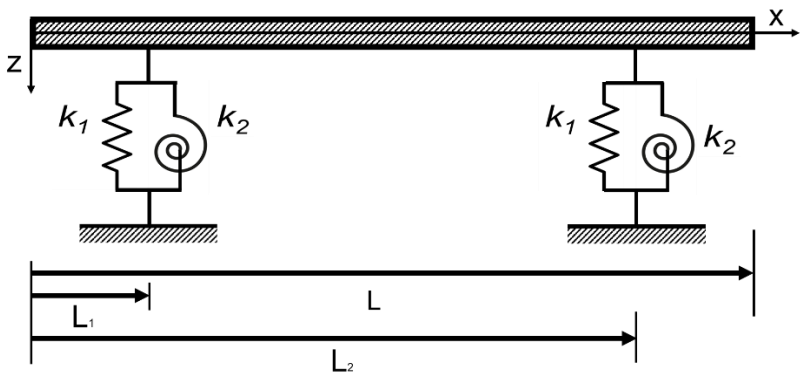


Fig. 1. Beam attached to ground by translational and rotational springs.

The following geometric and material properties were used: L (length)=0.6 m; b (base)=0.04 m; h (height)=0.003 m; ρ (density) =8000 kg/m3; E (Young's modulus) =100 GPa; k_1 (translational spring stiffness) = 1 kN/m; k_2 (rotational spring stiffness) = 101.7 Nm/rad; L_1 (length to springs) = 0.181 m; L_2 (length to springs) = 0.4 m. Modal damping was introduced into the system (η=0.01 for all modes). A force F (triangular pulse of length 10 ms and maximum amplitude of 50 N) applied at length L is used to excite the beam. The parameters to be inferred are the stiffness k_2 of both rotational springs and the location L_1 of the first rotational spring.

A Finite Element (FE) model is used to calculate the transversal velocity signals of the beam at several locations, to investigate the effects of the position of a single sensor on the utility functions. In particular, a 2-dimensional Euler-Bernoulli beam model is considered. This is discretized uniformly using 200 Euler-Bernoulli beam FEs with 2 degrees of freedom per node. Moreover, to simulate experimental conditions, for each transversal velocity signal measure at each node point, ten different realizations are created contaminating each signal using a white Gaussian noise with a noise to signal ratio (rms) of 5%.

The numerically contaminated velocity signals obtained at each possible sensor location (where the locations available are the ones at each node of the FE system) are post-processed using the Eigensystem Realization Algorithm (ERA) [28] to calculate the modal properties. Therefore, it was required to apply ERA 200 times to cover all the possible sensor locations in the system. These modal properties are then used as the data observed in the likelihood function. The likelihood function is then approximated by using the kernel smoothing function (`ksdensity` function of MATLAB [29]) on the set of modal properties obtained from ERA using the 10 different realizations of the contaminated velocity signals for each possible sensor location. Uniform priors were used for both the stiffness (100 Nm/rad to 103 Nm/rad) of the rotational springs and location (0.17 m to 0.19 m) of the rotational spring. The joint posterior distribution of k_2 and L_1 is calculated using two Bayesian model updating techniques [16]: Sequential Monte Carlo (SMC) sampling and the Transitional Markov Chain Monte Carlo (TMCMC). In the SMC sampling approach [16] the samples obtained from the prior were reused in all possible sensor locations to reduce the amount of forward simulations needed and to investigate how the bias resulting from this approach could affect the calculation of the utility functions. The results obtained with this implementation of SMC were compared with the result obtained using the unbiased TMCMC [20] that required new simulations each time a possible sensor location was considered. While SMC required 20,000 forward simulations to obtain acceptable estimations of the posterior distribution, the TMCMC required only 6,000 simulations but each time a new sensor position was considered the forward simulations could not be

reused.

Figures 2, 3 show the precision values obtained for the Bayesian D-posterior precision and Bayesian A-posterior precision utility functions as a function of a sensor location along the length of the beam when using SMC and TMCMC. In the following figures, for the TMCMC approach, no data is shown for positions above a value of around 0.5 m, this occurs because from that position the TMCMC algorithm fails to calculate the utility functions. However, the SMC approach is able to calculate the utility functions for those locations.

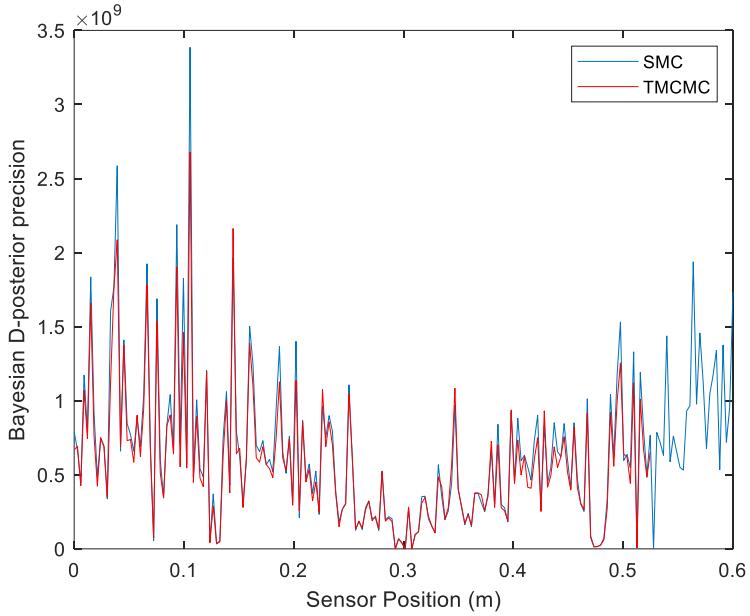


Fig. 2. Bayesian D-posterior precision values vs sensor location.

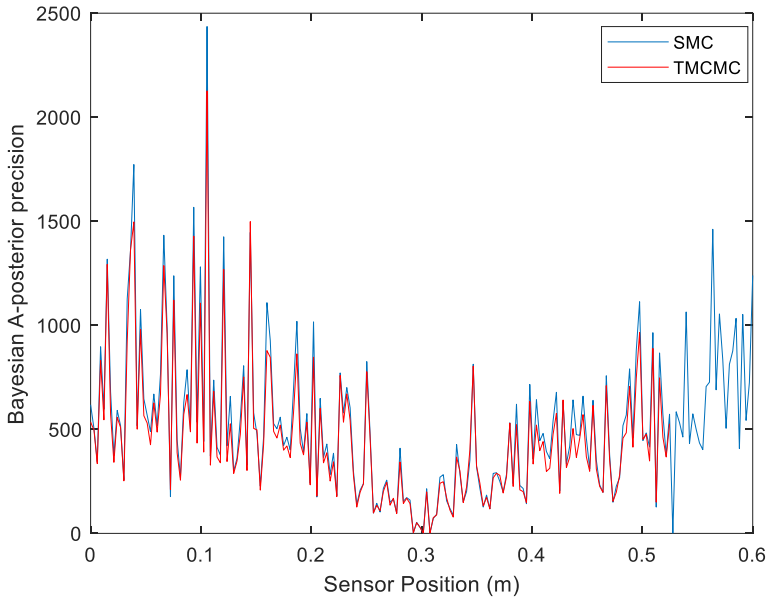


Fig. 3. Bayesian A-posterior precision values vs sensor location.

The results obtained with the EIG utility function are shown in fig. 4. These results were obtained by using the recursive copula splitting approach from [27] as using the Monte Carlo approximation shown in eq.(7) resulted in evaluating likelihoods at supports of low probability density which lead to arithmetic underflow.

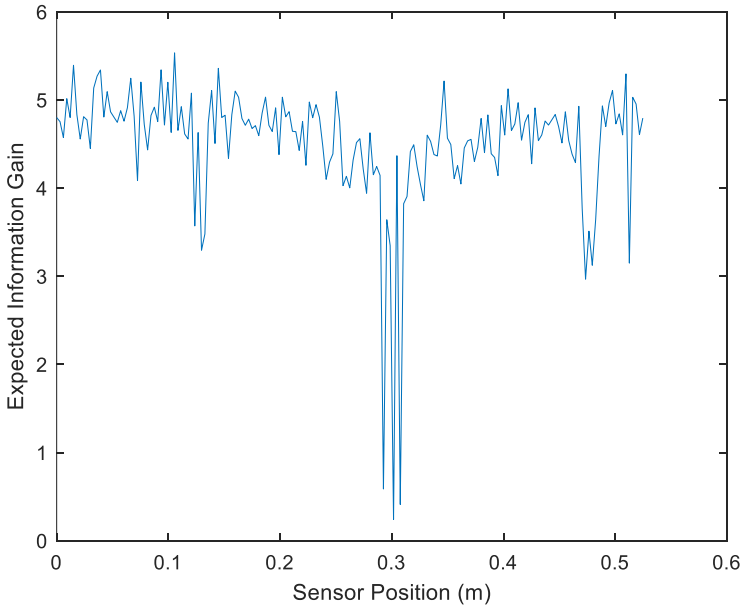


Fig. 4. Expected information gain vs sensor location.

It can be observed that the three utility functions used have identified the same best sensor location that is 0.106 m. Locations where the utility values were low were close to nodal points and hence the modal properties resulting from ERA were less accurate. For this beam, the results obtained using the three different utility functions investigated, have been found to be similar. However, it is expected that if a large number of physical parameters would have been inferred, the utility functions may have shown dissimilar results as the sensor location that maximises the joint posterior precision may have not been the same as the sensor location that maximises the marginal posterior precision.

It was also shown that for this case study the SMC and TMCMC provide similar results. The discrepancies found in the values of the utility function are largely due to the bias introduced by reusing samples from the prior and the choice of using a sequential importance sampling algorithm instead of a sequential importance resampling algorithm. If a sequential importance resampling algorithm had been chosen a lower bias would have been introduced in exchange for a higher computational cost.

4 Conclusions

The optimal sensor placement for the identification of two physical parameters of a beam attached to ground by translational and rotational springs has been investigated by considering three utility functions: Bayesian D-posterior precision and Bayesian A-posterior precision, and Expected Information Gain. It was shown that these utility functions led to the same best sensor location. As expected, poor values of the utility function were found for locations close to nodal points, as the modal properties estimated by ERA were less accurate. This result is expected as the measurements obtained at nodal points would not have as much information as other points along the beam system.

The utility functions chosen require the calculation of the posterior distribution. Therefore, the computational cost is reliant on the Bayesian Inference technique being used. However, the choice of the inference technique usually shows a trade-off between computational cost and accuracy. Reusing samples, as in SMC techniques, may limit the amount of likelihood evaluations but this is at the risk

of not evaluating samples close to regions of high probability densities. However, it was found that SMC and TMCMC lead to the same results for the case under investigation. The current challenge for the Bayesian Optimal Design approach would be the development of a fast inference technique that estimates the posterior at a limited computational cost.

References

[1] F. Magalhães, Á. Cunha, E. Caetano, Online automatic identification of the modal parameters of a long span arch bridge, Mechanical Systems and Signal Processing. 23 (2009) 316–329. https://doi.org/10.1016/j.ymssp.2008.05.003.

[2] C.R. Farrar, K. Worden, Structural health monitoring : a machine learning perspective, Wiley, Chichester, 2013.

[3] J.J. Moughty, J.R. Casas, A State of the Art Review of Modal-Based Damage Detection in Bridges: Development, Challenges, and Solutions, Applied Sciences. 7 (2017) 510. https://doi.org/10.3390/app7050510.

[4] D. Li, Herausgeber: Claus-Peter Fritzen Sensor Placement Methods and Evaluation Criteria in Structural Health Monitoring, 2011.

[5] J.L. Beck, L.S. Katafygiotis, Updating Models and Their Uncertainties. I: Bayesian Statistical Framework, Journal of Engineering Mechanics. 124 (1998) 455–461. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:4(455).

[6] L.S. Katafygiotis, J.L. Beck, Updating Models and Their Uncertainties. II: Model Identifiability, Journal of Engineering Mechanics. 124 (1998) 463–467. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:4(463).

[7] C. Papadimitriou, Optimal sensor placement methodology for parametric identification of structural systems, Journal of Sound and Vibration. 278 (2004) 923–947. https://doi.org/10.1016/j.jsv.2003.10.063.

[8] C. Papadimitriou, G. Lombaert, The effect of prediction error correlation on optimal sensor placement in structural dynamics, Mechanical Systems and Signal Processing. 28 (2012) 105–127. https://doi.org/10.1016/j.ymssp.2011.05.019.

[9] C. Argyris, C. Papadimitriou, G. Lombaert, Optimal sensor placement for response predictions using local and global methods, in: Conference Proceedings of the Society for Experimental Mechanics Series, Springer New York LLC, 2020: pp. 229–236. https://doi.org/10.1007/978-3-030-12075-7_26.

[10] C. Argyris, C. Papadimitriou, G. Samaey, G. Lombaert, A unified sampling-based framework for optimal sensor placement considering parameter and prediction inference, Mechanical Systems and Signal Processing. 161 (2021) 107950. https://doi.org/10.1016/j.ymssp.2021.107950.

[11] W. Ostachowicz, R. Soman, P. Malinowski, Optimization of sensor placement for structural health monitoring: a review, Structural Health Monitoring. 18 (2019) 963–988. https://doi.org/10.1177/1475921719825601.

[12] Y. Tan, L. Zhang, Computational methodologies for optimal sensor placement in structural health monitoring: A review, Structural Health Monitoring. (2019) 1–22. https://doi.org/10.1177/1475921719877579.

[13] E.G. Ryan, C.C. Drovandi, J.M. Mcgree, A.N. Pettitt, A Review of Modern Computational Algorithms for Bayesian Optimal Design, International Statistical Review. 84 (2016) 128–154. https://doi.org/10.1111/insr.12107.

[14] D. v. Lindley, Bayesian Statistics, Society for Industrial and Applied Mathematics, 1972. https://doi.org/10.1137/1.9781611970654.

[15] C. ben Issaid, Bayesian Optimal Experimental Design Using Multilevel Monte Carlo, 2015. https://repository.kaust.edu.sa/handle/10754/552705 (accessed June 8, 2021).

[16] A. Lye, A. Cicirello, E. Patelli, Sampling methods for solving Bayesian model updating problems: A tutorial, Mechanical Systems and Signal Processing. 159 (2021) 107760. https://doi.org/10.1016/j.ymssp.2021.107760.

[17] E. Simoen, G. de Roeck, G. Lombaert, Dealing with uncertainty in model updating for damage assessment: A review, Mechanical Systems and Signal Processing. 56 (2015) 123–149. https://doi.org/10.1016/j.ymssp.2014.11.001.

[18] J.L. Beck, S.-K. Au, Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation, (n.d.). https://doi.org/10.1061/ASCE0733-93992002128:4380.

[19] H.-F. Lam, J.-H. Yang, S.-K. Au, Markov chain Monte Carlo-based Bayesian method for structural model updating and damage detection, Structural Control and Health Monitoring. 25 (2018) e2140. https://doi.org/10.1002/stc.2140.

[20] J. Ching, Y.-C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging., Journal of Engineering Mechanics. 133 (2007) 816. https://doi.org/10.1061/(ASCE)0733-9399(2007)133:7(816).

[21] D.M. Blei, A. Kucukelbir, J.D. McAuliffe, Variational Inference: A Review for Statisticians, Journal of the American Statistical Association. 112 (2017) 859–877. https://doi.org/10.1080/01621459.2017.1285773.

[22] Z. Xu, Q. Liao, Gaussian process based expected information gain computation for bayesian optimal design, Entropy. 22 (2020) 258. https://doi.org/10.3390/e22020258.

[23] K. Chaloner, I. Verdinelli, Bayesian experimental design: A review, Statistical Science. 10 (1995) 273–304. https://doi.org/10.1214/ss/1177009939.

[24] X. Huan, Y.M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, Journal of Computational Physics. 232 (2013) 288–317. https://doi.org/10.1016/j.jcp.2012.08.013.

[25] K.J. Ryan, Estimating Expected Information Gains for Experimental Designs with Application to the Random Fatigue-Limit Model, Journal of Computational and Graphical Statistics. 12 (2003) 585–603. https://doi.org/10.1198/1061860032012.

[26] J. Beck, B.M. Dia, L.F.R. Espath, Q. Long, R. Tempone, Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain, Computer Methods in Applied Mechanics and Engineering. 334 (2018) 523–553. https://doi.org/10.1016/j.cma.2018.01.053.

[27] G. Ariel, Y. Louzoun, Estimating differential entropy using recursive copula splitting, Entropy. 22 (2020) 236. https://doi.org/10.3390/e22020236.

[28] J.N. Juang, R.S. Pappa, An eigensystem realization algorithm for modal parameter identification and model reduction, Journal of Guidance, Control, and Dynamics. 8 (1985) 620–627. https://doi.org/10.2514/3.20031.

[29] MATLAB, (2020).

4.2 Conclusions

In the paper II attached in this chapter, the sensor placement problem is investigated using a Bayesian experimental design framework. Information collected by the sensors is used to define the locations and number of sensors that minimise the prior uncertainty associated with the parameters to be estimated. Consequently, the primary objective of the framework is the minimisation of the uncertainty related to specific physical parameters defined by the practitioner. This is accomplished by comparison of various metrics known as utility functions.

In paper II, an example of sensor placement involving a beam connected to the ground by two supports, each built with two springs, one rotational and one translational is illustrated. The example investigates the optimal sensors' locations along the beam according to three different utility functions, so two physical parameters can be identified. It was found that the readings of sensors located at areas close to nodal points produce poor estimations of the modal properties, and therefore, their information content is low. It was also found that for all three utility functions, the optimal locations to place the sensors were the same.

Theoretically, the algorithm can be applied for an increasing number of sensors. This would be done by seeking the configurations that maximise the utility functions for those greater number of sensors. However, as the number of sensors is increased, the computational cost of the calculation escalates to the point of becoming unfeasible. An alternative procedure to obtain the optimal sensor configuration would be to use a sequential approach. This procedure would begin with a unique sensor for which the position that maximises the chosen utility function is found. Then, without changing the position selected for the first sensor, a second sensor location is chosen in such a manner that it maximizes the utility function. The process is iterated until the chosen number of sensors is reached. Although the scheme might not result in the most optimal sensor configuration, it reduces significantly the computational cost associated with the sensor location optimization.

An important issue identified in this paper is the robustness of the selection of the optimal sensor location. This phenomenon may be studied considering the values of the utility function in the proximity of the optimal location. It may occur that the value of the utility function drops significantly in the proximity of the optimal location, and therefore, the placement of the sensor will be very sensitive to misplacement. In paper II, the key issue identified is the requirement of a quick inference method that may allow the calculation of the posterior of the relevant uncertain parameters at a low computational cost. This key issue is tackled in the next chapter.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	On The Investigation of Utility Functions on Optimal Sensor Locations
Publication Status	☒Published □ Accepted for Publication □Submitted for Publication □Unpublished and unsubmitted work written in a manuscript style
Publication Details	Igea F., Chatzis M.N.., Cicirello A., On the investigation of Utility functions on Optimal Sensor Location. Proceedings of the 4th ECCOMAS thematic conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2021), 2021.

Student Confirmation

Student Name:	Felipe Igea		
Contribution to the Paper	Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing.		
Signature	Felipe Igea	Date	16/09/23

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title:	Dr Alice Cicirello		
Supervisor comments	The student made a substantial contribution to the publication, and the description above is accurate		
Signature		Date	18/09/23

Chapter 5 - Limited information in operating conditions – updating our physics-based model

To assess the behaviour of an engineering system characterised by uncertain parameters, statistically updated models may be used. Observations taken from an experiment may be used as inputs in a statistical model updating framework with the purpose of defining the values or intrinsic variability of those latent parameters [19].

The engineering problem of inferring some uncertain parameters that is encountered in some physical systems characterised by an expensive-to-run detailed deterministic model, is tackled in this chapter. The detailed deterministic model is considered as a black-box function. The likelihood function calculated during the inference process is determined by analysing the differences between the features derived from the physics-based model, and the ones generated by the gathered experimental data. The objective is to present a methodology that can be implemented even in the cases where there are restrictions in computational and/or time resources, and therefore, the number of runs the simulation code can be performed is limited. This approach allows a computationally efficient Bayesian model updating approach.

Two main disadvantages arise when sampling-based methods are employed: a first drawback is that usually, these techniques demand the adjustment of a significant number of parameters to achieve sampling efficiency of the algorithm; a second weakness is the necessity of using a significant number of samples, and as a consequence, the high computational burden implied. The applicability of these sampling methods might be limited if tight time constraints or restricted computational budget are present, as those techniques involve multiple evaluations of the expensive-to-run models used to describe the physics of the real system. Additionally, some limitations in the scalability and range of applicable problems [53,54,61–63,73] are found in the MCMC techniques [74–76].

Some engineering problems in structural dynamics imply multimodal distributions of some model parameters. Therefore, a methodology able to manage this kind of distribution of the model

parameters is required. Precise capture of multimodal distributions may not be performed unless many simulations are run. This is a common problem in engineering as only a limited number of evaluations is possible, and this increases the chance of poor predictions of the uncertain parameters.

With the objective of overcoming the above-mentioned limitations, VI is investigated. The Machine Learning community has employed VI [55] with the purpose of approximating posterior distributions using an optimisation-based perspective to the inference process.

5.1 Paper III – 'Structural Model Updating Using Variational Inference'

This paper is reproduced as originally published with minor revisions:

'Igea F., Chatzis M.N., Cicirello A., Structural Model Updating Using Variational Inference, Structural Health Monitoring 2021, Enabling Next-Generation SHM For Cyber-Physical Systems, 2021'.

https://doi.org/10.12783/shm2021/36282

5.2 Paper IV – 'Cyclical Variational Bayes Monte Carlo for Efficient Multi-Modal Posterior Distributions Evaluation'

This paper is reproduced as originally published with minor revisions:

'Igea F., Cicirello A., Cyclical Variational Bayes Monte Carlo for Efficient Multi- Modal Posterior Distributions Evaluation, Mechanical Systems and Signal Processing, 2023'.

https://doi.org/10.1016/j.ymssp.2022.109868

Structural Model Updating Using Variational Inference

Felipe Igea[1]*, Manolis N. Chatzis[1] and Alice Cicirello[1,2]

[1] Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

[2] Department of Engineering, Faculty of Civil Engineering and Geosciences, Department of Engineering Structures, Delft University of Technology, Stevinweg 1, Delft 2628, NL

*Corresponding author: felipe.igea@eng.ox.ac.uk

Abstract

Monte Carlo sampling approaches are frequently used for probabilistic model updating of physics-based models under parametric uncertainty due to their high accuracy. The model updating framework produces a model that represents the real system more accurately than the prior knowledge or assumptions. This statistically updated model may prove useful if Structural Health Monitoring (SHM) techniques are to be applied. However, the updating of the models requires the use of a high number of samples, implying a high computational cost. Another additional disadvantage of these methods is that most of them require the calibration of a high number of parameters for their algorithm to become sampling efficient.

Variational inference (VI) is an alternative approach for inference often used by the machine learning community. An optimization algorithm is employed to choose from a family of distributions the member that best approximates the posterior. In the method described in this paper the variational posterior that maximises the evidence lower bound (ELBO) is chosen.

An approach based on VI is proposed and implemented on two different numerical examples to infer the uncertain parameters by postulating a variational posterior distribution given by a multivariate Gaussian approximation. It has been found that the number of samples required for the calculation of the posterior is reduced compared with Monte Carlo sampling approaches, however this occurs at the

cost of some accuracy. The methodology will be helpful for the development of enhanced SHM strategies that require fast inference under a limited computational budget.

1. Introduction

Structural Health Monitoring (SHM) includes non-intrusive methods for detection of abnormal conditions on a structure [1]. To achieve so, a Bayesian Inference framework can be used to update a model using measurements obtained from sensors deployed in the structure so that possible structural damages can be observed [2]. In particular, Model Updating (MU) techniques can be employed for the detection and assessment of possible structural damages [3].

The application of stochastic MU methods based on sampling techniques might be severely limited if the behaviour of the structure is described by a model of high complexity [4,5], as in these cases the methodology implies a high computational cost or lengthy lead times. In particular, in the structural dynamics community stochastic MU methods have relied on the use of sampling Monte Carlo methods such as the Markov Chain Monte Carlo (MCMC) technique [3]. These techniques can introduce a bias [6], and even though the Markov Chain converges to the posterior distribution, the number of samples required to achieve convergence may not be known before the algorithm is run [3]. Therefore, their applicability for optimal sensor placement statistical schemes [7], and for Bayesian experimental design [8] is compromised.

As an alternative to the computationally expensive MCMC techniques, Machine Learning (ML) users have developed an optimization approach for the inference problem called Variational Inference (VI) [9]. The VI methods suggest several distributions from which, based on the available data, the distribution that approximates best the posterior is selected at a lower computational cost. These methods differ from approaches such as variations of the Kalman Filter [10,11] that are commonly used for online time series forecasting and input estimation. The main differences that VI methods have are that the postulated posterior is not limited to Gaussian distributions, how the sampling is chosen and the optimisation routine used to update the unknown variables is based on the

maximisation of the Evidence Lower Bound (ELBO).

In this paper, an efficient methodology to infer uncertain parameters for structural systems described by deterministic models, based on the VI methods, requiring a limited number of runs, is proposed.

2. Bayesian Inference framework

The Bayesian Inference framework allows the combination of a physics-based model with uncertain model parameters $\boldsymbol{\theta}$ and information $\boldsymbol{y}_{obs}$ obtained in the form of measurements on some observed variables. Before any measurements are performed [3], the prior knowledge of the uncertain parameters $\boldsymbol{\theta}$ is described by the prior pdf $p(\boldsymbol{\theta})$. As new information becomes available in the form of measurements on some variables $\boldsymbol{y}_{obs}$, a likelihood function $p(\boldsymbol{y}_{obs} \mid \boldsymbol{\theta})$ can be used to describe how acceptable is the model to describe the observed measurements given the uncertain parameters $\boldsymbol{\theta}$, is chosen. The likelihood function $p(\boldsymbol{y}_{obs} \mid \boldsymbol{\theta})$ is constructed as a function of both the model and the measurements performed [3]. Using the Bayes theorem [3], the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs})$ of the uncertain parameters of the model $\boldsymbol{\theta}$ are calculated using eq.(1):

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}}) = \frac{p(\boldsymbol{\theta})\, p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta})}{p(\boldsymbol{y}_{\text{obs}})} \tag{1}$$

The posterior pdf $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs})$ is normalised using the evidence term $p(\boldsymbol{y}_{obs})$. The evidence term $p(\boldsymbol{y}_{obs})$ is usually neglected as it is not necessary in the inference of the uncertain parameters [3].

The accuracy that can be reached using sampling methods is mainly limited by the number of samples used by the sampling algorithm [3]. As a result, those methods' precision is restricted only by the computational power and time available.

3. Variational inference for model updating

Variational inference (VI) is a method that frames the probabilistic inference of the posterior probability density function (pdf) into an optimization approach [9]. Usually, in this approach the

posterior pdf is approximated assuming that the posterior pdf belongs to a family of densities $\mathcal{Q}$, and the member $q(\boldsymbol{\theta})$ that best approximates the posterior density is chosen using an optimization scheme [9]:

$$q^*(\boldsymbol{\theta}) = \arg\min_{q(\boldsymbol{\theta})\in\mathcal{Q}} KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) \tag{2}$$

where KL is the Kullback-Leibler (KL) divergence (relative entropy), defined as:

$$KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) \equiv \int_{\mathcal{S}} q(\boldsymbol{\theta}) \log\left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})}\right) d\boldsymbol{\theta} \tag{3}$$

The KL divergence is used to measure how dissimilar is a distribution compared to a second reference distribution. It is a non-negative measure and it is non-symmetric.

The calculation of the optimal member $q(\boldsymbol{\theta})$ and its parameters that minimize eq.(3), cannot be analytically computed as the evidence term should be known beforehand [9].

In this paper, to approximate the posterior, a multivariate Gaussian with a full covariance matrix able to capture highly correlated posteriors for which parameters are obtained by maximizing of the evidence lower bound (ELBO) using stochastic gradient descent [12], is assumed, as described in what follows.

3.1 Postulated posterior

The multivariate Gaussian approximation of the posterior facilitates the capture of posteriors $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs})$ that are frequently encountered in engineering applications. The $q(\boldsymbol{\theta})$ is therefore expressed as:

$$q(\boldsymbol{\theta}) \equiv q_{\phi}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}, \Sigma) \tag{4}$$

In eq.(4) above, Σ is a full covariance matrix, and μ is the mean. Parameterization of the variational posterior $q(\boldsymbol{\theta})$ is performed in terms of $\boldsymbol{\phi} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

3.2 Evidence lower bound (ELBO) equation

It is not feasible to calculate analytically the KL divergence between the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs})$ and its approximation $q(\boldsymbol{\theta})$ in a direct manner [9]. Therefore, it is common [9] to rewrite the objective function in terms of the evidence lower bound (ELBO) shown by eq.(5) that equals to a constant minus the KL divergence between the prior $p(\boldsymbol{\theta})$ and posterior approximation $q(\boldsymbol{\theta})$.

$$ELBO(q) = \mathbb{E}\left[\log p(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}})\right] - \mathbb{E}\left[\log q(\boldsymbol{\theta})\right] = \mathbb{E}\left[\log p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta})\right] - KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta})) \quad (5)$$

This leads to the following optimization problem:

$$\arg\max_{q(\boldsymbol{\theta}) \in \mathcal{Q}} ELBO(q) \equiv \arg\min_{q(\boldsymbol{\theta}) \in \mathcal{Q}} KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) \quad (6)$$

For most of the problems found on structural dynamics, the models are highly complex, and therefore, an analytical solution for eq.(6) cannot be found, as no simple expression of the expectation $\mathbb{E}\left[\log p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})\right]$ exists.

3.3 Implementation of proposed method

The proposed approach can be summarized in the following steps:

i. A set of simulations of the physics-based model to be updated are run.
    a. For the samples chosen, prior and likelihood values are calculated.
ii. Calculation of ELBO value.
iii. Update the terms of the variational posterior.
iv. Check if stopping criteria have been met, if not repeat from step i.

A flowchart of the main blocks involved in the proposed method are shown in fig. 1.

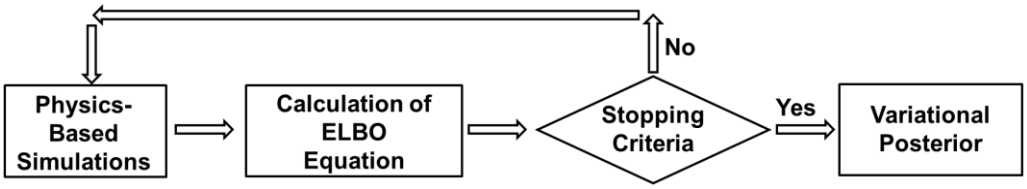


Fig. 1. Flowchart of proposed approach.

At each iteration a prescribed number of iterations is set to calculate the ELBO value. The stopping criteria can be chosen to be based on a total number of simulations or number of iterations or when the algorithm has reached convergence. To assess if convergence has been reached, a parameter defined as patience term [12] is specified. This patience term is the number of successive times that it is admissible for the ELBO to have an equal or greater value to the preceding ELBO.

The stochastic Gaussian variational approximation algorithm proposed by Zhou et al [12], is used to solve the optimization problem in eq. (6). The method factorizes the covariance matrix employing a low rank plus diagonal (LR+D) structure, guaranteeing the unique optimal variational parameters through the introduction of two geometrical manifold constraints on those variational parameters [12]. The detailed steps of the algorithm can be found in [12].

4. Numerical results

Numerical results obtained using the proposed method for two different examples are shown: the first a two-degree-of-freedom (2-DoF) mass-spring system and the second a uniform mass beam system. For both examples, results obtained with the proposed method and MCMC are compared.

4.1 Two degree of freedom system

Let us consider the 2-DoF mass-spring system shown in fig. 2 below. The 2-DoF shear mass-spring system has springs of equal stiffnesses $k_1 = k_2 = 2\,N/m$; and masses $m_1 = 1kg$ and $m_2 = 5\,kg$. The natural frequencies that correspond to the specified properties of the two-degree-of freedom system are given by: $f_1 = 0.4359\ Hz, f_2 = 2.0518\ Hz$. To simulate experimental conditions where the natural frequencies are taken as measurements to update the physics-based model, two assumptions are made:

the system is assumed to experience free vibration and the two natural frequencies of the system follow likelihood functions with means the actual deterministic values of those two natural frequencies, and standard deviations of 2% of the values of the natural frequencies. In this example, the 2nd mass is assumed to be unknown and using the 'experimental measurements' the posterior pdf of the m_2 parameter is updated.

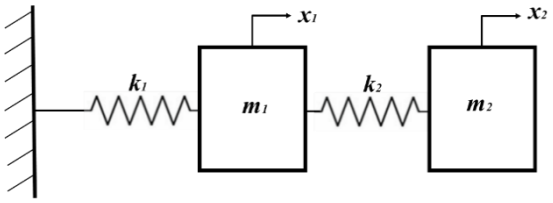


Fig. 2. Two degree of freedom system.

For the proposed method, to achieve convergence on the parameters of the Gaussian variational approximation, with the patience term set to five, it was required to run 34 iterations of 5 samples each (5 simulations of the physics-based model), giving a total of 170 simulations.

When running the MCMC approach using the Metropolis-Hastings algorithm, 2000 samples were needed for convergence (in mean and standard deviation) to the posterior distribution.

Results shown in the Table 1 indicate that to obtain values of comparable accuracy the proposed approach requires a significantly lower number of samples than the MCMC method. The number of samples refers to the number of forward evaluations of the physics-based model.

It can be seen from fig. 3b that after approximately eight iterations the ELBO value stabilizes. During the first eight iterations, a high variation of the values of the mean and standard deviation of the variational Gaussian approximation is observed in fig. 3c and fig. 3d. After approximately twenty-five iterations, the mean and standard deviation reach values significantly similar to the final estimations.

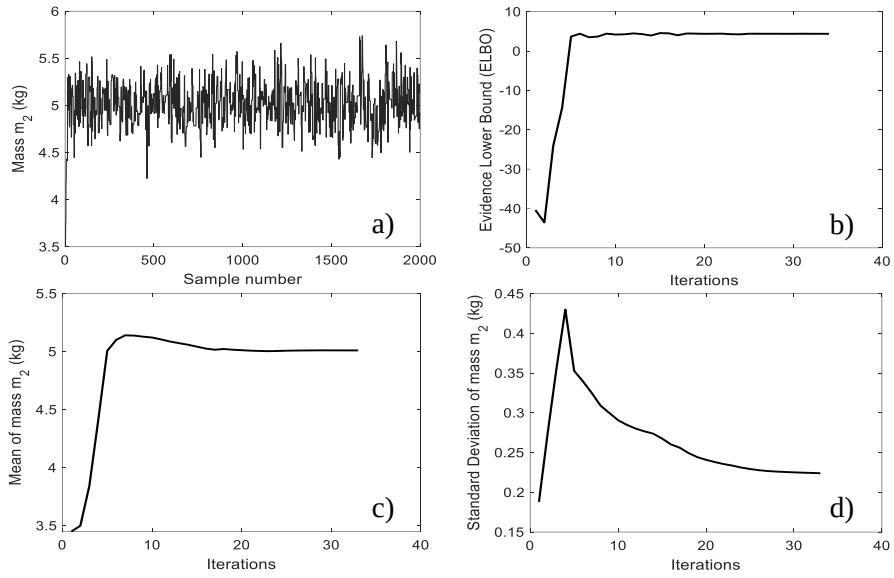


Fig. 3. Results for 2-DoF shear mass-spring system using MCMC and proposed approach. a) MCMC trace plot for 2nd mass. b) Evidence lower bound (objective function) at each iteration. c) Mean of the normal distribution of the 2nd mass at each iteration. d) Standard deviation of the normal distribution of the 2nd mass at each iteration.

Table 1: Comparison of results

Method	Mean	Std.	N. of samples	Iterations	ELBO
MCMC	5.03	0.213	2,000	-	-
Proposed approach	5.01	0.226	170	34	4.38

4.2 Uniform mass beam system

A uniform mass beam system connected to ground via two sets of translational and rotational springs positioned along the length of the beam, as shown on fig. 4 below, is investigated in this section. This case study has been chosen as it can represent a variety of practical situations where a component is attached by some fixtures, but there are uncertainties in the physical properties of the fixtures that may be due to its assembly, boundary conditions and/or manufacture.

An FE model is used to calculate the natural frequencies of the system assuming free vibration conditions. Moreover, to simulate experimental conditions, it was assumed that the likelihood function of each natural frequency has as the mean the values of deterministic system, and as its standard deviation a 2% magnitude. The following geometric and material properties were used: L (length)=0.6 m; b (base)=0.04 m; h (height)=0.003 m; ρ (density) =8000 kg/m3; E (Young's modulus) =100 GPa; k_1 (translational spring stiffness) = 1 kN/m; k_2 (rotational spring stiffness) = 500 Nm/rad; L_1 (length to springs) = 0.2 m; L_2 (length to springs) = 0.4 m. The natural frequencies that correspond to the specified properties of the uniform mass beam system are given by: $f_1 = 9.26\ Hz$, $f_2 = 26.80\ Hz$, $f_3 = 48.74\ Hz$, $f_4 = 86.17\ Hz$, $f_5 = 210.06\ Hz$, $f_6 = 295.75\ Hz$, $f_7 = 407.80\ Hz$ and $f_8 = 623.59\ Hz$. In this example, the density of the uniform mass beam ρ and the rotational springs stiffnesses k_2 are assumed to be unknown and using the 'experimental measurements' their posterior pdf is updated.

As shown on fig. 5, a total of 850 simulations (85 iterations of 10 samples each) has been required to achieve convergence on the parameters of the Gaussian variational approximation, where a patience value of ten was used. A total of 2000 samples (forward evaluations of the physics-based models) were required for the posterior distribution to converge when the MCMC method was used.

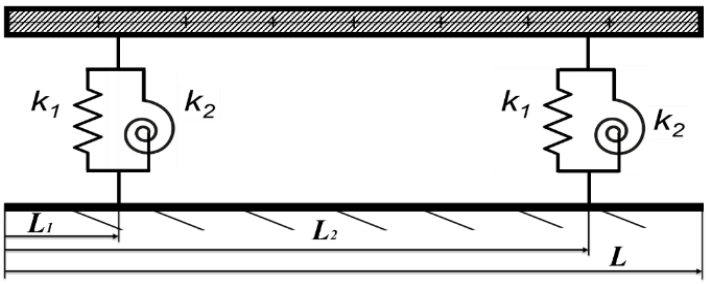


Fig. 4. Beam system attached to ground by translational and rotational springs adapted from [13].

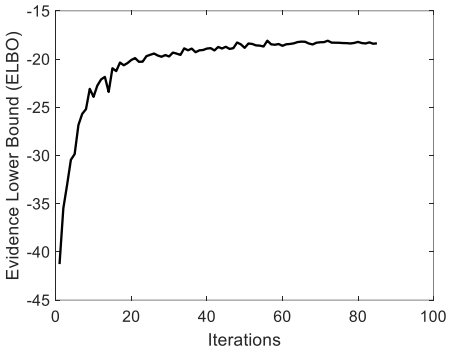


Fig. 5. Results for uniform mass beam system of the evidence lower bound (objective function) at each iteration for the proposed approach.

Table 2: Comparison of results

Method	Mean	Cov.	Std.	N. of samples	Iterations	ELBO
MCMC	$\begin{bmatrix} 8023.5 \\ 511.8 \end{bmatrix}$	$\begin{bmatrix} 26810.9 & 6254.2 \\ 6254.2 & 2677.6 \end{bmatrix}$	$\begin{bmatrix} 163.7 \\ 51.7 \end{bmatrix}$	2,000	-	-
Proposed approach	$\begin{bmatrix} 8006.2 \\ 508.0 \end{bmatrix}$	$\begin{bmatrix} 20405.2 & 4329.1 \\ 4329.1 & 2119.0 \end{bmatrix}$	$\begin{bmatrix} 142.8 \\ 46.0 \end{bmatrix}$	850	85	-18.37

As in the previous example, Table 2 above shows that to find results of comparable accuracy, the number of samples required by the proposed approach, is lower than the one needed when the MCMC method is used. The proposed approach is able to deal with several highly correlated parameters due to the use of a full covariance matrix.

5. Conclusion

An approach has been proposed to estimate the posterior distribution of the uncertain parameters of physics-based model when new data becomes available. In particular, Variational Inference (VI) for model updating is used as an alternative to sampling approaches to reduce the number of model evaluations. In the cases studied (a two-degree of freedom system and a beam system attached to

ground by translational and rotational springs) is has been shown that the number of samples required to obtain an estimation of the posterior with the proposed approach are lower than when using the MCMC approach. However, a reduction of computational time is usually seen at the expense of lower accuracy due to the reduced number of simulations and/or postulated posterior that are not able to capture the true posterior.

The proposed approach uses a multivariate Gaussian postulated posterior with a full covariance matrix able to capture highly correlated posteriors. This approach is useful for cases where the uncertain parameters can be approximated by a multivariate Gaussian distribution. A possible extension to this approach would be the use of Gaussian mixtures for the postulated posterior, as they are able to capture complex smooth posteriors and/or multimodal posteriors. The method would also benefit from the development of a sample-efficient algorithm that chooses samples in the vicinity of the high values of the unnormalized posterior, and from the reuse of samples in such a manner that the amount of simulations would be further reduced. These topics are currently under investigation.

References

[1] C.R. Farrar, K. Worden, Structural health monitoring : a machine learning perspective, Wiley, Chichester, 2013.

[2] A. Rytter, Vibrational Based Inspection of Civil Engineering Structures, (1993).

[3] A. Lye, A. Cicirello, E. Patelli, Sampling methods for solving Bayesian model updating problems: A tutorial, Mechanical Systems and Signal Processing. 159 (2021) 107760. https://doi.org/10.1016/j.ymssp.2021.107760.

[4] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, D.M. Blei, Automatic Differentiation Variational Inference, Journal of Machine Learning Research. 18 (2016) 1–45. http://arxiv.org/abs/1603.00788 (accessed May 12, 2021).

[5] R. Ranganath, S. Gerrish, D.M. Blei, Black Box Variational Inference, Journal of Machine Learning Research. 33 (2013) 814–822. http://arxiv.org/abs/1401.0118 (accessed May 20, 2021).

[6] J.K. Kruschke, Markov Chain Monte Carlo, in: Doing Bayesian Data Analysis, Elsevier, 2015: pp. 143–191. https://doi.org/10.1016/B978-0-12-405888-0.00007-6.

[7] C. Argyris, S. Chowdhury, V. Zabel, C. Papadimitriou, Bayesian optimal sensor placement for crack identification in structures using strain measurements, Structural Control and Health Monitoring. 25 (2018) e2137. https://doi.org/10.1002/stc.2137.

[8] E.G. Ryan, C.C. Drovandi, J.M. Mcgree, A.N. Pettitt, A Review of Modern Computational Algorithms for Bayesian Optimal Design, International Statistical Review. 84 (2016) 128–154. https://doi.org/10.1111/insr.12107.

[9] D.M. Blei, A. Kucukelbir, J.D. McAuliffe, Variational Inference: A Review for Statisticians, Journal of the American Statistical Association. 112 (2017) 859–877. https://doi.org/10.1080/01621459.2017.1285773.

[10] M.N. Chatzis, E.N. Chatzi, S.P. Triantafyllou, A Discontinuous Extended Kalman Filter for non-smooth dynamic problems, Mechanical Systems and Signal Processing. 92 (2017) 13–29. https://doi.org/10.1016/J.YMSSP.2017.01.021.

[11] S.E. Azam, E. Chatzi, C. Papadimitriou, A. Smyth, Experimental validation of the Kalman-type filters for online and real-time state and input estimation:, Http://Dx.Doi.Org/10.1177/1077546315617672. 23 (2015) 2494–2519. https://doi.org/10.1177/1077546315617672.

[12] B. Zhou, J. Gao, M.-N. Tran, R. Gerlach, Manifold Optimization Assisted Gaussian Variational Approximation, Journal of Computational and Graphical Statistics. (2019). https://arxiv.org/abs/1902.03718v2 (accessed September 12, 2021).

[13] F. Igea, M.N. Chatzis, A. Cicirello, ON THE INVESTIGATION OF THE EFFECT OF POPULATION UNCERTAINTY ON OPTIMAL SENSOR LOCATIONS, 4th International Conference on Uncertainty Quantification in Computational Sciences and Engineering. (2021) 168–177. https://doi.org/10.7712/120221.8030.19059.

Cyclical Variational Bayes Monte Carlo for Efficient Multi-Modal Posterior Distributions Evaluation

Felipe Igea[1]* and Alice Cicirello[1,2]

[1] Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

[2] Department of Engineering, Faculty of Civil Engineering and Geosciences, Department of Engineering Structures, Delft University of Technology, Stevinweg 1, Delft 2628, NL

*Corresponding author: felipe.igea@eng.ox.ac.uk

Abstract

Multi-modal distributions of some physics-based model parameters are often encountered in engineering due to different situations such as a change in some environmental conditions, and the presence of some types of damage and non-linearity. In statistical model updating, for locally identifiable parameters, it can be anticipated that multi-modal posterior distributions would be found. The full characterization of these multi-modal distributions is important as methodologies for structural condition monitoring in structures are frequently based in the comparison of the damaged and healthy models of the structure. The characterization of posterior multi-modal distributions using state-of-the-art sampling techniques would require a large number of simulations of expensive-to-run physics-based models. Therefore, when a limited number of simulations can be run, as it often occurs in engineering, the traditional sampling techniques would not be able to capture accurately the multi-modal distributions. This could potentially lead to large numerical errors when assessing the performance of an engineering structure under uncertainty.

Therefore, an approach is proposed for drastically reducing the number of models runs while yielding accurate estimates of highly multi-modal posterior distributions. This approach introduces a cyclical

annealing schedule into the Variational Bayes Monte Carlo (VBMC) method to improve the algorithm's phase of exploration and the finding of high probability areas in the multi-modal posteriors throughout the different cycles.

Three numerical and one experimental investigations are used to compare the proposed cyclical VBMC with the standard VBMC algorithm, the monotonic VBMC and the Transitional Ensemble Markov Chain Monte Carlo (TEMCMC). It is shown that the standard VBMC fails in capturing multi-modal posteriors as it is unable to escape already found regions of high posterior density. In the presence of highly multi-modal posteriors, the proposed cyclical VBMC algorithm outperforms all the other approaches in terms of accuracy of the resulting posterior, and number of model runs required.

***Keywords*:** Bayesian Inference; Variational Inference; Bayesian Quadrature; Gaussian Process; Model Updating; Cyclical Annealing;

1 Introduction

Statistical model updating techniques are frequently used in engineering to quantify the inherent variability of some uncertain latent parameters, or to identify the unknown values of latent parameters used in physics-based models in the light of measurements of some observable quantities [1]. These statistically updated models can then be used to evaluate the behaviour of an engineering system under uncertainties. For example, the statistically updated model can be used for assessing the performance of a structure with uncertain input parameters under various loading and environmental conditions, and/or to assess the remaining useful life of such structure [1–4].

Multi-modality on the distributions of some physics-based model's parameters of an engineering system is frequently found [5]. Multi-modality in the latent parameters of a physics-based model may be encountered due to different reasons such as changes in the environmental conditions [6] (e.g., change in stiffness due to varying temperature), the presence of some types of damage and non-

linearity [7] (e.g., change in localised stiffness due to opening and closing of a crack). In statistical model updating it is also expected to find multi-modal distributions for locally identifiable problems [8] (e.g., multiple stiffnesses combinations result in the same observations).

Several papers have illustrated practical engineering examples, in which either the input or response parameters follow multi-modal distributions. In these engineering examples, as either the input or response parameters are shown to follow multi-modal distributions, it is expected that the distribution of the latent uncertain parameters that are affected by either the input or response will also result in a multi-modal distribution. The vibratory load undertaken by the blade of a wind turbine [9], shows a multi-modal distribution when under stochastic excitations. Stresses found at the start and shutdown of generator turbine rotors also show multi-modal distributions [9]. Using long-term monitoring data, it was verified [10,11] that the structural fatigue stress of a steel bridge follows a bi-modal distribution. For nanostructured zirconia coatings, it was observed [12] that the Knoop microhardness follows a bi-modal distribution. Papers [13,14] demonstrated that the axle load spectra can be used for estimating the relative pavement damage of roads. The axle load spectra follow a bimodal distribution as it considers the trucks' weights when unloaded and loaded.

This paper is focused on the inference of multi-modal uncertain parameters of expensive-to-run detailed physics-based models frequently encountered in engineering problems. Within the standard statistical model updating framework [1], the misfit between the features extracted from the measurements and those obtained from the model are used to calculate the likelihood function that is used in the inference scheme. However, because of computational budget and/or time constraints, the number of model evaluations that can be carried out, may be limited. This would significantly hinder the accuracy of the resulting posterior even when applying state-of-the-art statistical model updating approaches. For example, statistical model updating is often implemented by using sampling-based techniques [1]. However, these techniques, including Markov Chain Monte Carlo (MCMC), show a trade-off between computational cost and accuracy, as the convergence of the Markov Chain to the posterior distribution is improved as the chain size lengthens, and therefore the number of physics-

based model runs increases [1]. Moreover, Monte Carlo-based techniques introduce a bias, and the number of runs required to achieve convergence is generally unknown when starting the algorithm [1,15].

Alternatively, Variational inference [16] has been used by the machine learning community to estimate posterior distribution approximations employing an optimization approach to reduce the number of model runs required for the inference problem. In simple terms, most variational inference methods propose a family of distributions where the member of the family that best approximates the posterior is chosen [16]. Compared to sampling techniques such as MCMC [17–19], the recent variational inference techniques [20–24] are more numerically scalable and may be used in a wider range of problems due to significant advances in the optimization process [16]. Nevertheless, MCMC based techniques are still the preferred method, as they guarantee convergence to the correct posteriors [23]. However, the disadvantage of these techniques is their high computational cost in order to yield accurate posterior distribution estimates, especially in the presence of highly multi-modal posteriors.

The Variational Bayes Monte Carlo (VBMC) [24,25] has been recently developed to provide an efficient estimation of the model evidence and of the posterior. The method combines active-sampling Bayesian quadrature [26,27] with variational inference [16]. In a nutshell: (a) a postulated posterior is obtained using a Gaussian mixture; (b) the parameters of the Gaussian mixture are obtained using the evidence lower bound (ELBO) as the objective function to be maximised; (c) the expensive to evaluate log unnormalized posterior distribution is replaced by a statistical surrogate model constructed using a GP [28]; (d) active sampling is carried out using ‘smart’ acquisition functions applied to the GP model to perform a guided local refinement of the GP model; (e) the Bayesian quadrature [26,27] is implemented to carry out fast integrations in the variational objective; (f) a warm-up process is introduced to avoid the algorithm getting initially stuck in areas of very low probability under the true posterior. During the initial phases of the warm-up, significant improvements of the ELBO are rapidly obtained; (g) the algorithm adaptively adjusts the number of

components in the variational mixture, adding or removing components based on the level of improvement found on the solution. As a result, the VBMC framework [24,25] is highly efficient. However, the application of VBMC to statistical model updating in engineering problems requires addressing the following challenges: (i) How to select the limited number of initial simulations to build the initial GP? (ii) How to select the new samples to account for multi-modality in the posterior distribution?

To tackle these challenges, the cyclical VBMC approach is proposed. Both first (i) and second (ii) issues, are tackled by introducing an artificial temperature parameter that anneals the unnormalized posterior. This parameter improves the exploration abilities and mode coverage of the algorithm, so the limitations introduced by the limited number of samples and a poor initialization are overcome. This annealing schedule enhances the exploration phase of the cycle and the discovery of regions of high probability density in multi-modal posteriors, as it avoids the algorithm getting stuck in the initially found regions of high probability.

The core of the inference strategy employed in the standard VBMC is the same as the one shown in [29], and it is based on Bayesian quadrature and Variational Inference with a postulated multivariate Gaussian mixture. However, both approaches [29] and [24,25], are such that once a mode of the distribution is found, they are not capable to explore further the uncertain variable domain to identify other modes.

Compared to the work by Ni et al [29], the main differences introduced by the proposed algorithm are: (a) the use of a different acquisition function that selects new points prioritizing the areas of greater probability density compared to the acquisition function based on the absolute value of the mean divided by the standard deviation of the GP surrogate model; (b) variational whitening is performed to deal with posteriors that are highly correlated; (c) convergence criteria based on ELBO compared to the use of criteria related to the vector of variational parameters and the values of the Gaussian mixture weights; (d) the introduction of a warm-up process; (e) the adaptive adjustment of components in the variational mixture; (f) a cyclical annealing schedule to improve the exploration

capabilities of the algorithm for dealing with multi-modal posteriors.

The performance of the proposed cyclical VBMC method is assessed with respect to the standard VBMC [24,25], the monotonic VBMC, and the state-of-the-art sampling approach Transitional Ensemble Markov Chain Monte Carlo (TEMCMC) [30] considering both multi-modal and unimodal posteriors of physics-based models' parameters.

The paper is structured as follows. In section 2, the Bayesian model updating framework and variational inference are reviewed. The main building blocks on the cyclical VBMC algorithm are described in detail in section 3. The results obtained from three numerical examples and one experimental investigation are presented in section 4. The conclusions of the proposed cyclical VBMC algorithm are then discussed in section 5.

2 Bayesian model updating framework

The Bayesian model updating strategy enables the combination of a physics-based model that includes uncertain parameters $\boldsymbol{\theta}$ which cannot be directly observed (also known as latent variables), described by probability density functions, the so-called prior distribution, with new information obtained from measurements of some observable quantities $\boldsymbol{y}_{obs}$ [2,3]. These measurements can be expressed in different forms such as time history, modal properties, etc. This approach results into an updated physics-based model with parameters described as probability density functions, the so-called posterior distributions. This statistical updated model, that is more representative of the real system, can then be used to investigate the behaviour of the system under different loading conditions in order to predict its performance with respect to safety, quality, design or cost constraints [1–4].

In particular, a prior probability density function $p(\boldsymbol{\theta})$ that reflects the prior knowledge of the uncertain parameters $\boldsymbol{\theta}$ before any measurements on some observable variables $\boldsymbol{y}$ are taken, is assigned to the parameters. Then a likelihood function $p(\boldsymbol{y}_{\text{obs}}|\boldsymbol{\theta})$ that reflects the level of acceptability of the physics-based model, given a set of uncertain parameters $\boldsymbol{\theta}$ to describe the

measurements is constructed. This is done by using features extracted (e.g., natural frequencies) from the response obtained with the physics-based model, and the corresponding ones extracted from some measurements. The approach results into an updated statistical physics-based model with its latent variables described as posterior probability density functions $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})$ that can be calculated using [2,3]:

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}}) = \frac{p(\boldsymbol{\theta})\, p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta})}{p(\boldsymbol{y}_{\text{obs}})} \tag{1}$$

where $p(\boldsymbol{y}_{\text{obs}})$ is defined as the evidence, and it serves as a normalization constant for the posterior probability density functions. The posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})$ can be computed analytically if the prior and likelihood distributions are part of the conjugate family. However, this is not necessarily always the case, and therefore, numerical integration may be necessary.

As the evidence term in Bayesian Inference is a numerical constant, and it is independent of the uncertain parameters $\boldsymbol{\theta}$ [1], sampling techniques (e.g., MCMC) [8,31,32] can be used to obtain samples from the posterior distribution using the following proportional relationship:

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}}) \propto p(\boldsymbol{\theta})\, p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta}) \tag{2}$$

Variational inference takes an alternative approach to sampling methods by minimising the KL divergence between the best member $q(\boldsymbol{\theta})$ of a postulated family of densities Q and the posterior density, and therefore bypassing the calculation of the evidence term [16]. The posterior distribution is then obtained by transforming the statistical inference problem into an optimization problem. The optimization scheme chooses the member of the family $q(\boldsymbol{\theta})$ that is 'closest' to the posterior density by converting the minimization of the KL divergence in a maximization of the evidence lower bound (ELBO) objective function. An extensive review of variational inference can be found in [16].

In computer science, several approaches [21–25,33,34] have been developed to circumvent the need of an analytical expression of the ELBO equation, and to reduce the number of evaluations of the

physics-based model to be carried out. One of these approaches is the Variational Bayesian Monte Carlo (VBMC**)** [24,25]. In this paper, a variant of the approach called cyclical VBMC is proposed for addressing the statistical updating problems in engineering where multi-modal posteriors are expected.

3 Cyclical Variational Bayesian Monte Carlo framework

The cyclical VBMC approach is based on the VBMC algorithm, it has been developed to deal with multi-modal posteriors in an efficient way by introducing an artificial temperature parameter that anneals the unnormalized posterior. The proposed method overcomes the drawbacks of limited function evaluations and a poor initialization of the VBMC algorithm by introducing the cyclical schedule that improves the exploration abilities and mode coverage of the algorithm.

Given an expensive-to-evaluate computational model of an engineering system, for which prior information on the unknown latent parameters and measurements obtained from the engineering system are available, the proposed approach aims at minimising the number of function evaluations compared to state of the art Bayesian sampling approaches, while obtaining an accurate description of the posterior. The approach consists of two main parts: the initialization of the algorithm and the procedure used to update the parameters in the posterior variational distribution. These two main parts are shown in fig. 1 and fig. 2 respectively.

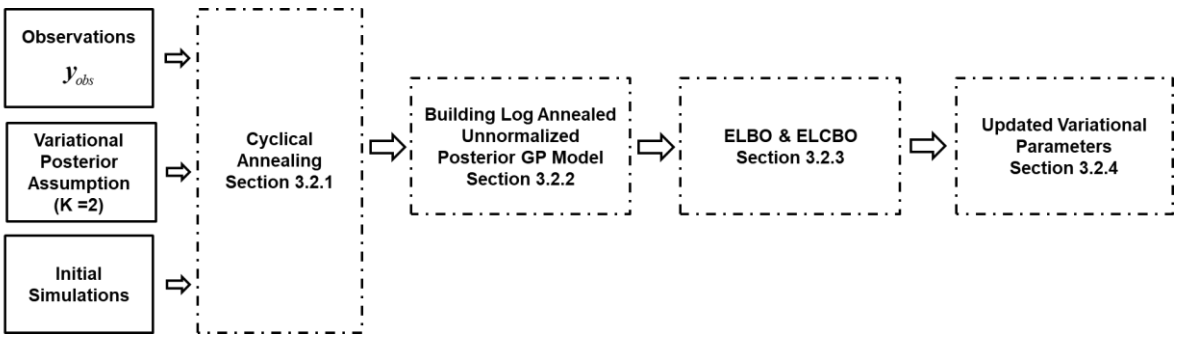


Fig. 1. Initialization blocks of Cyclical VBMC algorithm.

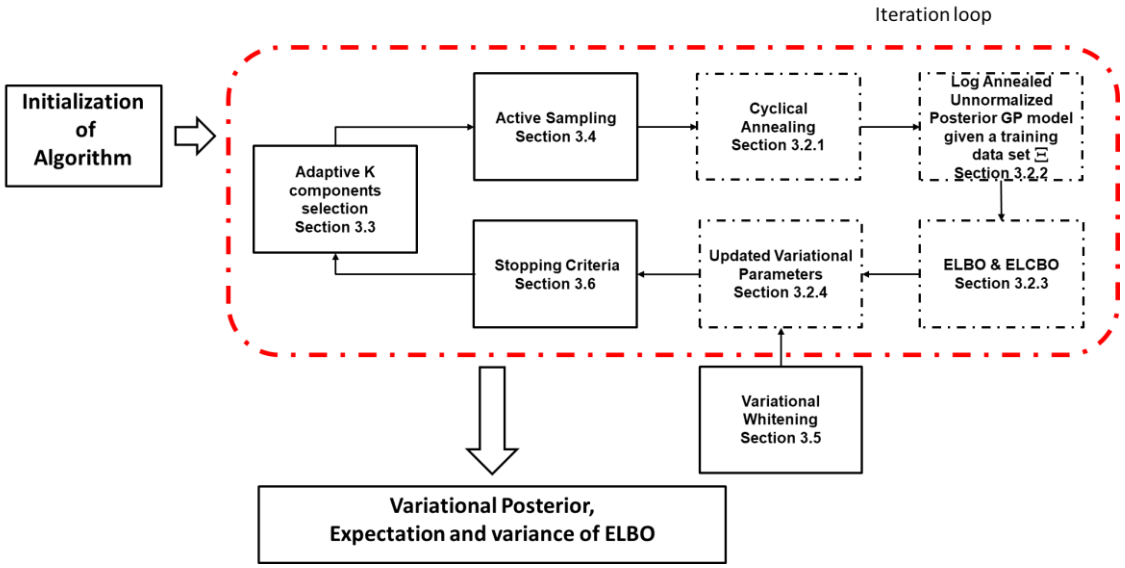


Fig. 2. Cyclical VBMC algorithm blocks.

The initialization of the algorithm shown in fig. 1 begins with assuming a variational posterior that is flexible, and able to capture smooth posteriors. This is done by using a Gaussian mixture as the postulated posterior. The multivariate Gaussian mixture model provides a flexible way of describing any continuous density by taking linear combinations of Gaussian distributions (with adjusted means, covariances and the linear combination coefficients). Almost any continuous density can be approximated to arbitrary accuracy [35], therefore, this is chosen as the postulated posterior. Then, an initial set of parameters that is given as input to the physics-based simulation is chosen, and its output is calculated. Given an assumed prior and likelihood function, the logarithm of the unnormalized posterior is calculated. Cyclical annealing is also introduced by replacing the log unnormalized posterior values with the annealed log unnormalized posterior values. A Gaussian Process (GP) regression model using as training points the logarithm of the annealed unnormalized posterior values is employed to build a probabilistic surrogate of the logarithm of the annealed unnormalized posterior. Using Bayesian quadrature [26,27], the GP can then be used to calculate the ELBO and evidence lower confidence bound (ELCBO) values. Finally, the updated variational parameters are obtained.

The second part of the algorithm shown in fig. 2, consists of a total of T iterations, and it starts with

active sampling to select samples for the physics-based model that are run at locations that maximize an acquisition function. The acquisition function is chosen in such a manner that sampling is encouraged at high probability regions of the log unnormalized posterior. A prescribed number of cycles and total iterations is set to produce both, an exploration phase and an exploitation phase. The GP regression model is built and the GP's hyperparameters are automatically set by using the maximum-a-posteriori estimates. In this paper it is done by employing initially slice sampling [36], and subsequently gradient based optimization, as recommended to improve computational efficiency in [24,25]. Bayesian Quadrature [26,27] is used to calculate the value of the ELBO. The ELCBO is also calculated, and employed to evaluate the variational approximation's improvement and also as a convergence diagnostic. For each iteration, the stochastic gradient ascent is used to update the parameters of the variational posterior. Variational whitening is also performed every few iterations to deal with highly correlated posteriors. Once the stopping criteria have been fulfilled, the method returns the variational approximation of the posterior, and the expectation and variance of the ELBO.

More details about the approach setup, core building blocks (dashed blocks of fig. 1 and fig. 2), adaptive K components selection, active sampling, variational whitening and stopping criteria are given in the following subsections.

3.1 Approach setup

Given a set of observations $\boldsymbol{y}_{obs}$ and a model $PM(\boldsymbol{x},\boldsymbol{\theta})$, $\boldsymbol{x}$ represents a vector with fixed model properties known in advance, and $\boldsymbol{\theta}$ is a vector of uncertain model parameters. The setup used in the cyclical VBMC algorithm is described in the following subsections.

3.1.1 Variational approximation of the posterior $q(\boldsymbol{\theta})$

The selection of the variational posterior $q(\boldsymbol{\theta})$ is flexible [22], and it should be made with the intent to capture multi-modal posteriors $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs})$ that can be encountered in engineering applications.

Without loss of generality, the variational posterior $q(\boldsymbol{\theta})$ can be expressed with a non-parametric

approximation that is provided by a Gaussian mixture with K components [24,25]:

$$q(\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_k, \gamma_k^2 \Sigma) \tag{3}$$

where the mixture weight, mean and scale factor are respectively given by $w_k, \boldsymbol{\mu}_k, \gamma_k^2$, and Σ is the covariance matrix:

$$\Sigma = diag\left[\lambda^{(1)^2}, \ldots, \lambda^{(d)^2}\right] \tag{4}$$

with d representing the number of unknown parameters to be inferred (i.e., length of the vector of uncertain parameters $\boldsymbol{\theta}$).

The variational posterior is parameterized in terms of the vector of parameters $\boldsymbol{\phi} \equiv \left(w_1, \ldots, w_K, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K, \gamma_1, \ldots, \gamma_K, \boldsymbol{\lambda}\right)$. As a result, the number of parameters to optimise in the variational posterior $q_\phi(\boldsymbol{\theta})$ is given by $d + \left(d+2\right)K$, which is the length of the vector $\boldsymbol{\phi}$.

3.1.2 Selection of initial physics-based simulations

Given the uncertain parameters $\boldsymbol{\theta}$, the initial step requires samples to be generated. If available, the plausible lower bound (PLB) and the plausible upper bound (PUB) which limit the region of the parameter space of high posterior probability mass should be specified. A set of points (as a rule-of-thumb a total of 10 points) situated in the plausible box would be uniformly distributed at random [24,25].

It might occur that the PLB and the PUB are not known. In that case, a set of initial points can be chosen using different sampling methods such as Latin Hypercube Sampling (LHS) [37]. LHS is used in the examples shown in section 4.

The initial points $\boldsymbol{\Theta}_0 = \left[\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{n_{init}}\right]$ are used as inputs for the physics-based model to obtain the output response, where n_{init} is the total number of initial points. The output response can be used to construct features that are then used to evaluate the likelihood values given an assumed likelihood

function. The likelihood function reflects the level of agreement between the features obtained by the mathematical model and the measurements. Then, the unnormalized posterior of the initial points can be calculated to build the GP of the annealed logarithm of the unnormalized posterior as described in subsection 3.2.2.

3.2 Core building blocks

3.2.1 Cyclical annealing

The annealing process is used to flatten the objective function (the ELBO), and to reduce the chance of the algorithm getting stuck in some local optima of the parameters of the variational posterior. The annealing process produces a deterministic deformation of the objective function [38], by means of a temperature parameter [30]:

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})^{\frac{1}{temp}} \propto \left[p(\boldsymbol{\theta})\, p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta}) \right]^{\frac{1}{temp}} \tag{5}$$

A fixed temperature implies that the true objective is optimized at a constant schedule, as implemented in the standard VBMC algorithm. Monotonic annealing schedules, in which the temperature is progressively reduced, are the most frequently used. The temperature decreases until the algorithm reaches the true posterior [39].

Although the cyclical annealing schedule [39–42] has been used to deal with multi-modal posteriors in the machine learning field. In this paper, it is the first time that the cyclical annealing schedule is introduced into the combined Variational Inference and Bayesian Quadrature framework (VBMC method) to yield a better representation of the posterior through the introduction of an exploration phase with an improved target guidance. The introduction of the annealing schedule into the VBMC method is of great interest as it is able to remove its inability to deal with multi-modal posteriors. This occurs because the standard VBMC is unable to escape already found regions of high posterior density. The theoretical foundations of the proposed strategy are described in [43].

Specifically, two phases, exploration and exploitation may be found as the temperature is decreased

from its maximum to its minimum within each cycle. These phases are cyclically repeated for a prescribed number of times to improve convergence. This enables the algorithm to explore areas of high probability density that may otherwise have not been found. Specifically, during the exploration phase, "paths" would start forming in regions where sampling would take place, producing a high coverage of the support of the target distribution. During the exploitation phase, sampling takes place at regions of high probability density. Therefore, the cyclical schedule gradually improves convergence by reopening paths, and by leveraging on the previous cycles as warm re-starts.

The temperature parameter is the inverse of the parameter β_t:

$$temp = 1/\beta_t \tag{6}$$

The parameter β_t is defined in the interval $[0,1]$, and it is calculated for each iteration step in the cyclical VBMC algorithm. According to [39], the β_t can be expressed as:

$$\beta_t = \begin{cases} \dfrac{\tau}{S}, & \tau \le S \\ 1, & \tau > S \end{cases} \tag{7}$$

where:

$$\tau = \frac{\mathrm{mod}\left(t-1, \lceil S/M \rceil\right)}{S/M} \tag{8}$$

and $t = 1:1:T$ is the iteration number, T is the number of total iterations for the annealing schedule, M is the number of cycles, and S is a control parameter. The exceptional case of $\beta_t = 0$ is circumvented by defining $temp$ as an interval variable $temp \in [1,50]$. The control parameter S is set to 0.5 as described in [39].

As a rule-of-thumb, if the user has a maximum number of simulations available N_{sim} (e.g., 1000), 20% of these simulations (200) will be allocated for carrying out the cyclical annealing schedule. Therefore, the total number of iterations for the cyclical annealing is obtained by considering the total

number of simulations assigned per iteration (for example, 5 simulations per iteration would lead to a total of $T = 40$ iterations of the cyclical annealing schedule). The choice of the number of cycles M depends on the trade-off between exploration and exploitation that the user wants to investigate. For example, if $M = 5$, it would mean that 8 iterations form 1 cycle, in which the temperature (eq.(6)) decreases from its maximum to 1. Once the total number of iterations for the cyclical annealing schedule has been reached, the temperature is set to 1.

To introduce the cyclical annealing schedule into the algorithm, the log unnormalized posterior $\log p(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}})$ is replaced with the annealed log unnormalized posterior $\log p_{annealed}(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}})$, that is defined as:

$$\log p_{annealed}(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}}) = \frac{1}{temp} \log p(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}}) \tag{9}$$

3.2.2 Gaussian Process (GP) of the annealed logarithm unnormalized posterior

For the proposed cyclical VBMC, cyclical tempering is introduced into the algorithm, by simply replacing the log unnormalized posterior $f \equiv \log p(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}})$ with the annealed log unnormalized posterior $f \equiv \log p_{annealed}(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}})$.

The annealed log unnormalized posterior $f \equiv \log p_{annealed}(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}})$, is approximated using a GP regression model [28]:

$$f \sim GP\left(m_{gp}(\boldsymbol{\theta}), k_{gp}(\boldsymbol{\theta}, \boldsymbol{\theta}')\right) \tag{10}$$

where $m_{gp}(\boldsymbol{\theta})$ is the mean function, and a covariance matrix is defined in terms of a kernel function $k_{gp}(\boldsymbol{\theta}, \boldsymbol{\theta}')$. The typical choice when little is known about the function to be approximated [28] is to use the squared exponential kernel that is expressed as:

$$k_{gp}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sigma_f^2 \Lambda \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}')^T \Sigma_l^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}')\right) \tag{11}$$

Where σ_f is the output length scale, and:

$$\Lambda \equiv (2\pi)^{\frac{d}{2}} \prod_{i=1}^{d} l^{(i)} \tag{12}$$

is the normalization of the Gaussian, $\boldsymbol{l}$ is the vector of input length scales, the superscript $^{(i)}$ refers to the *i*-th dimension and Σ_l is a diagonal covariance matrix:

$$\Sigma_l = diag(l^{(i)^2}, \ldots, l^{(d)^2}) \tag{13}$$

The likelihood is assumed to be Gaussian with an observation noise $\sigma_{obs} > 0$ (to obtain numerical stability [44]), and the shape of the mean function is assumed to be given by [24,25]:

$$m_{gp}(\boldsymbol{\theta}) = m_0 - \frac{1}{2} \sum_{i=1}^{d} \frac{\left(\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}_{\max}^{(i)}\right)^2}{r^{(i)^2}} \tag{14}$$

Where m_0 is the mean's maximum value, $\boldsymbol{\theta}_{\max}$ is the location of the mean's maximum value, and $\boldsymbol{r}$ is the length scales' vector [24,25].

The hyperparameters that define the GP are collected in a vector $\boldsymbol{\psi} = \left[\boldsymbol{l}, \sigma_f, \sigma_{obs}, m_0, \boldsymbol{\theta}_m, \boldsymbol{r}\right]$, of dimension $3d+3$. These hyperparameters $\boldsymbol{\psi}$, are themselves defined in terms of a uniform distribution or a truncated Student's $\boldsymbol{t}$ distribution with mean μ, standard deviation σ, and $\nu = 3$ degrees of freedom. Some of these GP hyperparameters $\left[\boldsymbol{l}, \sigma_f, \sigma_{obs}, \boldsymbol{r}\right]$ are defined in the log space. The same distributions used in references [25] have been directly implemented and are given in Table 1.

Table 1: Prior's hyperparameters.

Hyperparameter	Description	Prior mean	Prior scale
$\log l^{(i)}$	Input length scale (i-th dimension)	$\log\left[\sqrt{\frac{d}{6}}L^{(i)}\right]$	$\log\sqrt{10^3}$
$\log\sigma_f$	Output scale	Uniform	-
$\log\bar{\sigma}_{obs}$	Base observation noise	$\log\sqrt{10^{-5}}$	0.5
m_0	Mean function maximum	Uniform	-
$\theta_{\max}^{(i)}$	Mean function location (i-th dimension)	Uniform	-
$\log r^{(i)}$	Mean function scale (i-th dimension)	Uniform	-

$L^{(i)}$ is defined by:

$$L^{(i)} = PUB^{(i)} - PLB^{(i)} \tag{15}$$

Where the PLB and PUB limit the region of the parameter space of high posterior probability mass.

The calculation of the maximum-a-posteriori (MAP) of the hyperparameters may be performed by any appropriate algorithm. Another possible option to calculate the hyperparameters would be to use the maximum-likelihood-estimate (MLE) method. In this paper, the MAP of the hyperparameters is first calculated using slice sampling, and the estimation is subsequently switched to a gradient based optimization approach when the variability of the expected log unnormalized posterior is below a threshold [24,25].

By conditioning, the resulting GP predictive posterior mean function $\bar{f}_{\Xi}(\boldsymbol{\theta})$ and posterior covariance function $C_{\Xi}(\boldsymbol{\theta},\boldsymbol{\theta}')$ for a training data set $\Xi=\{\boldsymbol{\Theta},\boldsymbol{h},\boldsymbol{\sigma}_{obs}\}$ (for N training inputs $\boldsymbol{\Theta}=[\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_N]$, and N observations $\boldsymbol{h}=f(\boldsymbol{\Theta})$ with observation noise $\sigma_{obs}>0$) is given in closed-form [28] as:

$$\bar{f}_{\Xi}(\boldsymbol{\theta})\equiv\mathbb{E}\left[f(\boldsymbol{\theta})\,|\,\Xi,\boldsymbol{\psi}\right]=k(\boldsymbol{\theta},\boldsymbol{\Theta})\left[k(\boldsymbol{\Theta},\boldsymbol{\Theta})+\Sigma_{obs}(\boldsymbol{\Theta})\right]^{-1}\left(\boldsymbol{h}-m(\boldsymbol{\Theta})\right)+m(\boldsymbol{\Theta}) \tag{16}$$

$$C_{\Xi}(\boldsymbol{\theta},\boldsymbol{\theta}')\equiv Cov\left[f(\boldsymbol{\theta}),f(\boldsymbol{\theta}')\,|\,\Xi,\boldsymbol{\psi}\right]=k(\boldsymbol{\theta},\boldsymbol{\theta}')-k(\boldsymbol{\theta},\boldsymbol{\Theta})\left[k(\boldsymbol{\Theta},\boldsymbol{\Theta})+\Sigma_{obs}(\boldsymbol{\Theta})\right]^{-1}k(\boldsymbol{\Theta},\boldsymbol{\theta}') \tag{17}$$

The observation noise matrix has the following form:

$$\Sigma_{obs}=diag(\sigma_{obs}^2(\boldsymbol{\theta}_1),\ldots,\sigma_{obs}^2(\boldsymbol{\theta}_N)) \tag{18}$$

As the annealed log unnormalized posterior $\log p_{annealed}(\boldsymbol{\theta},\boldsymbol{y}_{\text{obs}})$ is approximated using a GP model, an analytical computation of the integral involved in the ELBO and ELCBO equation can be derived using Bayesian Quadrature [26,27], as described in what follows.

3.2.3 The Evidence Lower Bound (ELBO) and Evidence Confidence Lower Bound (ELCBO)

The ELBO can now be expressed as [24]:

$$ELBO(\boldsymbol{\phi},f)=\mathrm{E}_q\left[\log p(\boldsymbol{\theta},\boldsymbol{y}_{\text{obs}})\right]-\mathrm{E}_q\left[\log q(\boldsymbol{\theta})\right]=\mathrm{E}_{f|\Xi}\left[\mathrm{E}_{\phi}\left[f\right]\right]+\mathcal{H}\left[q(\boldsymbol{\theta})\right] \tag{19}$$

Where $\mathcal{H}\left[q(\boldsymbol{\theta})\right]$ is the entropy of the variational posterior [24]. The integrals in eq.(19) can be analytically computed [24,25] with the Bayesian MC statistical method also known as Bayesian quadrature, so that [26,27]:

$$\mathrm{E}_{f|\Xi}[Z]=\int\bar{f}_{\Xi}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{20}$$

$$\mathrm{V}_{f|\Xi}[Z]=\int\int C_{\Xi}(\boldsymbol{\theta},\boldsymbol{\theta}')\pi(\boldsymbol{\theta})\pi(\boldsymbol{\theta}')d\boldsymbol{\theta}d\boldsymbol{\theta}' \tag{21}$$

where:

$$Z=\int_{\mathrm{R}^d}g(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{22}$$

In the proposed cyclical VBMC algorithm the function $g(\boldsymbol{\theta})$ is given by the annealed log unnormalized posterior $\log f(\boldsymbol{\theta}) \equiv \log p_{annealed}(\boldsymbol{\theta}, \boldsymbol{y}_{\text{obs}})$ and $\pi(\boldsymbol{\theta})$ is the variational approximation to the posterior $q(\boldsymbol{\theta})$ [11,13].

The variational approximation's $\mathcal{H}[q(\boldsymbol{\theta})]$ entropy is calculated using Monte Carlo sampling, and the gradient is propagated using a reparameterization trick [34,45], which allows stochastic gradient ascent [46] to be used to optimize the ELBO equation.

The evidence lower confidence bound (ELCBO) is [24]:

$$ELCBO(\boldsymbol{\phi}, f) = \mathrm{E}_{f|\Xi}\left[\mathrm{E}_{\phi}[f]\right] + \mathcal{H}[q(\boldsymbol{\theta})] - \beta_{LCB}\sqrt{\mathrm{V}_{f|\Xi}\left[\mathrm{E}_{\phi}[f]\right]} \tag{23}$$

where the term β_{LCB} represents a risk-sensitivity term.

The ELCBO is the probabilistic lower bound of the ELBO, and it can be used to judge the variational approximation's improvement. As $\mathrm{V}_{f|\Xi}\left[\mathrm{E}_{\phi}[f]\right]$ in eq.(23) decreases, the ELCBO value will converge to the ELBO value [24].

The first two terms of the ELCBO equation are estimated as described before. The risk sensitivity term [24] is usually set to $\beta_{LCB} = 3$.

3.2.4 Update of variational parameters

The variational posterior is parameterized in terms of the variational parameters in the vector $\boldsymbol{\phi}$. These variational parameters are updated by solving an optimization problem [24,25]:

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} \left\{ ELBO(\boldsymbol{\phi}, f) \right\} \tag{24}$$

This optimization is carried out using a stochastic gradient descent algorithm based on a variant of the Adam optimization algorithm [47] to obtain the updated variational posterior.

3.3 Adaptive K components selection

A warm-up stage is used in the initial iterations of the algorithm. In this phase, the variational posterior is specified in terms of a $K = 2$ Gaussian mixture with $w_1 \equiv w_2 = 0.5$. The warm-up phase finishes when the improvement of the ELCBO for three consecutive iterations is smaller than 1, this implies that the variational solution is stabilizing.

An adequate number of components K should be used to capture the true posterior [22]. The number of components in the Gaussian mixture used for the approximation is adaptively chosen as described in [24]. For this purpose, a component can be added to or removed from the Gaussian mixture. A component is added to the Gaussian mixture after the ELCBO of the current iteration is greater than the ELCBO found in the last four iterations. This is done as long as during the last iteration no mixture component was removed. An additional condition can also be set to speed up the approximation as explained in [24,25]. A component of the Gaussian mixture can also be randomly removed from the variational solution, if it simultaneously occurs that the mixture weight is smaller than 0.01, and the difference between the ELCBO of the variational solution in this iteration, and the ELCBO of the variational solution after removal of that component, is smaller than 0.01. More information can be found in [24,25].

3.4 Active sampling

Active sampling is employed to select a number of prescribed samples within each iteration at locations which maximize an acquisition function.

These samples are the input parameters for which the physics-based model is evaluated. The acquisition function a_{pro}, is chosen in such a manner that sampling is encouraged at high probability regions of the log unnormalized posterior [24,25]:

$$a_{pro}(\boldsymbol{\theta}) = s_{\Xi}^2(\boldsymbol{\theta}) \exp\left(f_{\Xi}(\boldsymbol{\theta})\right) q_{\phi}(\boldsymbol{\theta}) \tag{25}$$

where $s_{\Xi}^2(\boldsymbol{\theta})$ is the variance of the GP posterior, $\exp\left(f_{\Xi}(\boldsymbol{\theta})\right)$ is the exponentiated GP posterior

mean for a given training set Ξ, and $q_\phi(\boldsymbol{\theta})$ is the variational approximation of the posterior at $\boldsymbol{\theta}$.

Therefore, this acquisition function favours mostly exploitation of the knowledge obtained in the previous iterations. To add an exploration phase, cyclical annealing is also introduced. This enables the algorithm to explore areas of high probability density that may otherwise have not been found. As shown in the first and second numerical example, cyclical annealing allows sampling of multi-modal regions as an exploration phase occurs in the initial iterations of each annealing cycle. Due to the exploitation nature of the acquisition function, the last iterations of each annealing cycle sample at the already found modes.

3.5 Variational whitening

To deal with highly correlated posterior distributions, variational whitening is introduced in the cyclical VBMC algorithm. This is carried out using a linear transformation of the inference space to a new space where the covariance matrix results in a unit diagonal matrix [25]. The transformation matrix W (rotation and scaling) is obtained using singular value decomposition (SVD) of the covariance matrix of the variational posterior $q(\boldsymbol{\theta})$. Variational whitening occurs after the reliability index $\rho(t)$ described in the next subsection is lower and or equal to 3. It is applied in increasingly spaced intervals within iterations as illustrated in [25].

3.6 Stopping criteria

To determine the number of required iterations, the algorithm uses a reliability index $\rho(t) \geq 0$, that suggests the stability of the variational solution. The algorithm is finished if the value $\rho(t) \leq 1$ is found at the end of $n_{stable} = 8$ consecutive iterations, where a maximum of one intermediate iteration may be unstable, or if a predetermined number of iterations n_{max} is reached.

The value at iteration t of the reliability index is calculated as the average of the three reliability features $\rho_j(t)$ for $j = 1, 2, 3$:

$$\rho(t) = \frac{\rho_1(t) + \rho_2(t) + \rho_3(t)}{3} \tag{26}$$

The value of the reliability index $\rho_1(t)$ is calculated as a function of the KL divergence between the previous and the current variational posterior. The reliability index $\rho_2(t)$ is a function of the change of ELBO between two consecutive iterations, and $\rho_3(t)$ is a function of the estimation of the variance of the ELBO. These indices give an overall measure of how the variational posterior is converging throughout the iterations and are defined as [24,25]:

$$\rho_1(t) = \frac{\left|\mathrm{E}\left[ELBO(t)\right] - \mathrm{E}\left[ELBO(t-1)\right]\right|}{\Delta_{SD}} \tag{27}$$

$$\rho_2(t) = \frac{\sqrt{\mathrm{V}\left[ELBO(t)\right]}}{\Delta_{SD}} \tag{28}$$

$$\rho_3(t) = \frac{KL(q_t \| q_{t-1}) + KL(q_{t-1} \| q_t)}{2\Delta_{KL}} \tag{29}$$

The parameters Δ_{KL} and Δ_{SD} should be chosen in such a manner that the values of the individual reliability features meet the inequality $\rho_j \lesssim 1$, where $j = 1,2,3$, for the values of $\rho(t)$ considered representative of good solutions. In the cyclical VBMC algorithm, the values of Δ_{SD} and Δ_{KL} are respectively set at 0.1 and $0.01\sqrt{d}$.

3.7 Steps of the approach

The proposed approach can be summarised in the following steps:

1. Initialization of Algorithm (fig. 1):
    a. Initial training set for physics-based simulation is run.
    b. Cyclical Annealing is introduced (eq.(9)).
    c. Logarithm of (annealed) unnormalized posterior of the initial set is calculated.

d. GP surrogate model of the (annealed) logarithm unnormalized is built using the initial training set values calculated in 1.b.
e. ELBO and ELCBO are calculated (eq.(19) and eq.(23)).

2. Second part of the algorithm (fig. 2):
    a. Selection of new samples using an acquisition function (eq.(25)), these are used to actively update the GP surrogate model.
    b. Cyclical Annealing is introduced (eq.(9)).
    c. The GP surrogate model of the (annealed) logarithm unnormalized posterior is built.
    d. Calculation of ELBO and ELCBO value (eq.(19)) and eq.(23)).
    e. Update the variational parameters (variational whitening may also be applied at this step).
    f. Check if stopping criteria have been met, if not repeat from step 2.a.

Where the main outputs of the algorithm are the variational posterior, the expected value of the ELBO, and the variance of the ELBO.

4 Results obtained from the case studies investigated

This section has the purpose of comparing the proposed cyclical VBMC algorithm with the standard VBMC algorithm, the monotonic VBMC algorithm and the Transitional Ensemble Markov Chain Monte Carlo (TEMCMC) sampling algorithm. The functions `plotmatrix` and `ksdensity` from MATLAB [48] were used to plot the posterior distributions obtained with the TEMCMC algorithm.

Four case studies have been chosen to showcase the advantages and disadvantages of the cyclical VBMC for different distribution modality: one highly multi-modal (first example), one mildly multi-modal (second example), one unimodal (third example) numerical examples and a multi-modal experimental case study.

The number of function evaluations (number of samples), the number of iterations used to achieve convergence (this will determine the number of function evaluations needed), and the number of

modes found, are used to compare the performance of the algorithms on the multi-modal examples. On the unimodal example, for the same purpose, the samples mean, coefficient of variation, number of function evaluations and number of iterations are used. For all the VBMC implementations, 300,000 samples of the variational posterior are taken to compute the sample mean and sample coefficient of variation.

Ten initial samples are picked using LHS [37] for all the case studies that use a form of the VBMC algorithm, and for every iteration that occurs within the algorithm, five samples are chosen using the acquisition function, and are evaluated. The samples chosen, correspond to evaluations of the physics-based model.

The monotonic annealing schedule used in the monotonic VBMC is calculated for a total number of iterations $T = 40$, with one cycle $M = 1$ and a parameter $S = 0.5$. However, for the cyclical annealing schedule in the cyclical VBMC, the number of cycles is $M = 5$. The monotonic schedule maximum temperature of fifty is subsequently decreased in each iteration until a minimum temperature of one is reached. The same concept is applied to the cyclical annealing schedule that has five cycles where a pre-set maximum temperature of fifty is subsequently decreased in each iteration until a minimum temperature of one is reached in each cycle.

Throughout the examples it will be shown that the monotonic VBMC and cyclical VBMC require a higher amount of samples evaluations compared to the standard VBMC for problems with low dimensionality (i.e., low number of inferred parameters). This is expected as a total number of iterations $T = 40$ is prescribed for both algorithms, meaning that forty is the lowest number of iterations possible.

4.1 Himmelblau multi-modal posterior

A multi-modal problem based on [30,49] is introduced in this subsection. The posterior used (4 peaks, 2-dimensional) can be observed on fig. 3 and it has as its mathematical expression [49] the Himmelblau's function $HB(\theta_1, \theta_2)$:

$$HB(\theta_1,\theta_2)=\left(\theta_1^2+\theta_2-11\right)^2+\left(\theta_1+\theta_2^2-7\right)^2 \tag{30}$$

The $HB(\theta_1,\theta_2)$ is frequently used to assess the performance of optimization algorithms.

It has four distinct solutions of local minima at $\{\theta_1,\theta_2\}_1=\{3,2\}$, $\{\theta_1,\theta_2\}_2=\{-2.805,3.131\}$, $\{\theta_1,\theta_2\}_3=\{-3.779,-3.283\}$ and $\{\theta_1,\theta_2\}_4=\{-3.584,-1.848\}$.

The posterior of interest is then defined as follows [30,49]:

$$p(\boldsymbol{\theta}\,|\,\boldsymbol{y}_{\text{obs}})\propto\exp\left[-HB\left(\theta_1,\theta_2\right)\right] \tag{31}$$

That ensures that the local minima of $HB(\theta_1,\theta_2)$ become regions of high probability density, producing the 4 peaks shown in fig. 3. The likelihood function is modelled as the exponential function of $-HB(\theta_1,\theta_2)$, and thus takes the same mathematical form as the posterior [30,49]. The uniform priors $\theta_1\sim U(-5,5)$ and $\theta_2\sim U(-5,5)$ have been used.

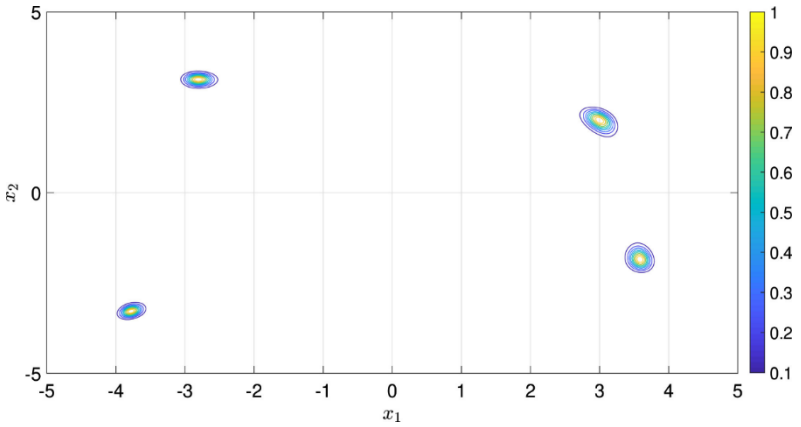


Fig. 3. Contour plot of the Himmelblau's function using eq.(31) taken from [30]. The values of the posterior are given by the numbers on the colour chart.

The results found using the standard VBMC algorithm after running several iterations are illustrated in fig. 4. The final 1-D and 2-D marginal posterior distributions obtained by the standard VBMC algorithm are shown in fig. 5. The figures show that only one mode has been found, as due to the nature of the algorithm, the active sampling used is unable to escape from that mode. In other words,

the algorithm proceeds to only sample in the vicinity of that mode due to its exploitation nature. A total of 75 function evaluations were required, and only one mode was found.

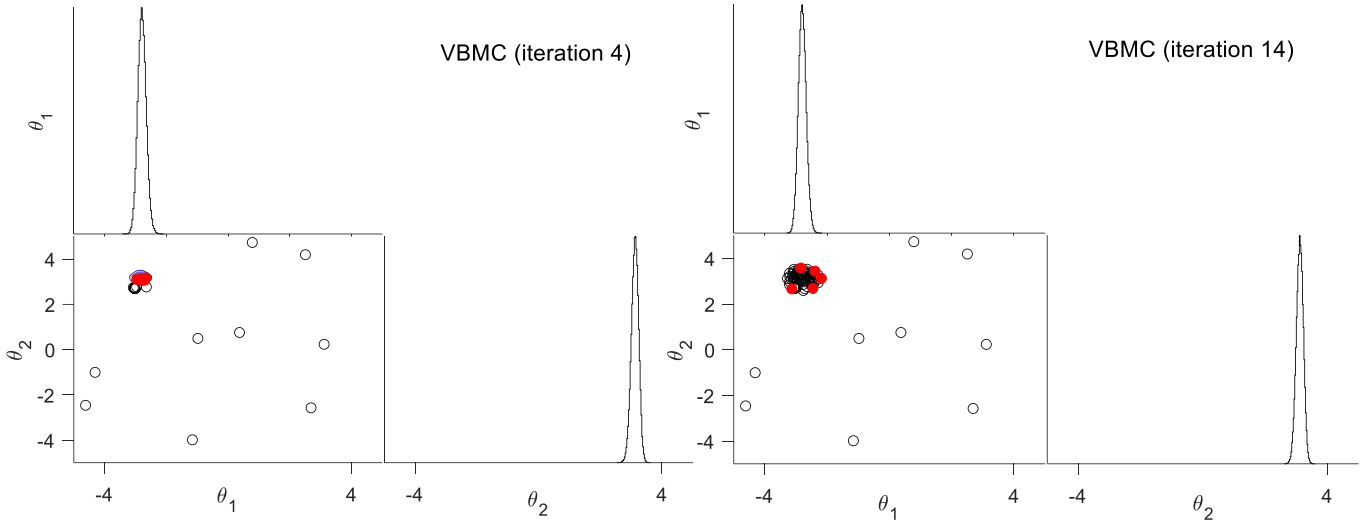


Fig. 4. Resulting 1-D and 2-D marginal posterior distributions from VBMC at 4th and 14th iterations. Red dots indicate samples taken at the current iteration. Black circles indicate samples used for the GP model of the unnormalized posterior at each iteration.

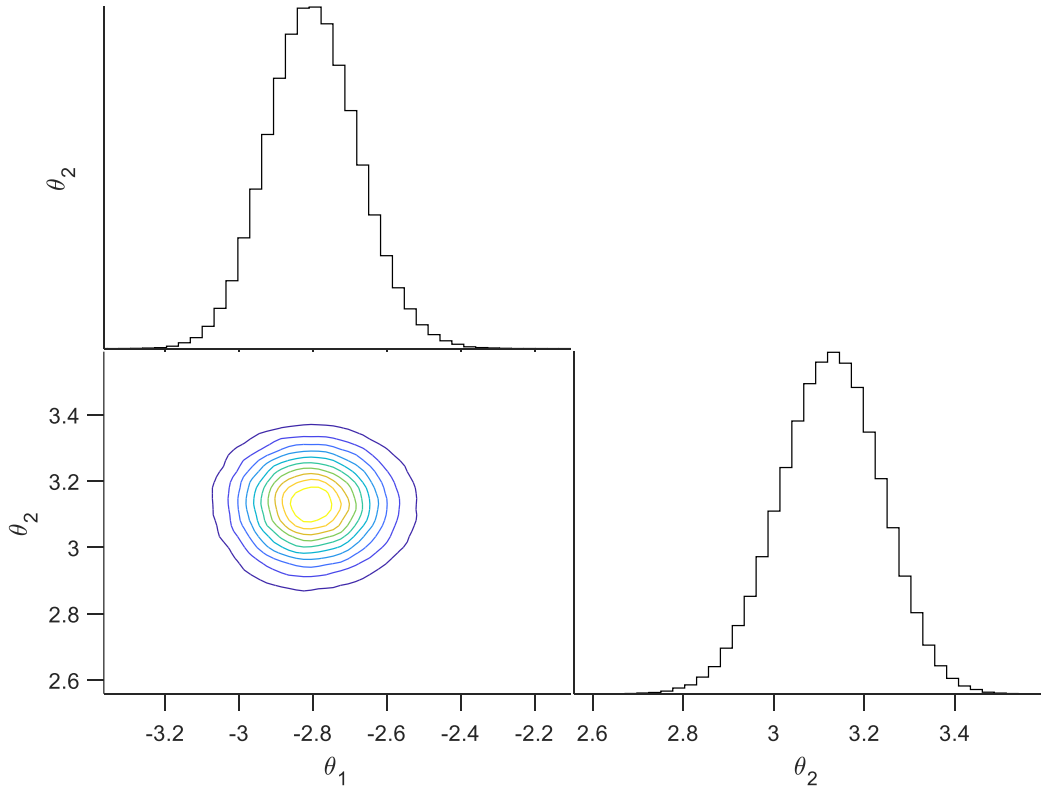


Fig. 5. Final 1-D and 2-D marginal posterior distributions from VBMC algorithm.

The results obtained with the monotonic and the cyclical VBMC algorithm are shown in fig. 6 and fig. 7, it is found that the overall results of these two schedules significantly differ.

It can be seen in fig. 6 that in the monotonic VBMC algorithm, for the first few iterations, the samples are chosen following an exploration approach. In the final iterations of the monotonic algorithm shown in fig. 7, the samples chosen are close to the two modes found. The resulting refined postulated posterior of the algorithm when using a monotonic annealing schedule is only able to account for two modes. The monotonic VBMC algorithm needs a total of 235 function evaluations to converge to the 2-mode estimated posterior shown in fig. 8.

It can be seen in fig. 6, that in the proposed cyclical VBMC algorithm, for the first few iterations, the samples are chosen following both an exploration and an exploitation approach. For the final iterations of the cyclical algorithm, it can be seen in fig. 7, that the samples chosen are close to the four modes found. The resulting refined postulated posterior of the algorithm, using the cyclical annealing schedule shown in fig. 8, is able to account for all four modes. A total of 300 function evaluations were needed to converge to the estimated posterior.

The empirical cumulative density functions (ECDFs), shown in fig. 9 were calculated using samples obtained from both the cyclical VBMC, and the TEMCMC algorithm, applying the function `cdfplot` from Matlab [48]. For the TEMCMC algorithm, the number of samples used to calculate the ECDFs were progressively increased until convergence occurred. It was found that the converged ECDFs obtained by the TEMCMC algorithm, required a much larger number of samples than the cyclical VBMC algorithm.

The computational cost is significantly reduced for the three VBMC algorithms compared to the TEMCMC sampling algorithm, where 5000 evaluations of the likelihood function were needed as shown in [30]. The proposed cyclical VBMC algorithm is the only VBMC variant that is able to find the four modes of the Himmelblau posterior. The numerical results for the Himmelblau multi-modal posterior are summarised in Table 2.

Table 2: Comparison of numerical results for the Himmelblau multi-modal posterior

Method	N. of samples	N. of Total Iterations for Convergence	N. of modes found
VBMC	75	14	1
Monotonic VBMC	235	46	2
Cyclical VBMC	300	59	4
TEMCMC [37]	5000	5	4

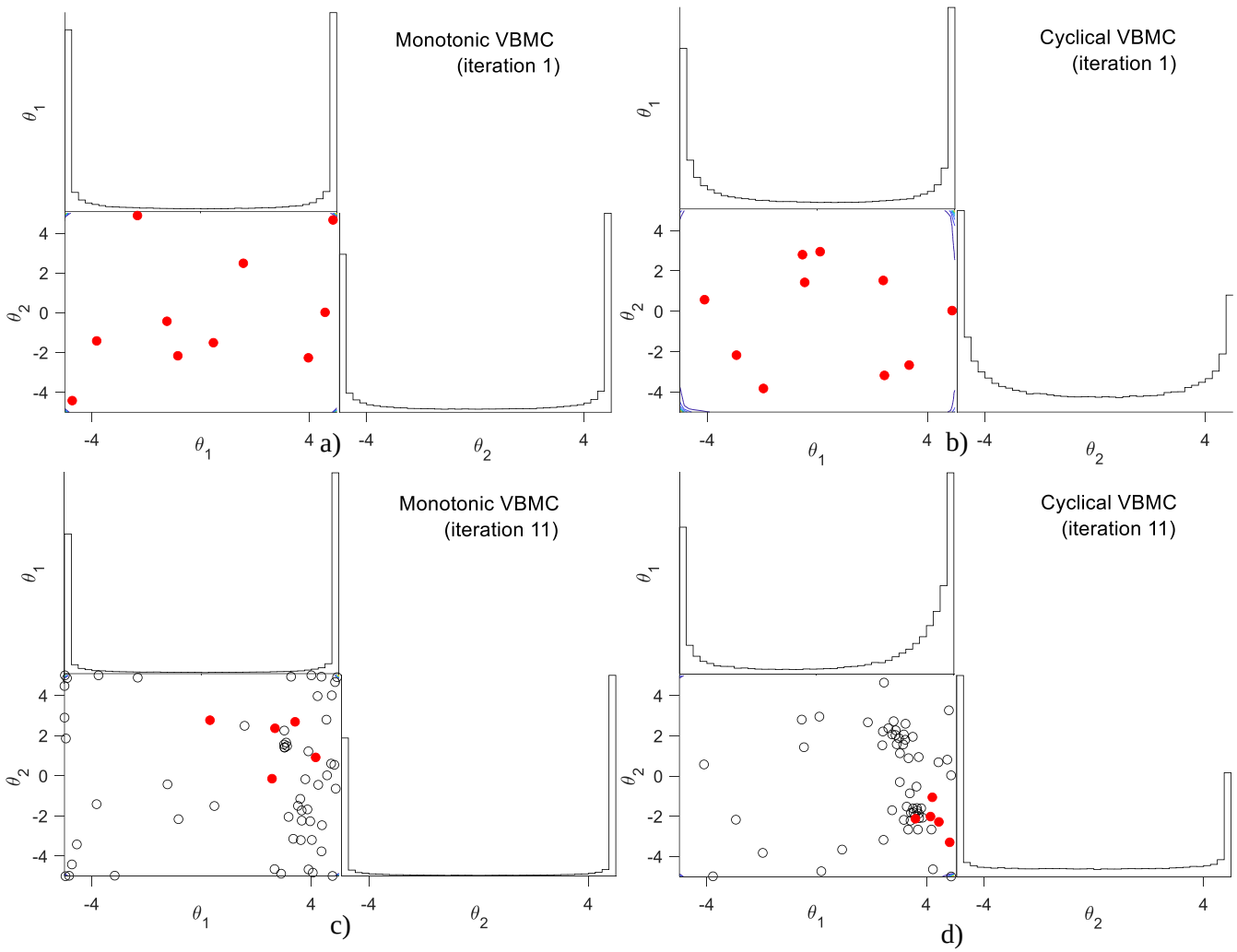


Fig. 6. Resulting 1-D and 2-D marginal posterior distributions from VBMC at 1st and 11th iterations. Red dots indicate samples taken at the current iteration. Black circles indicate samples used for the GP model of the unnormalized posterior at each iteration. a) Monotonic annealing schedule at 1st iteration; b) Cyclical Annealing Schedule at 1st iteration; c) Monotonic annealing schedule at 11th iteration; d) Cyclical Annealing Schedule at 11th iteration.

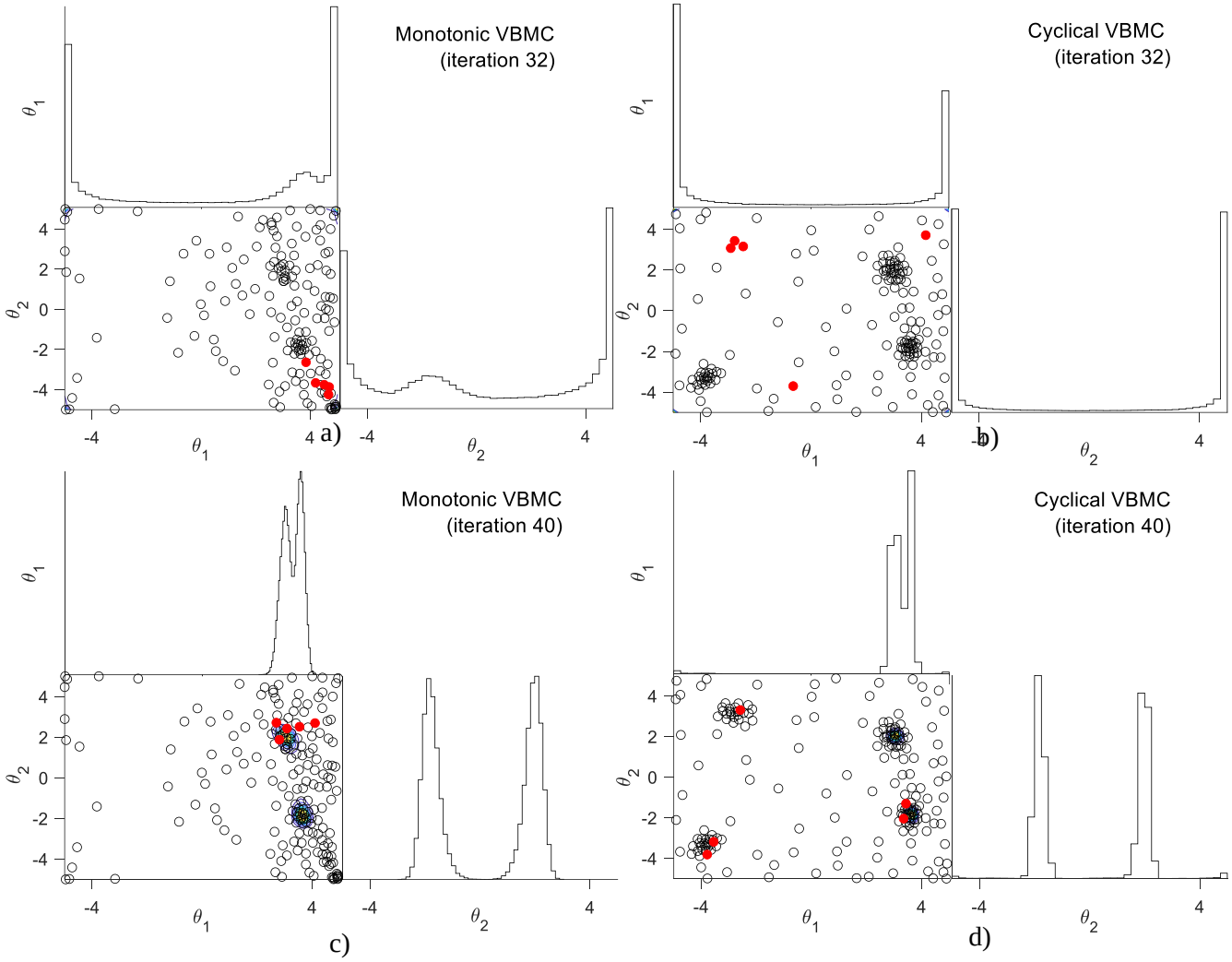


Fig. 7. Resulting 1-D and 2-D marginal posterior distributions from VBMC at 32nd and 40th iterations. Red dots indicate samples taken at the current iteration. Black circles indicate samples used for the GP model of the unnormalized posterior at each iteration. a) Monotonic annealing schedule at 32nd iteration; b) Cyclical Annealing Schedule at 32nd iteration; c) Monotonic annealing schedule at 40th iteration; d) Cyclical Annealing Schedule at 40th iteration.

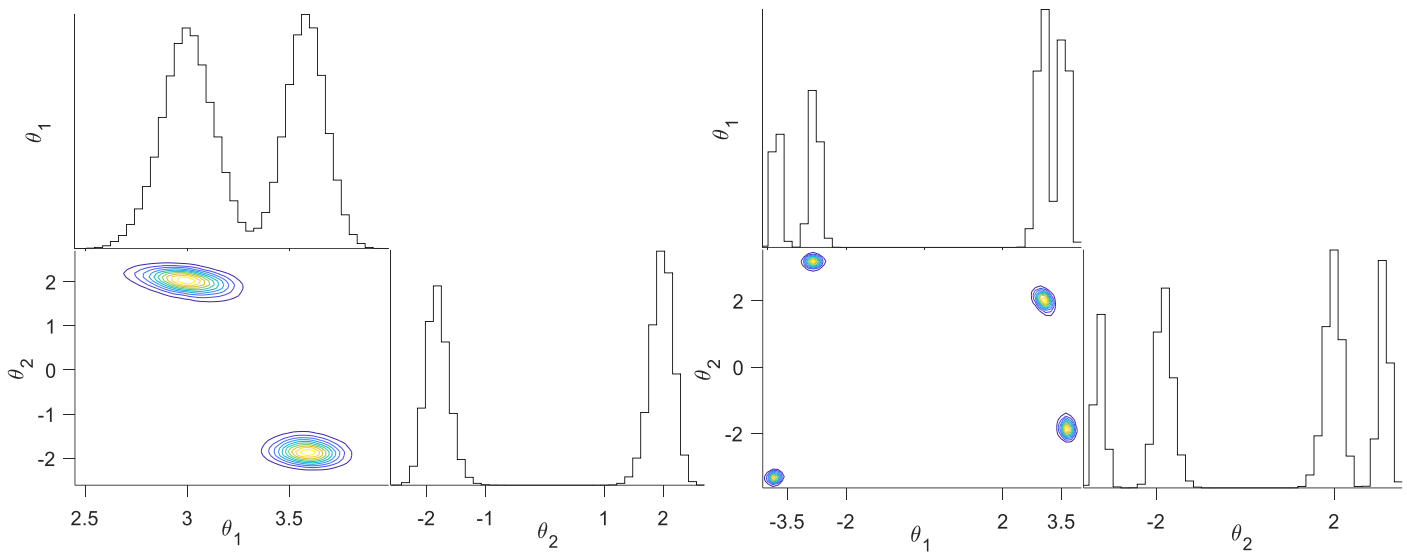


Fig. 8. Final 1-D and 2-D marginal posterior distributions from VBMC algorithm. a) Monotonic annealing schedule; b) Cyclical Annealing Schedule.

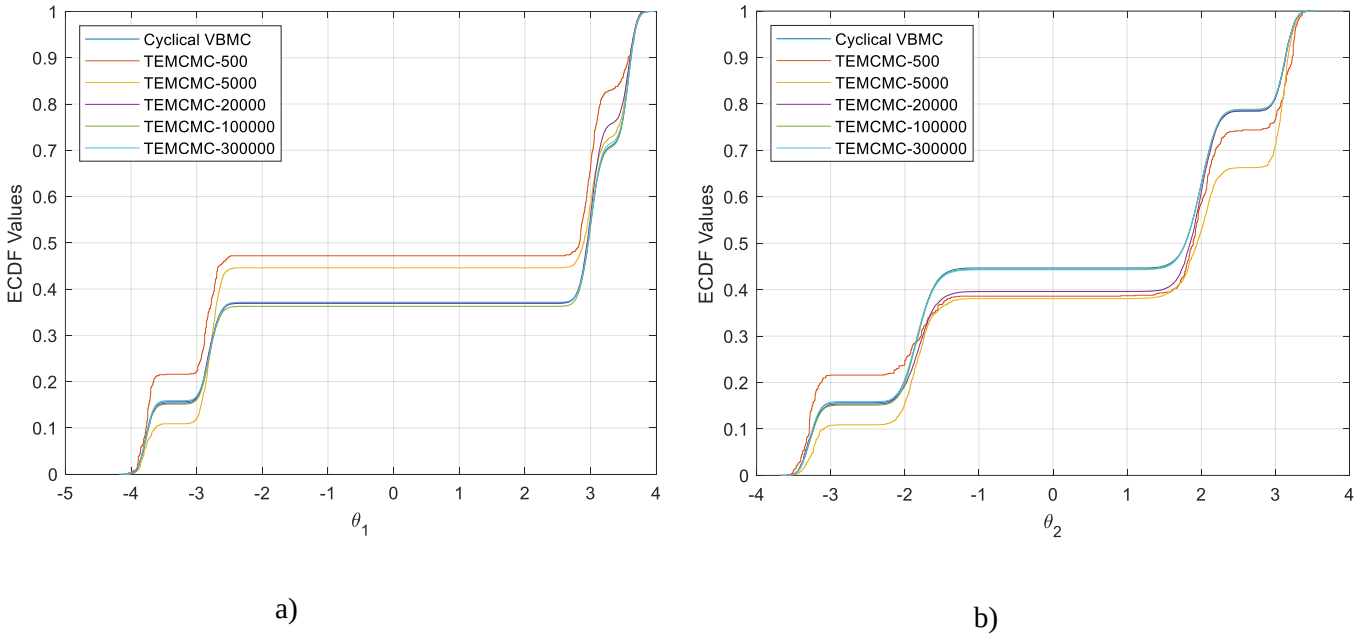


Fig. 9. Marginal ECDFs using Cyclical VBMC and TEMCMC. a) Parameter θ_1; b) Parameter θ_2.

4.2 Mass-spring system (multi-modal posterior)

In this example taken from [50], a 2-dimensional multi-modal Bayesian model updating system that may be found in engineering problems, is used to compare the differences in the performance of the aforementioned algorithms. For the purposes of this subsection, the numerical performance will be based on the number of samples used to compute the posterior, the number of modes found in each algorithm and the empirical cumulative density function.

As shown in fig. 10, a 2-degrees of freedom (2-DoF) system with masses $m_1 = 16.531 \times 10^3\, kg$, $m_2 = 16.131 \times 10^3\, kg$ joined by springs with stiffness $k_1 = \bar{k}\theta_1, k_2 = \bar{k}\theta_2$, where $\bar{k} = 29.7 \times 10^6\, N/m$ is defined, and θ_1 and θ_2 are the uncertain parameters to be inferred. For the spring constants, the uniform priors $\theta_1 \sim U(0.01, 3)$ and $\theta_2 \sim U(0.01, 3)$ have been used.

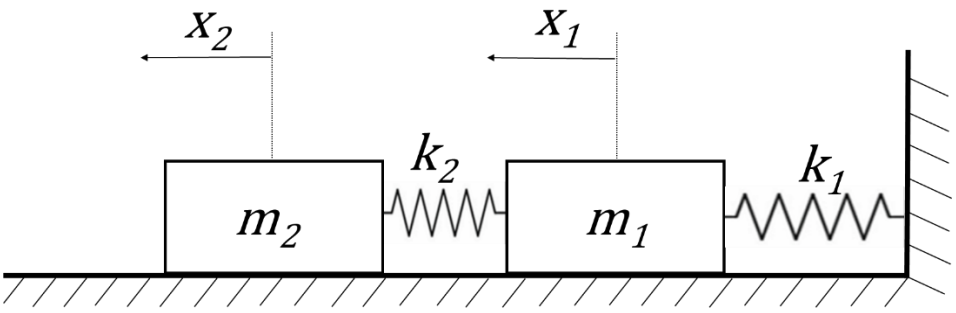


Fig. 10. First mass-spring system.

Two independent likelihood functions are used with standard deviations $\sigma_i = 0.02\omega_i$ (2% of the deterministic values of the natural frequencies) and with means that equal the deterministic values of the natural frequencies.

From fig. 11, it can be observed that the standard VBMC is only able to find one mode, and that the active sampling is again unable to escape from that mode. The standard algorithm uses a total of 75 function evaluations to obtain the resulting posterior distribution that only accounts for a single mode. The final 1-D and 2-D marginal posterior distributions obtained by the standard VBMC algorithm are shown in fig. 12.

For the monotonic VBMC algorithm, an exploration phase is shown in the first few iterations. An exploitation phase that samples in the vicinity of the two found modes is illustrated in fig. 13.

However, for the cyclical VBMC algorithm both exploration and exploitation occur in the early iterations, as shown in fig. 13. The final 1-D and 2-D marginal posterior distributions obtained by the monotonic and cyclical VBMC algorithm are shown in fig. 14, where it is possible to observe that both methods account for the two modes.

In fig. 16 the ECDFs for the cyclical VBMC, monotonic VBMC, and TEMCMC are shown. The resulting ECDFs are found to be similar, with a slight difference observed for the ECDFs obtained from the TEMCMC algorithm.

Table 3 summarises the number of evaluations and iterations needed for the three analysed VBMC algorithms, and the sampling method TEMCMC (fig. 15). It can be seen that when using the three VBMC algorithms, the computational cost is significantly reduced compared to the TEMCMC sampling algorithm, where 5000 evaluations of the likelihood function were needed to obtain samples from the posterior distribution.

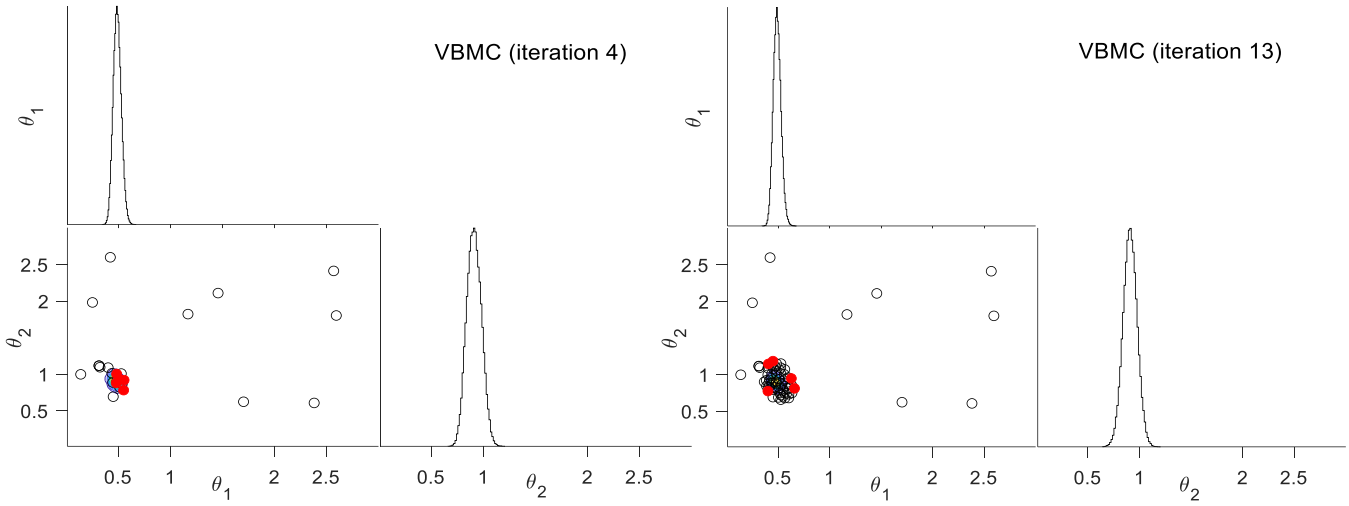


Fig. 11. Resulting 1-D and 2-D marginal posterior distributions from VBMC at 4th and 13th iterations. Red dots indicate samples taken at the current iteration.

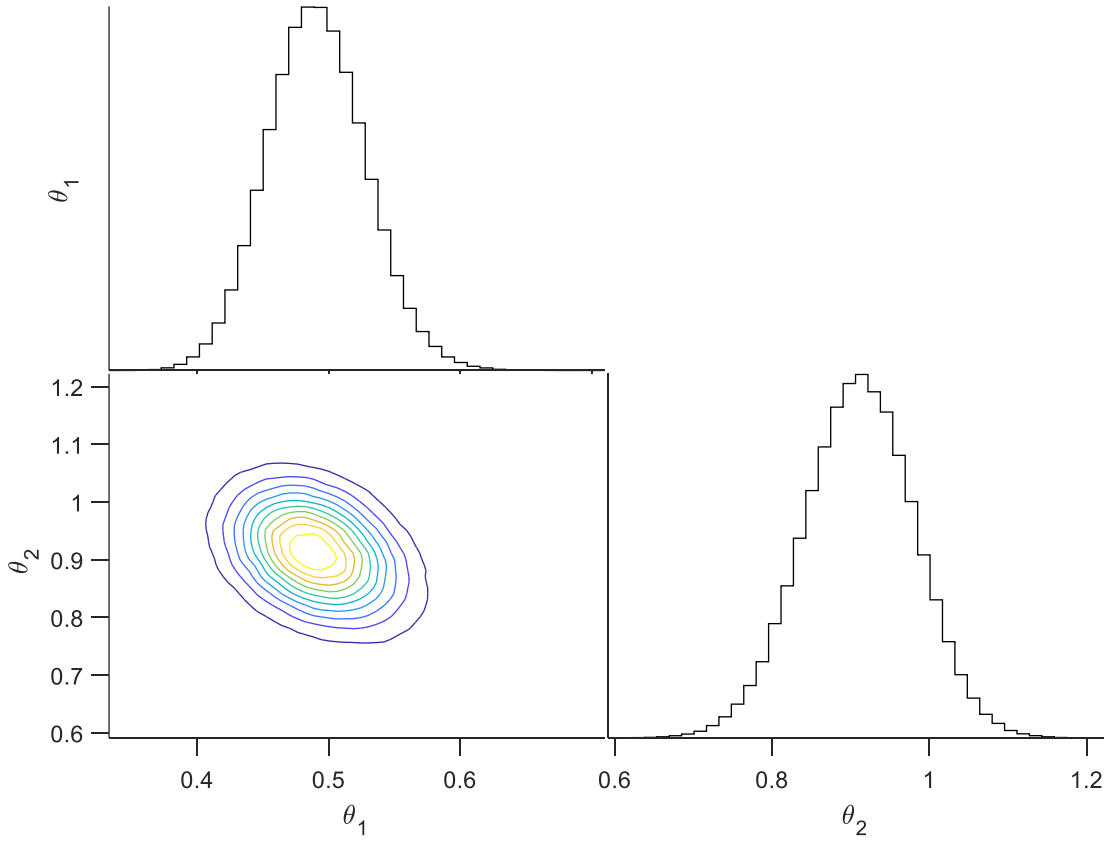


Fig. 12. Final 1-D and 2-D marginal posterior distributions from VBMC algorithm.

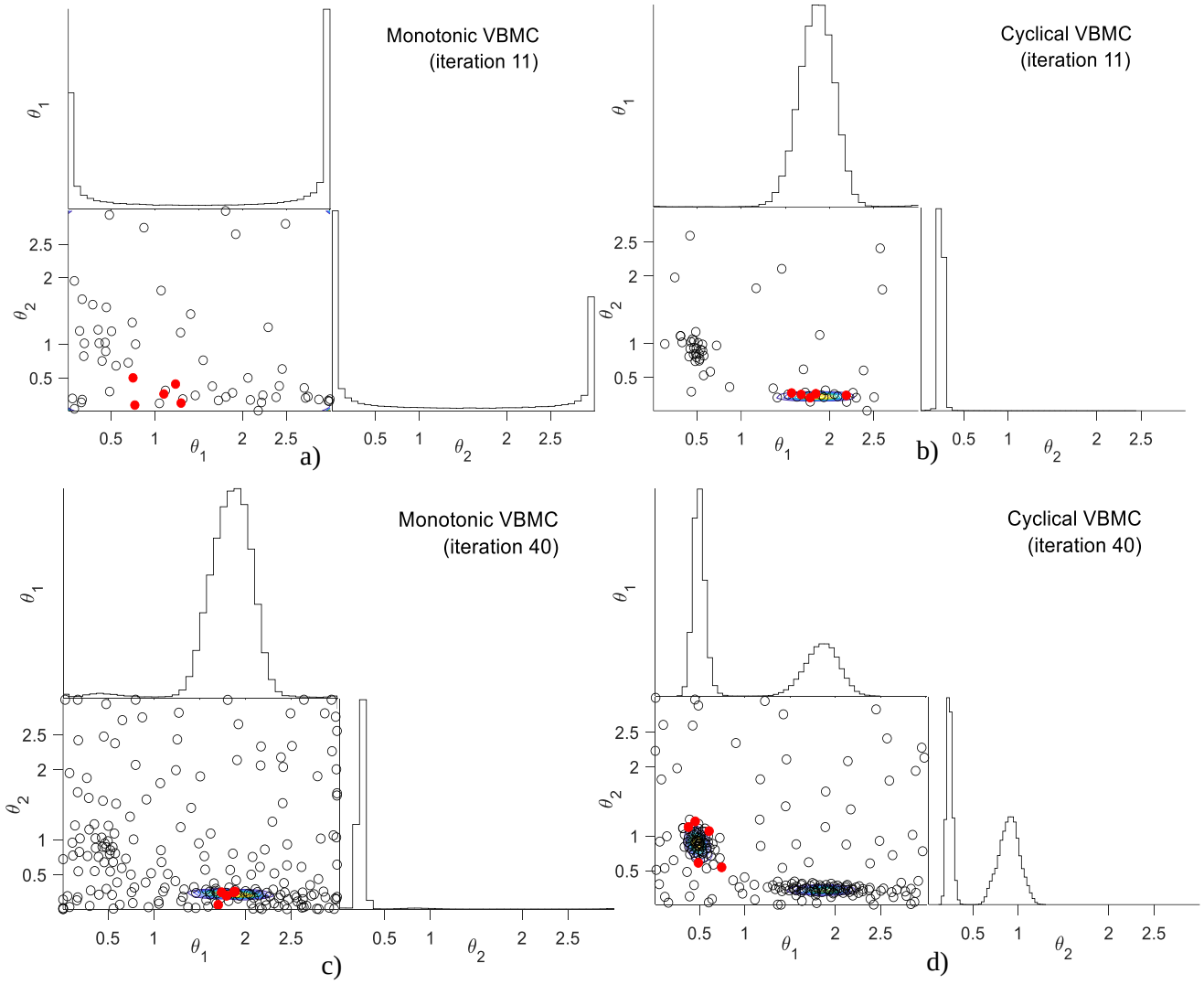


Fig. 13. Resulting 1-D and 2-D marginal posterior distributions from VBMC. Red dots indicate samples taken at the current iteration. Black circles indicate samples used for the GP model of the unnormalized posterior at each iteration. a) Monotonic annealing schedule at 11th iteration; b) Cyclical Annealing Schedule at 11th iteration; c) Monotonic annealing schedule at 40th iteration; d) Cyclical Annealing Schedule at 40th iteration.

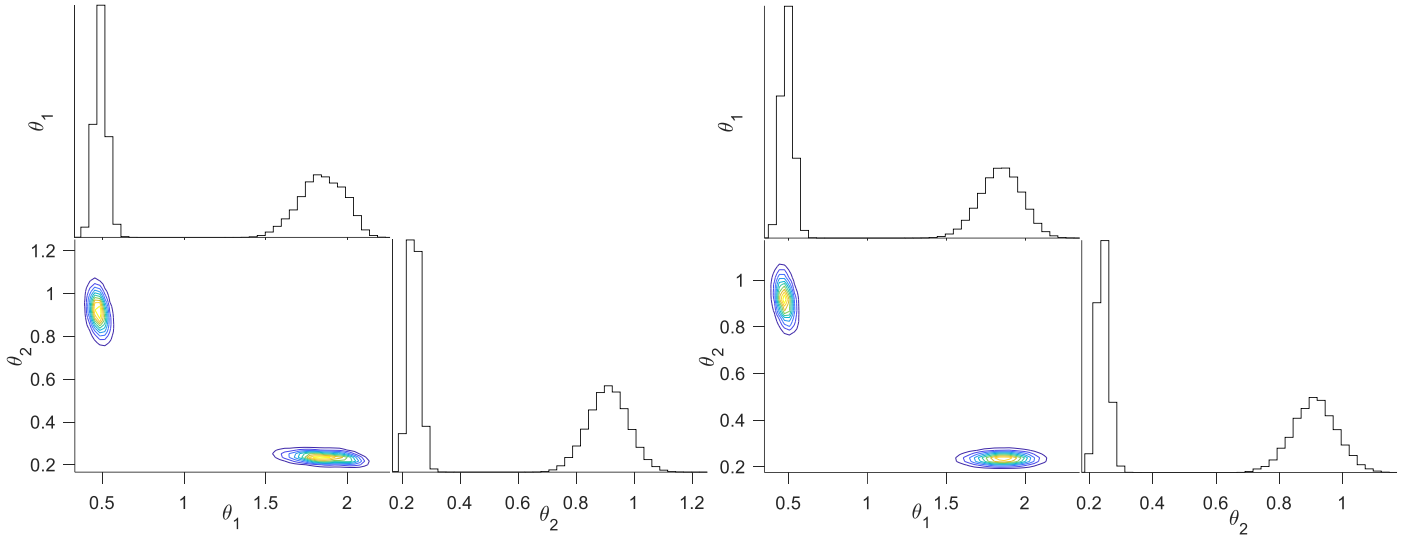


Fig. 14. Final 1-D and 2-D marginal posterior distributions from VBMC algorithm. a) Monotonic annealing schedule; b) Cyclical Annealing Schedule.

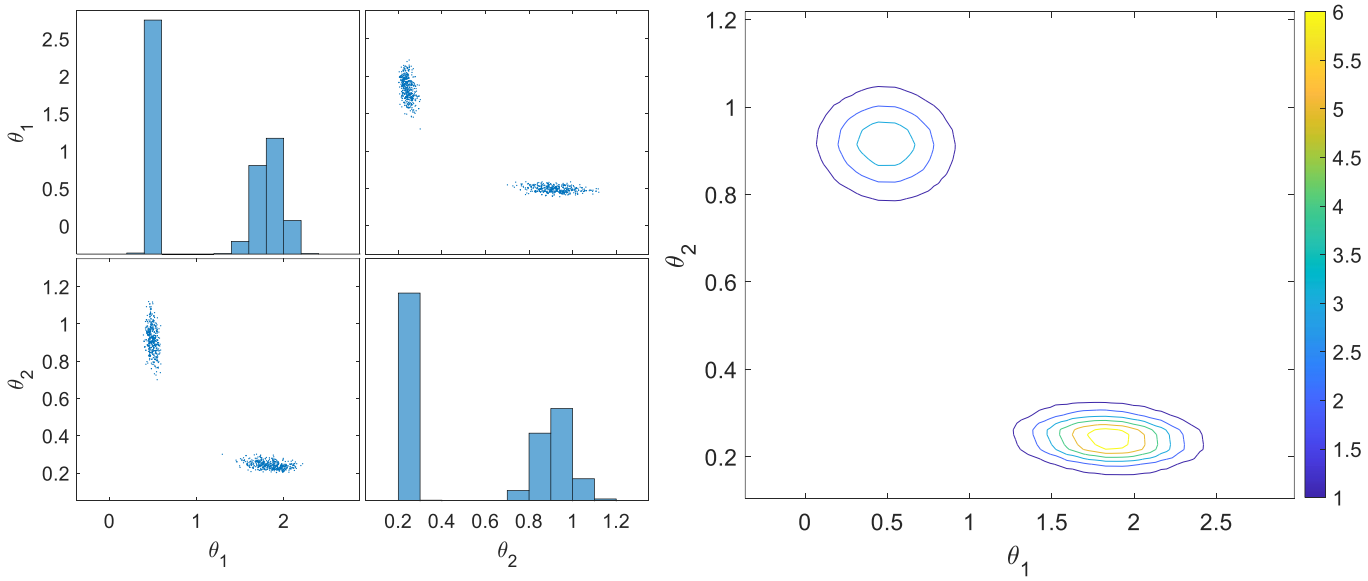


Fig. 15. Resulting plots from TEMCMC algorithm. a) Scatterplot; b) 2-D Posterior distribution. The values of the 2-D posterior are given by the numbers on the colour chart.

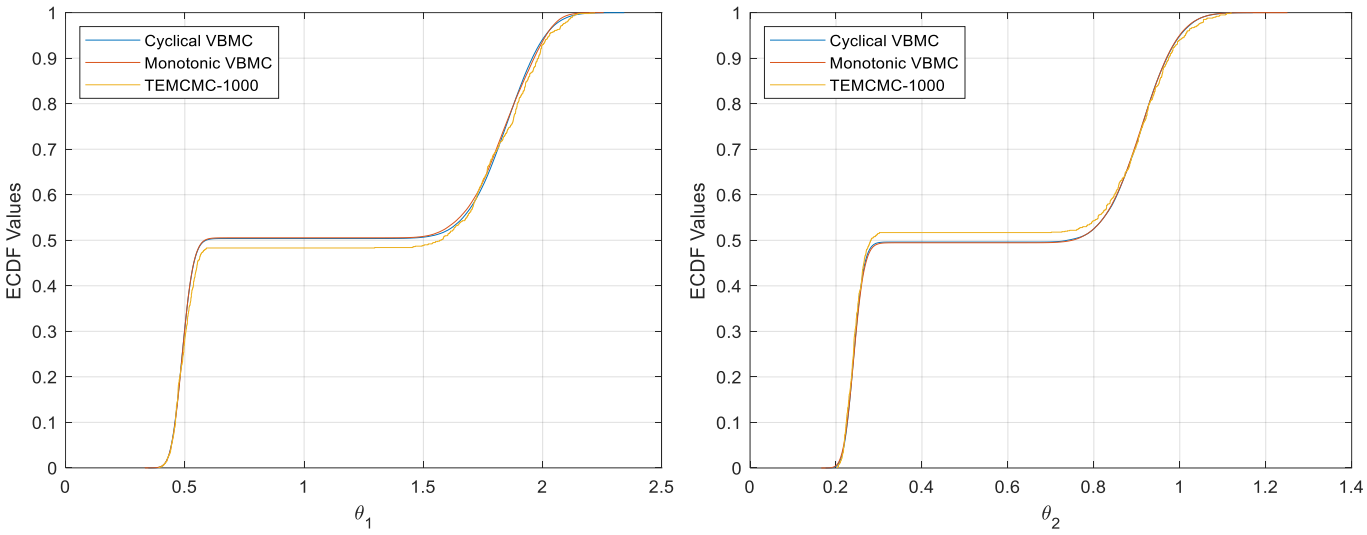


Fig. 16. Marginal ECDFs using Cyclical VBMC, Monotonic VBMC and TEMCMC.

Table 3: Comparison of numerical results for the mass-spring (multi-modal) system.

Method	N. of samples	N. of Total Iterations for Convergence	N. of modes found
VBMC	70	13	1
Monotonic VBMC	275	54	2
Cyclical VBMC	260	51	2
TEMCMC	5000	5	2

4.3 Mass-spring system (unimodal posterior)

In this example taken from [30], for a 4-dimensional Bayesian model updating system, the two Degrees-of-Freedom (DoF) system shown in fig. 17 is used to compare the performances of all algorithms.

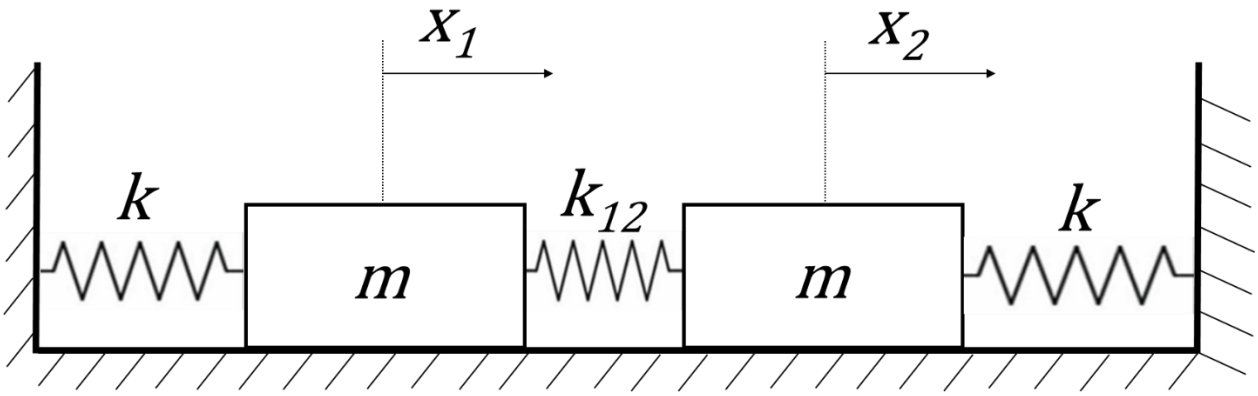


Fig. 17. Second mass-spring system.

The 2-DoF system has equal masses $m = 0.5\,kg$ attached to edge springs with stiffness $k = 0.6\,N/m$, and the stiffness of the spring between the two masses is $k_{12} = 1\,N/m$.

The natural frequencies that correspond to the specified properties of the two-degree-of freedom system are given by [30]:

$$\hat{\omega}_1 = \sqrt{\frac{k}{m}} \tag{32}$$

$$\hat{\omega}_2 = \sqrt{\frac{k + 2k_{12}}{m}} \tag{33}$$

The values of the natural frequencies $\hat{\omega}_1$ and $\hat{\omega}_2$ are corrupted with noise as shown below [30]:

$$\omega_1 = \hat{\omega}_1 + \varepsilon_1 \tag{34}$$

$$\omega_2 = \hat{\omega}_2 + \varepsilon_2 \tag{35}$$

Where ε_1 and ε_2 are the noise terms, that follow Gaussian statistical distributions. The mean of both

Gaussian distributions is $0\,Hz$, and their standard deviations are respectively $\sigma_1 = 0.1\hat{\omega}_1 = 0.110\,Hz$ and $\sigma_2 = 0.1\hat{\omega}_2 = 0.228\,Hz$. The likelihood function is then given by the following equation [30]:

$$P\left(y_{obs}\,|\,\theta\right) = \prod_{n=1}^{15} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{\left(\omega_{1,n} - \hat{\omega}_1\right)}{2\sigma_1^2} - \frac{\left(\omega_{2,n} - \hat{\omega}_{2,n}\right)}{2\sigma_2^2} \right] \tag{36}$$

In this example, the parameters $\{k, k_{12}, \sigma_1, \sigma_2\} \equiv \{\theta_1, \theta_2, \theta_3, \theta_4\}$ are assumed to be unknown. The uniform priors $k \sim U(0.1, 4)$ [N/m] and $k_{12} \sim U(0.1, 4)$ [N/m] have been used for the stiffnesses. The prior uniforms taken for the standard deviations σ_1 and σ_2 are $\sigma_1 \sim U\left(10^{-5}, 1\right)$ [Hz] and $\sigma_2 \sim U\left(10^{-5}, 1\right)$ [Hz]. The posterior probability density function of the parameters is updated using the 'experimental measurements', for this example, the fifteen individual experimental 'measurements' of ω_1 and ω_2 used are found in [30].

The final posterior using 1D and 2D marginal distributions are shown in fig. 18, fig. 19 and fig. 20.

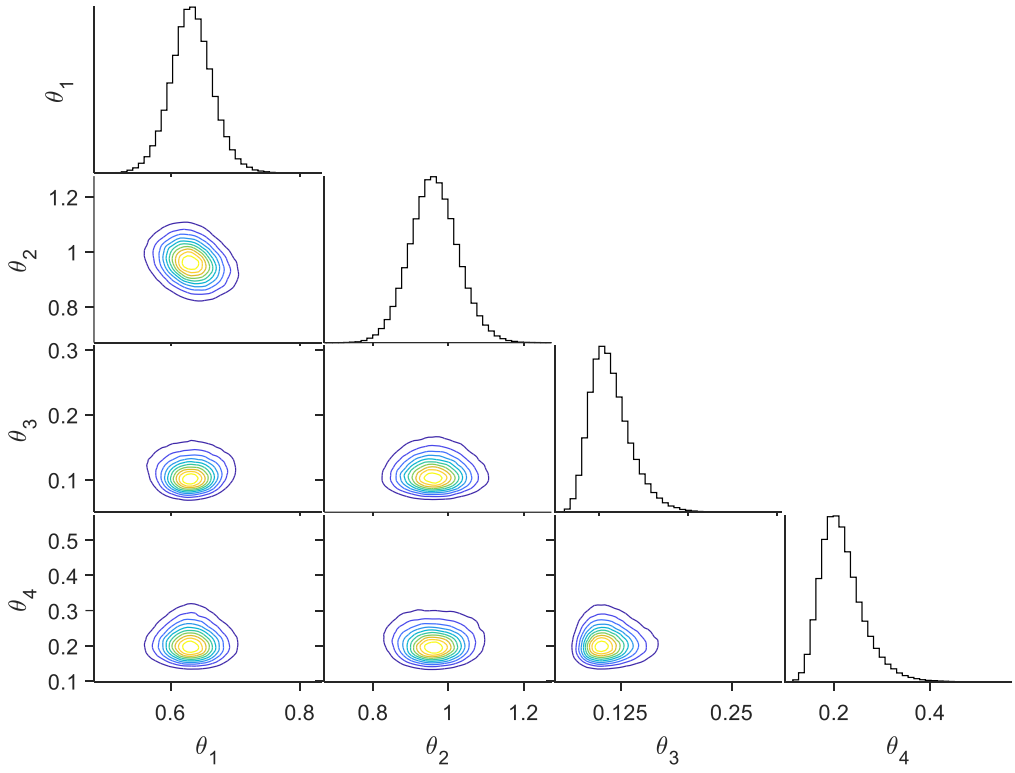


Fig. 18. Final 1-D and 2-D marginal posterior distributions from standard VBMC algorithm.

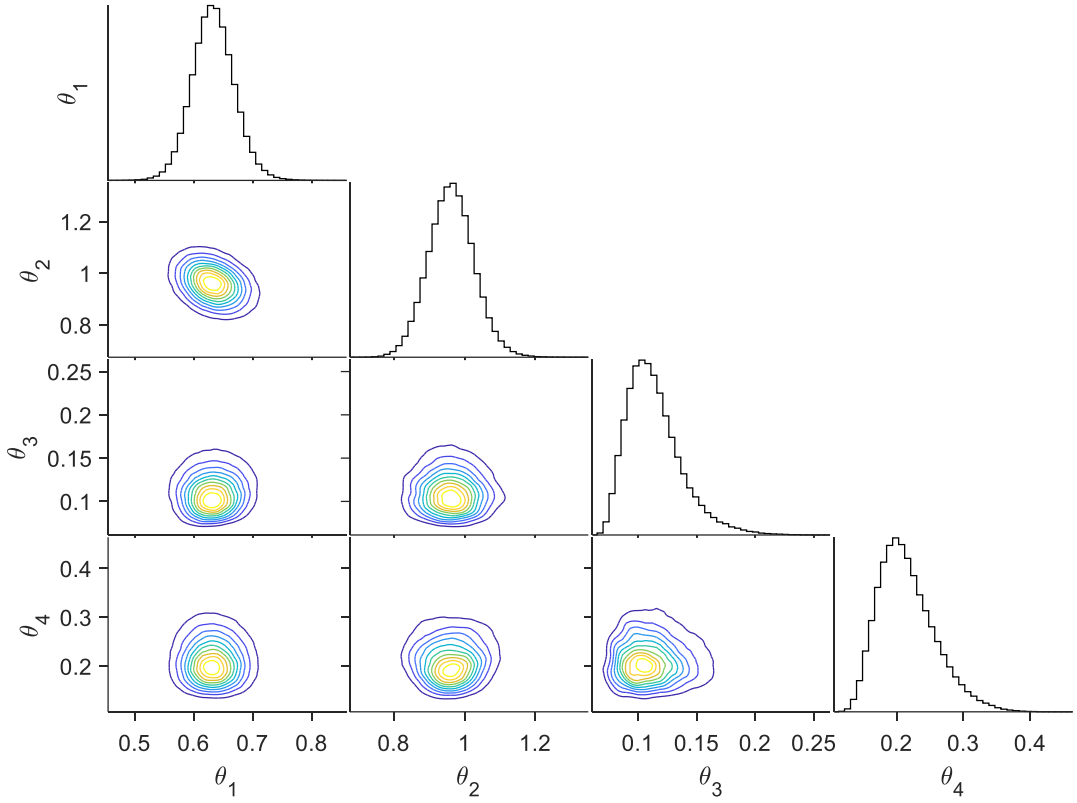


Fig. 19. Final 1-D and 2-D marginal posterior distributions from monotonic VBMC algorithm.

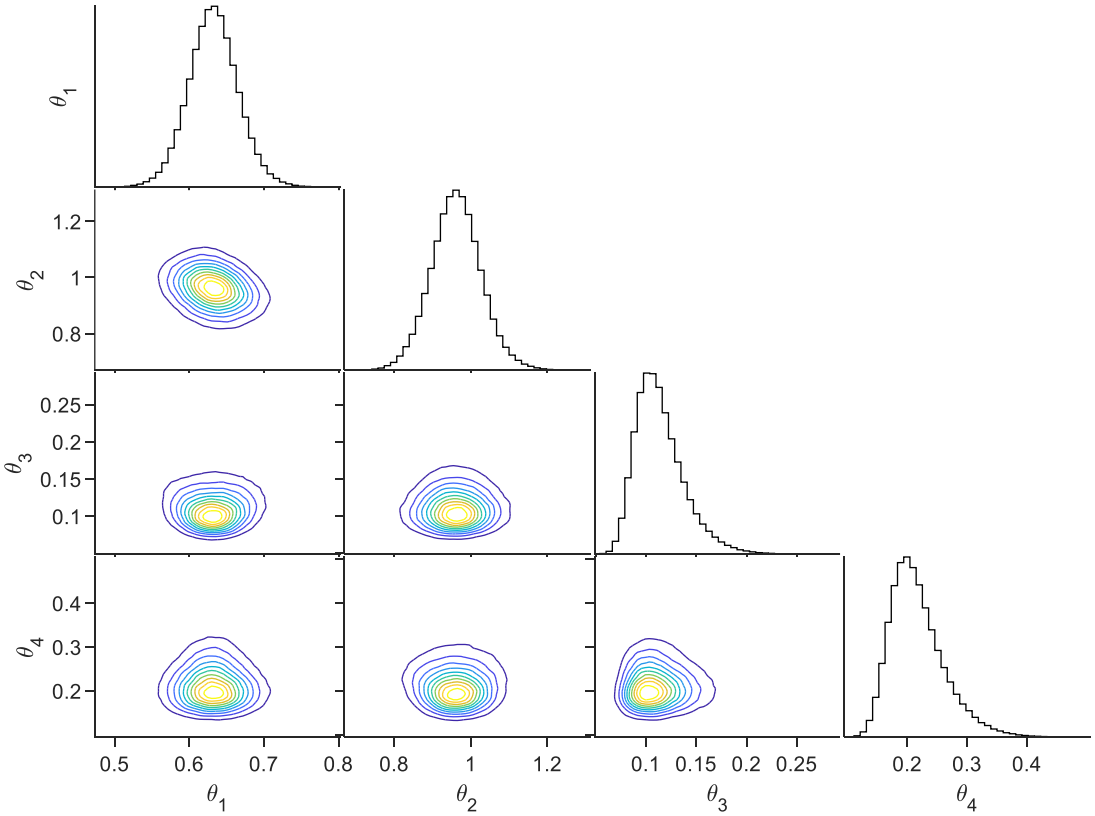


Fig. 20. Final 1-D and 2-D marginal posterior distributions from cyclical VBMC algorithm.

Table 4: Comparison of numerical results for the mass-spring system (unimodal posterior).

Method	Sample Mean	Sample C.O.V [%]	N. of samples	N. of Total Iterations for Convergence
VBMC	$\begin{bmatrix} 0.633\text{N/m} \\ 0.962\text{N/m} \\ 0.114\text{Hz} \\ 0.217\text{Hz} \end{bmatrix}$	$\begin{bmatrix} 5.21 \\ 6.67 \\ 20.66 \\ 20.82 \end{bmatrix}$	220	43
Monotonic VBMC	$\begin{bmatrix} 0.633\text{N/m} \\ 0.963\text{N/m} \\ 0.114\text{Hz} \\ 0.216\text{Hz} \end{bmatrix}$	$\begin{bmatrix} 5.68 \\ 6.73 \\ 20.70 \\ 19.71 \end{bmatrix}$	265	52
Cyclical VBMC	$\begin{bmatrix} 0.632\text{N/m} \\ 0.962\text{N/m} \\ 0.114\text{Hz} \\ 0.217\text{Hz} \end{bmatrix}$	$\begin{bmatrix} 5.34 \\ 6.85 \\ 21.32 \\ 20.85 \end{bmatrix}$	260	51
TEMCMC [30]	$\begin{bmatrix} 0.625\text{N/m} \\ 1.013\text{N/m} \\ 0.121\text{Hz} \\ 0.229\text{Hz} \end{bmatrix}$	$\begin{bmatrix} 5.67 \\ 6.80 \\ 17.25 \\ 26.15 \end{bmatrix}$	5000	5
True Values	$\begin{bmatrix} 0.6\text{N/m} \\ 1\text{N/m} \\ 0.11\text{Hz} \\ 0.228\text{Hz} \end{bmatrix}$	-	-	-

Table 4, for a 4D dimensional problem, shows that a significant lower amount of model evaluations is needed for the three VMBC approaches to obtain similar results to the obtained with the TEMCMC algorithm. The results obtained using the TEMCMC algorithm are taken from [30].

4.4 Experimental validation: aluminium frame problem

An aluminium frame experimental case study presented in [30,51], and schematically shown in fig. 21 is used to compare the four statistical model updating algorithms.

The aluminium frame structure is composed of a total of seven beams. Out of those seven, four beams are vertical (two short internal ones and two long external ones) and three beams are horizontal (all with identical length). Two masses m_1 and m_2 of variable position also form part of the structure. Each of those two masses is attached to one of the two short vertical beams at distances from the bottom of the beam pm_1 and pm_2. The movable masses are used in the structure to induce an effect in the modal properties of the system comparable to the one induced by possible structural damages.

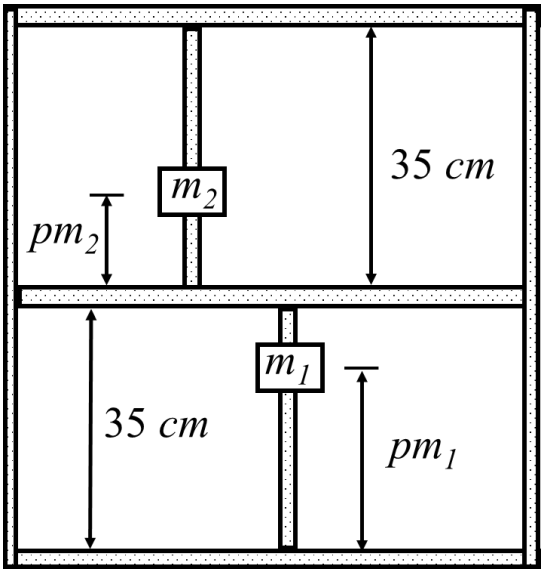


Fig. 21. Schematic representation of aluminium frame with moveable masses (adapted from [30]).

Experimental data was used to obtain the first six natural frequencies of the structure for eleven different combinations $\{pm_1, pm_2\}$. A summary of the results is found on Table 5.

Table 5: Experimental results (data from [52])

Mode		1	2	3	4	5	6
Modal shape		1st in-plane bending mode	1st out-of-plane bending mode	1st torsional mode	2nd in-plane bending mode	2nd out-of-plane bending mode	2nd torsional mode
Exp.	Positions (pm1,pm2) [cm]	Frequencies [Hz]					
		ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
1	(5, 5)	20.11	22.79	47.52	63.96	183.82	283.51
2	(5, 20)	18.72	20.46	46.97	72.24	214.84	296.32
3	(5, 35)	17.72	18.29	46.42	63.45	196.38	278.70
4	(20, 5)	19.40	22.39	46.32	61.78	173.49	259.76
5	(20, 20)	17.91	20.28	45.67	64.73	190.84	284.09
6	(20, 35)	16.71	18.21	45.18	56.53	177.97	264.44
7	(35, 5)	17.71	21.76	44.00	59.48	164.05	254.48
8	(35, 20)	16.91	19.82	43.15	60.06	175.75	279.10
9	(35, 35)	15.95	17.89	42.44	50.66	163.55	257.82
10	(11, 11)	19.58	21.73	47.00	67.54	196.21	285.95
11	(29, 29)	16.65	18.85	43.93	55.43	174.35	284.84

Following the same approach used on [30], a surrogate model of the expensive-to-evaluate Finite Element Model (FEM) of the aluminium frame structure is built. An Artificial Neural Network (ANN) is trained using a set of simulated values of the expensive-to-evaluate FEM from the database [52]. The architecture for the ANN and calibrated model of [30] is also used. The ANN's architecture includes three layers: input, hidden and output with respectively two, ten and six nodes. Further details on the calibration procedure employed can be found in the paper [30].

The set of input data presented in the form of a scatterplot are shown in fig. 22.

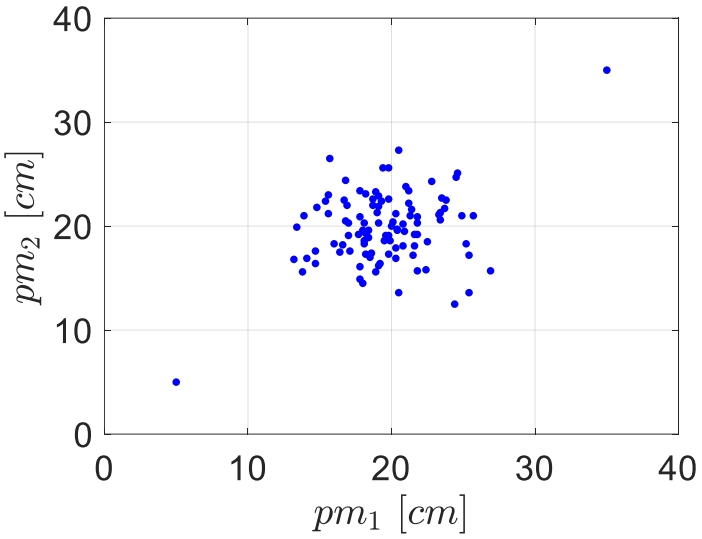


Fig. 22. Set of input data used for calibration of the ANN.

In this example, the parameters $\{pm_1, pm_2, \sigma_1, \ldots, \sigma_6\} \equiv \{\theta_1, \theta_2, \theta_3, \ldots, \theta_8\}$ are assumed to be uncertain and independent from each other. The uniform priors $pm_1 \sim U(5,35)$ [cm] and $pm_2 \sim U(5,35)$ [cm] have been used for the position of the masses. The uniform prior taken for all the measurement 'noises' σ_v corresponding to the natural frequency ω_v (for $v = 1, \ldots, 6$) are $\sigma_v \sim U(10^{-3}, 100)$ [Hz].

The overall likelihood function LF is given by the weighted addition of the three likelihood functions lf_1, lf_2, lf_3 defined below [30,51]:

$$lf_1 = \prod_{v=1}^{6} \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left[-\frac{(\omega_v - M_v)^2}{2\sigma_v^2} \right] \tag{37}$$

$$lf_2 = \prod_{v=1}^{6} 1 - \exp\left[-\frac{1}{(\omega_v - M_v)^2} \right] \tag{38}$$

$$lf_3 = \prod_{v=1}^{6} 1 - \exp\left[-\sqrt{\frac{1}{(\omega_v - M_v)^2}} \right] \tag{39}$$

$$LF = \frac{1}{3}\sum_{p=1}^{3} lf_p \tag{40}$$

Where M_v is the model output for ω_v.

In this example, only the last two experiments (10 and 11) of Table 5 were used as observations in the Bayesian model updating framework. These two experiments were selected to recreate an example where during an observation period, two different states of damage might occur. Therefore, these two different states of damage of the structure will have distinct natural frequencies that correspond to different values of the uncertain physical parameters to be inferred.

Figure 23, fig. 24, fig. 25 and fig. 26 show the final 1-D and 2-D marginal posterior distributions obtained from the algorithms. The modes of the posterior are not found at the positions of the masses on the experiments 10 and 11. This is due to two reasons: using an ANN that had not been trained with enough samples covering the entire parameter space as shown in fig 22, and the fact that the experiments do not bring enough information to make the Bayesian model updating problem globally identifiable. As a result, a highly multi-modal posterior of great interest to test the applicability of the three VBMC algorithms is found.

From fig. 23, it can be observed that the standard VBMC is unable to find one of the modes shown in the marginal posterior of θ_1. However, this missing mode on fig. 23 is clearly seen in the monotonic VBMC and cyclical VBMC of fig. 24 and fig. 25 respectively. It can also be seen in fig. 25 that the cyclical VBMC posterior has the highest degree of similarity to the posterior obtained in fig. 26 by the TEMCMC algorithm. It is also found that the three VBMC approaches face slight difficulties to obtain an approximation to the posterior compared to the previous numerical examples. This is due to the very multi-modal posterior found in this problem, where the greatest difficulty is found in the approximation of the tail behaviour of the true posterior.

In fig. 27 the ECDFs for the cyclical VBMC, monotonic VBMC, VBMC and TEMCMC are shown. The resulting ECDFs are found to be similar to the one obtained by the TEMCMC, with the cyclical

VBMC showing the highest similarity.

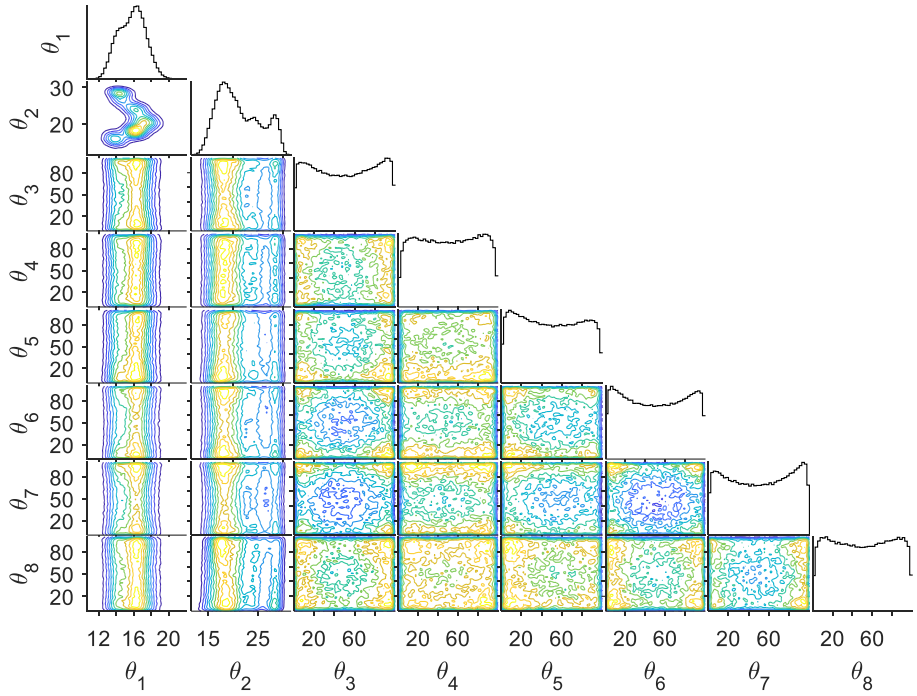


Fig. 23. Final 1-D and 2-D marginal posterior distributions from VBMC algorithm.

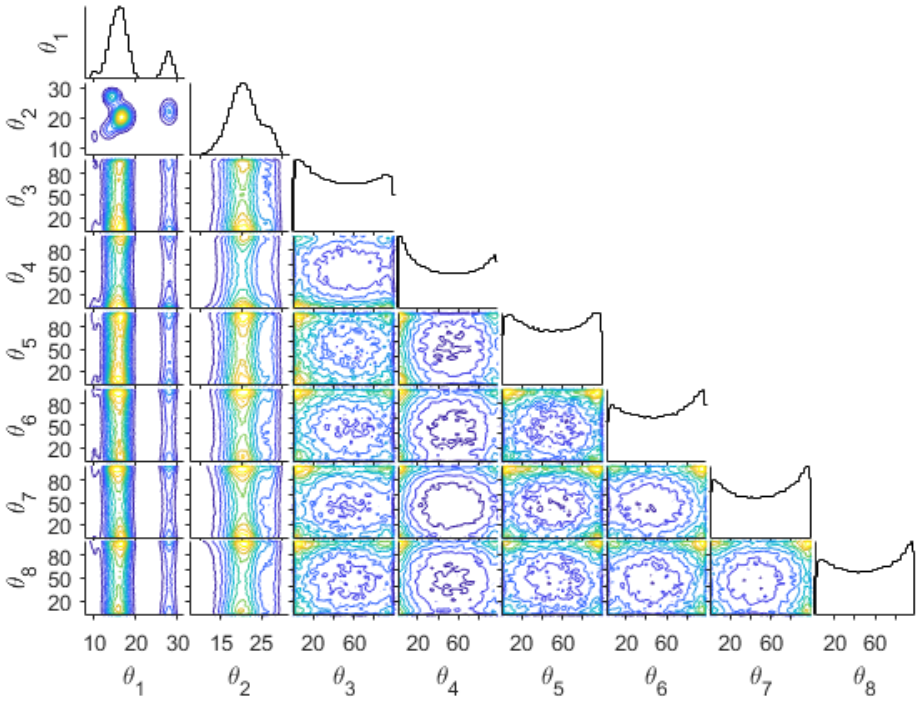


Fig. 24. Final 1-D and 2-D marginal posterior distributions from monotonic VBMC algorithm.

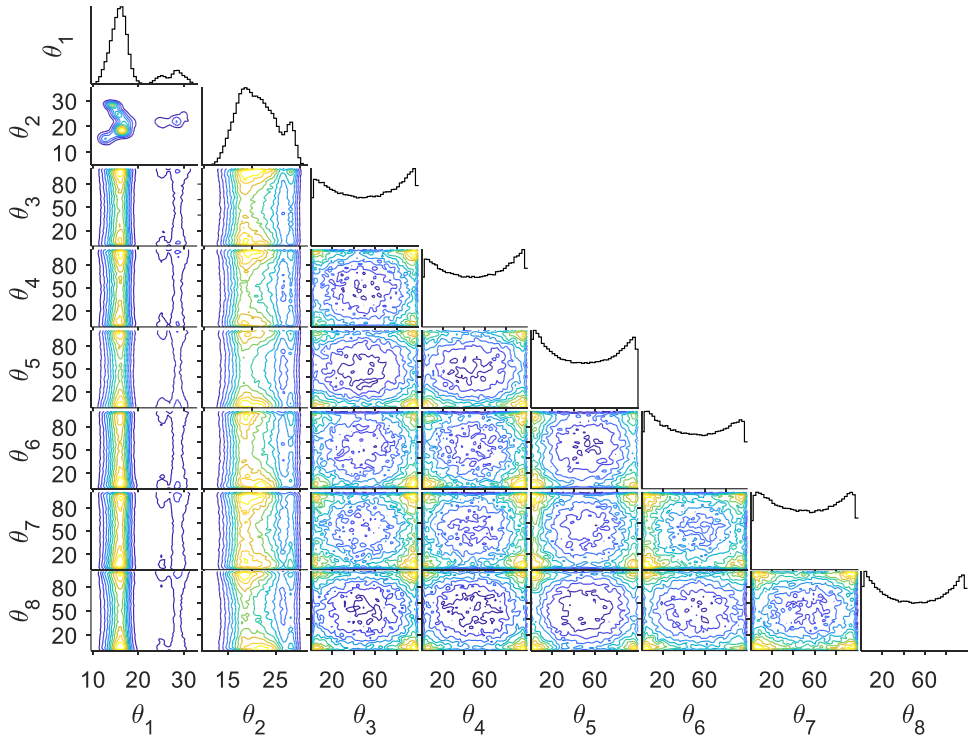


Fig. 25. Final 1-D and 2-D marginal posterior distributions from cyclical VBMC algorithm.

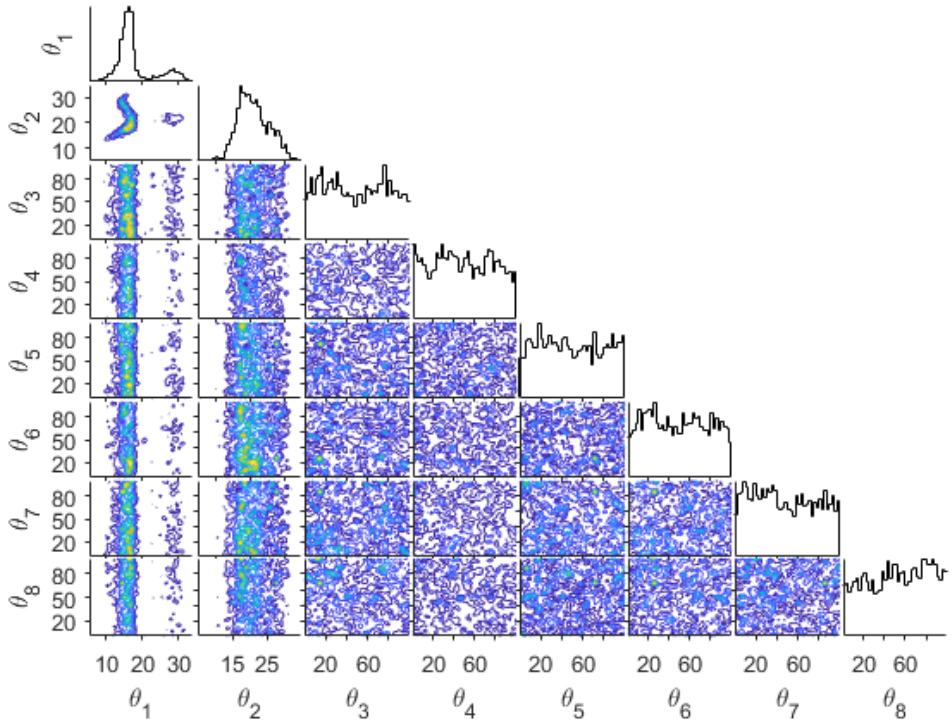


Fig. 26. Final 1-D and 2-D marginal posterior distributions from TEMCMC.

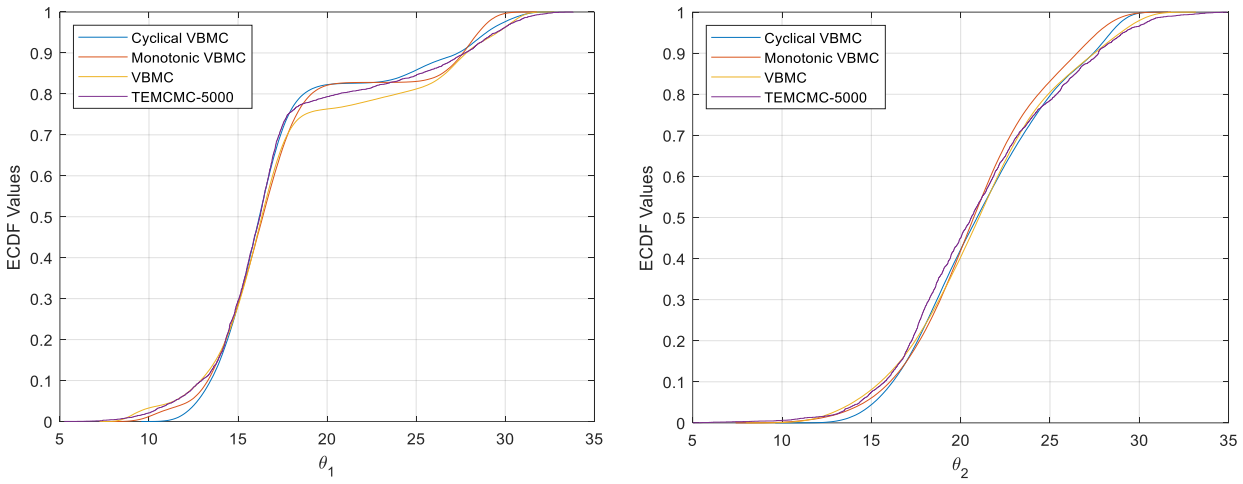


Fig. 27. Marginal ECDFs using Cyclical VBMC, Monotonic VBMC, VBMC and TEMCMC (5000 samples).

Table 6 compares the results obtained using the three VBMC approaches and TEMCMC. It can be seen that for a significantly lower amount of model evaluations, all the three VBMC algorithms show similar results to the TEMCMC algorithm.

Table 6: Comparison of numerical results for the multi-modal aluminium frame problem

Method	N. of samples	N. of Total Iterations for Convergence
VBMC	135	26
Monotonic VBMC	270	53
Cyclical VBMC	265	52
TEMCMC	20000	4

4.5 Discussion of results

The results obtained for the case studies illustrated in section 4 indicate that:

The standard VBMC algorithm displays an excellent performance for the unimodal posterior distribution example, however, for the three multi-modal posterior examples it gets stuck at the initially found mode. This is due to the nature of the algorithm: the active sampling used is unable to escape from that mode. In other words, the algorithm proceeds to only sample in the vicinity of that found mode due to its exploitation nature. A better approximation of the posterior may have been found if a higher number of function evaluations had been used for the initial training set, as that would have meant a better exploration of regions with high probability density. The disadvantage of running a higher number of function evaluations would be that as no guided exploration is used, a significant number of those function evaluations performed would be wasted.

For the three examples featuring the multi-modal posterior distributions, it was shown that the cyclical VBMC outperforms the other methods: it gradually improves convergence by reopening paths and by leveraging on the previous cycles as warm re-starts. When the monotonic VBMC algorithm is used, the paths are formed throughout the iterations, but they are not reopened, meaning that once the exploration phase has finished, exploitation of the previously found regions of high probability density occurs.

The second numerical example focuses on a mildly multi-modal distribution where the difference between the monotonic and cyclical VBMC results is expected to be modest.

In the other multi-modal examples, it was found that to obtain ECDFs of similar accuracy to the TEMCMC algorithm, the cyclical VBMC algorithm required a significantly lower amount of function evaluations of the model (at least an order of magnitude lower). It can be concluded that the annealing schedule improves the convergence of the approximated posterior to the true posterior. This improvement is at the expense of a slight additional computational cost compared to the standard VBMC when dealing with unimodal posterior distributions. This extra computational cost is certainly

justified when the interest is the accurate evaluation of highly multi-modal posteriors which are expected to be found in the Bayesian model updating problems of engineering applications. This is certainly of great interest for methods that require the evaluation of cumulative density functions, such as reliability analysis techniques.

5 Conclusions

In this paper, an approach based on variational inference for the estimation of the multi-modal posterior distributions of the latent parameters of an expensive-to-evaluate physics-based model, given available data, has been proposed. The proposed cyclical VBMC approach yields a non-parametric estimation of the posterior distribution of the identified parameters by combining the active-sampling Bayesian quadrature with a Gaussian-process based variational inference. Multi-modal smooth posteriors can be captured as it uses a multivariate Gaussian mixture postulated posterior. Variational whitening is also used in this proposed approach for a more accurate posterior approximation. The cyclical VBMC algorithm overcomes the constraints raised by poor initializations when the number of model runs that can be explored is small. This is done employing an artificial temperature parameter that cyclically anneals the unnormalized posterior, improving the mode coverage and exploration abilities of the procedure. Three numerical examples and one experimental investigation have shown the advantages of the cyclical VBMC when dealing with multi-modal posteriors and a limited number of physics-based model runs.

The proposed cyclical VBMC approach may benefit other engineering applications, including Bayesian Experimental Design [53–55], and optimal sensor placement frameworks based on information theory [56,57], since it may reduce the computational cost required for the statistical model updating part of these approaches when dealing with less than 20 uncertain parameters. These applications are currently being explored, with particular interest to the cases with more than 20 uncertain parameters, which are often encountered in engineering.

Acknowledgements

Felipe Igea and Alice Cicirello thanks the EPSRC and Schlumberger for an Industrial Case postgraduate scholarship (Grant ref. EP/T517653/1).

References

[1] A. Lye, A. Cicirello, E. Patelli, Sampling methods for solving Bayesian model updating problems: A tutorial, Mech Syst Signal Process. 159 (2021) 107760. https://doi.org/10.1016/j.ymssp.2021.107760.

[2] L.S. Katafygiotis, J.L. Beck, Updating Models and Their Uncertainties. II: Model Identifiability, J Eng Mech. 124 (1998) 463–467. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:4(463).

[3] J.L. Beck, L.S. Katafygiotis, Updating Models and Their Uncertainties. I: Bayesian Statistical Framework, J Eng Mech. 124 (1998) 455–461. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:4(455).

[4] E. Simoen, G. de Roeck, G. Lombaert, Dealing with uncertainty in model updating for damage assessment: A review, Mech Syst Signal Process. 56 (2015) 123–149. https://doi.org/10.1016/j.ymssp.2014.11.001.

[5] Z. Zhang, C. Jiang, X. Han, X.X. Ruan, A high-precision probabilistic uncertainty propagation method for problems involving multimodal distributions, Mech Syst Signal Process. 126 (2019) 21–41. https://doi.org/10.1016/J.YMSSP.2019.01.031.

[6] Y. Xia, B. Chen, S. Weng, Y.Q. Ni, Y.L. Xu, Temperature effect on vibration properties of civil structures: A literature review and case studies, J Civ Struct Health Monit. 2 (2012) 29–46. https://doi.org/10.1007/S13349-011-0015-7/FIGURES/17.

[7] A. Cicirello, On the response bounds of damaged Euler–Bernoulli beams with switching cracks under moving masses, Int J Solids Struct. 172–173 (2019) 70–83. https://doi.org/10.1016/J.IJSOLSTR.2019.05.003.

[8] J.L. Beck, S.-K. Au, Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation, (n.d.). https://doi.org/10.1061/ASCE0733-93992002128:4380.

[9] J. He, X. Guan, R. Jha, Improve the accuracy of asymptotic approximation in reliability problems involving multimodal distributions, IEEE Trans Reliab. 65 (2016) 1724–1736. https://doi.org/10.1109/TR.2016.2604121.

[10] Y.Q. Ni, X.W. Ye, J.M. Ko, Modeling of Stress Spectrum Using Long-Term Monitoring Data and Finite Mixture Distributions, J Eng Mech. 138 (2011) 175–183. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000313.

[11] Y.Q. Ni, X.W. Ye, J.M. Ko, Monitoring-Based Fatigue Reliability Assessment of Steel Bridges: Analytical Model and Application, Journal of Structural Engineering. 136 (2010) 1563–1573. https://doi.org/10.1061/(ASCE)ST.1943-541X.0000250.

[12] R.S. Lima, A. Kucuk, C.C. Berndt, Bimodal distribution of mechanical properties on plasma sprayed nanostructured partially stabilized zirconia, Materials Science and Engineering: A. 327 (2002) 224–232. https://doi.org/10.1016/S0921-5093(01)01530-1.

[13] S.W. Haider, R.S. Harichandran, M.B. Dwaikat, Closed-Form Solutions for Bimodal Axle Load Spectra and Relative Pavement Damage Estimation, J Transp Eng. 135 (2009) 974–983. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000077.

[14] D.H. Timm, S.M. Tisdale, R.E. Turochy, Axle Load Spectra Characterization by Mixed Distribution Modeling, J Transp Eng. 131 (2005) 83–88. https://doi.org/10.1061/(ASCE)0733-947X(2005)131:2(83).

[15] J.K. Kruschke, Markov Chain Monte Carlo, in: Doing Bayesian Data Analysis, Elsevier, 2015: pp. 143–191. https://doi.org/10.1016/B978-0-12-405888-0.00007-6.

[16] D.M. Blei, A. Kucukelbir, J.D. McAuliffe, Variational Inference: A Review for Statisticians, J Am Stat Assoc. 112 (2017) 859–877. https://doi.org/10.1080/01621459.2017.1285773.

[17] W.K. Hastings, Monte carlo sampling methods using Markov chains and their applications, Biometrika. 57 (1970) 97–109. https://doi.org/10.1093/biomet/57.1.97.

[18] W.R. Gilks, P. Wild, Adaptive Rejection Sampling for Gibbs Sampling, Appl Stat. 41 (1992) 337. https://doi.org/10.2307/2347565.

[19] S. Chib, Chapter 57 Markov chain Monte Carlo methods: computation and inference, in: Handbook of Econometrics, Elsevier, 2001: pp. 3569–3649. https://doi.org/10.1016/S1573-4412(01)05010-3.

[20] D.P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, M. Welling, Improving Variational Inference with Inverse Autoregressive Flow, Adv Neural Inf Process Syst. (2016) 4743–4751. http://arxiv.org/abs/1606.04934 (accessed May 12, 2021).

[21] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, D.M. Blei, Automatic Differentiation Variational Inference, Journal of Machine Learning Research. 18 (2016) 1–45. http://arxiv.org/abs/1603.00788 (accessed May 12, 2021).

[22] S. Gershman, M. Hoffman, D. Blei, Nonparametric variational inference, Proceedings of the 29th International Conference on Machine Learning, ICML 2012. 1 (2012) 663–670. http://arxiv.org/abs/1206.4665 (accessed May 12, 2021).

[23] T. Campbell, X. Li, Universal Boosting Variational Inference, ArXiv. (2019). http://arxiv.org/abs/1906.01235 (accessed May 12, 2021).

[24] L. Acerbi, Variational Bayesian Monte Carlo, Adv Neural Inf Process Syst. 2018-December (2018) 8213–8223. http://arxiv.org/abs/1810.05558 (accessed May 11, 2021).

[25] L. Acerbi, Variational Bayesian Monte Carlo with Noisy Likelihoods, Advances in Neural Information Processing Systems 34. (2020). http://arxiv.org/abs/2006.08655 (accessed May 11, 2021).

[26] C.E. Rasmussen, Z. Ghahramani, Bayesian Monte Carlo, n.d. http://www.gatsby.ucl.ac.uk (accessed May 13, 2021).

[27] A. O'Hagan, Bayes-Hermite quadrature, J Stat Plan Inference. 29 (1991) 245–260. https://doi.org/10.1016/0378-3758(91)90002-V.

[28] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, n.d. www.GaussianProcess.org/gpml (accessed May 15, 2021).

[29] P. Ni, J. Li, H. Hao, Q. Han, X. Du, Probabilistic model updating via variational Bayesian inference and adaptive Gaussian process modeling, Comput Methods Appl Mech Eng. 383 (2021) 113915. https://doi.org/10.1016/J.CMA.2021.113915.

[30] A. Lye, A. Cicirello, E. Patelli, An efficient and robust sampler for Bayesian inference: Transitional Ensemble Markov Chain Monte Carlo, Mech Syst Signal Process. 167 (2022) 108471. https://doi.org/10.1016/J.YMSSP.2021.108471.

[31] H.-F. Lam, J.-H. Yang, S.-K. Au, Markov chain Monte Carlo-based Bayesian method for structural model updating and damage detection, Struct Control Health Monit. 25 (2018) e2140. https://doi.org/10.1002/stc.2140.

[32] J. Ching, Y.-C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging., J Eng Mech. 133 (2007) 816. https://doi.org/10.1061/(ASCE)0733-9399(2007)133:7(816).

[33] R. Ranganath, S. Gerrish, D.M. Blei, Black Box Variational Inference, Journal of Machine Learning Research. 33 (2013) 814–822. http://arxiv.org/abs/1401.0118 (accessed May 20, 2021).

[34] A.C. Miller, N. Foti, R.P. Adams, Variational Boosting: Iteratively Refining Posterior Approximations, 34th International Conference on Machine Learning, ICML 2017. 5 (2016) 3732–3747. http://arxiv.org/abs/1611.06585 (accessed May 16, 2021).

[35] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[36] R.M. Neal, Slice sampling, Ann Stat. 31 (2003) 705–741. https://doi.org/10.1214/aos/1056562461.

[37] M.D. McKay, R.J. Beckman, W.J. Conover, A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics. 21 (1979) 239. https://doi.org/10.2307/1268522.

[38] S. Mandt, J. McInerney, F. Abrol, R. Ranganath, D. Blei, Variational Tempering, Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016. (2014) 704–712. https://arxiv.org/abs/1411.1810v4 (accessed November 29, 2021).

[39] H. Fu, C. Li, X. Liu, J. Gao, A. Celikyilmaz, L. Carin, Cyclical Annealing Schedule: A Simple Approach to Mitigating KL Vanishing, NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference. 1 (2019) 240–250. https://arxiv.org/abs/1903.10145v3 (accessed November 29, 2021).

[40] R. Zhang, C. Li, J. Zhang, C. Chen, A.G. Wilson, Cyclical Stochastic Gradient MCMC for Bayesian Deep Learning, (2019). https://arxiv.org/abs/1902.03932v2 (accessed November 30, 2021).

[41] G. Huang, Y. Li, G. Pleiss, Z. Liu, J.E. Hopcroft, K.Q. Weinberger, Snapshot Ensembles: Train 1, get M for free, 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings. (2017). https://arxiv.org/abs/1704.00109v1 (accessed November 30, 2021).

[42] I. Loshchilov, F. Hutter, SGDR: Stochastic Gradient Descent with Warm Restarts, 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings. (2016). https://arxiv.org/abs/1608.03983v5 (accessed November 30, 2021).

[43] S.H. Jacobson, S.N. Hall, L.A. McLay, J.E. Orosz, Performance Analysis of Cyclical Simulated Annealing Algorithms, Methodology and Computing in Applied Probability 2005 7:2. 7 (2005) 183–201. https://doi.org/10.1007/S11009-005-1482-2.

[44] R.B. Gramacy, H.K.H. Lee, Cases for the nugget in modeling computer experiments, Stat Comput. 22 (2012) 713–722. https://doi.org/10.1007/s11222-010-9224-x.

[45] D.P. Kingma, M. Welling, Auto-encoding variational bayes, in: 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, International Conference on Learning Representations, ICLR, 2014. https://arxiv.org/abs/1312.6114v10 (accessed May 16, 2021).

[46] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, International Conference on Learning Representations, ICLR, 2015. https://arxiv.org/abs/1412.6980v9 (accessed May 16, 2021).

[47] D.P. Kingma, J.L. Ba, Adam: A Method for Stochastic Optimization, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. (2014). https://arxiv.org/abs/1412.6980v9 (accessed December 7, 2021).

[48] MATLAB, (2020).

[49] C. Safta, M. Khalil, H.N. Najm, Transitional Markov Chain Monte Carlo Sampler in UQTk, (2020). https://doi.org/10.2172/1606084.

[50] M. Kitahara, S. Bi, M. Broggi, M. Beer, Nonparametric Bayesian stochastic model updating with hybrid uncertainties, Mech Syst Signal Process. 163 (2022) 108195. https://doi.org/10.1016/J.YMSSP.2021.108195.

[51] R. Rocchetta, M. Broggi, Q. Huchet, E. Patelli, On-line Bayesian model updating for structural health monitoring, Mech Syst Signal Process. 103 (2018) 174–195. https://doi.org/10.1016/J.YMSSP.2017.10.015.

[52] P. Liang, J.E. Mottershead, F.A. DiazDelaO, Model updating with the Kriging predictor: Effect of code uncertainty, Proceedings of ISMA 2016 - International Conference on Noise and Vibration Engineering and USD2016 - International Conference on Uncertainty in Structural Dynamics. (2016) 4363–4376.

[53] E.G. Ryan, C.C. Drovandi, J.M. Mcgree, A.N. Pettitt, A Review of Modern Computational Algorithms for Bayesian Optimal Design, International Statistical Review. 84 (2016) 128–154. https://doi.org/10.1111/insr.12107.

[54] Z. Xu, Q. Liao, Gaussian process based expected information gain computation for bayesian optimal design, Entropy. 22 (2020) 258. https://doi.org/10.3390/e22020258.

[55] X. Huan, Y.M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, J Comput Phys. 232 (2013) 288–317. https://doi.org/10.1016/j.jcp.2012.08.013.

[56] G. Capellari, E. Chatzi, S. Mariani, Structural Health Monitoring Sensor Network Optimization through Bayesian Experimental Design, ASCE ASME J Risk Uncertain Eng Syst A Civ Eng. 4 (2018) 04018016. https://doi.org/10.1061/ajrua6.0000966.

[57] C. Argyris, S. Chowdhury, V. Zabel, C. Papadimitriou, Bayesian optimal sensor placement for crack identification in structures using strain measurements, Struct Control Health Monit. 25 (2018) e2137. https://doi.org/10.1002/stc.2137.

5.3 Conclusions

In the paper III attached in this chapter, the use of Variational Inference (VI) for structural model updating is investigated. An optimization algorithm is employed to choose from a family of distributions the member that best approximates the posterior of the uncertain parameters. In the method described in this paper, the variational posterior that maximises the evidence lower bound (ELBO) is chosen. A variational posterior distribution defined by a multivariate Gaussian estimation is assumed with the purpose of inferring the uncertain parameters. This investigation shows that in comparison with MCMC using the Metropolis Hastings algorithm, a smaller number of samples is required with the suggested method. Nevertheless, it is also found that the accuracy of the proposed algorithm is lower than the obtained with the MCMC approach. The advantages of the method are more evident when fast inference, in a context of restrictive computational allocation is required.

The second VI work, paper IV, also attached in this chapter, illustrates the combination of a Gaussian-process based VI, with active-sampling Bayesian Quadrature, to produce non-parametric estimations of the posterior pdfs of the relevant parameters. The algorithm illustrated in paper IV, introduces in the VBMC algorithm a cyclical annealing schedule. This allows the algorithm to have exploration and exploitation phases in each annealing cycle. This is possible as the temperature is initially started at a high value, and is decreased to unity at the end of each cycle. These cycles are repeated for a prescribed number of times. The developed Cyclical VBMC method is able to efficiently manage multi-modal posteriors and the number of evaluations is reduced by roughly a factor of twenty compared to the sampling algorithm TEMCMC.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Structural Model Updating Using Variational Inference
Publication Status	☒Published □ Accepted for Publication □Submitted for Publication □Unpublished and unsubmitted work written in a manuscript style
Publication Details	Igea F., Chatzis M.N.., Cicirello A., Structural Model Updating Using Variational Inference, Structural Health Monitoring 2021, Enabling Next-Generation SHM For Cyber-Physical Systems, 2021.

Student Confirmation

Student Name:	Felipe Igea		
Contribution to the Paper	Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing.		
Signature	Felipe Igea	Date	16/09/23

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title:	Dr Alice Cicirello		
Supervisor comments	The student made a substantial contribution to the publication, and the description above is accurate		
Signature		Date	18/09/23

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Cyclical Variational Bayes Monte Carlo for Efficient Multi-Modal Posterior Distributions Evaluation
Publication Status	☒Published □ Accepted for Publication □Submitted for Publication □Unpublished and unsubmitted work written in a manuscript style
Publication Details	Igea F., Cicirello A., Cyclical Variational Bayes Monte Carlo for Efficient Multi-Modal Posterior Distributions Evaluation, Mechanical Systems and Signal Processing, 2023.

Student Confirmation

Student Name:	Felipe Igea		
Contribution to the Paper	Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing.		
Signature	Felipe Igea	Date	16/09/23

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title:	Dr Alice Cicirello		
Supervisor comments	The student made a substantial contribution to the publication, and the description above is accurate		
Signature		Date	18/09/2023

Chapter 6 - Limited information in operating conditions – updating our physics-based model under prior uncertainty

Bayesian Inference may show great sensitivity to the choice of priors for the latent parameters especially for cases where the availability of data is limited. As a result, some of the processes that require the evaluation of the posterior can be affected by the choice of priors, and therefore, certain recommendations related to the structure (e.g., maintenance is required) may change for different assumed priors. There are some situations e.g., when experts may show conflicting opinions, where it is not possible to agree on a proposal for the prior to be used for the latent variables. For these situations, it would be ideal to know how predicted posteriors are affected by the choice of priors in the vicinity of a given nominal prior. Accordingly, the development of a methodology able to find out the optimal and worse distributions in the vicinity of a specific prior w.r.t. a functional, would be of great significance.

The paper V attached, develops a method in the context of Bayesian Inference under prior uncertainty, able to assess the robustness of the posterior prediction to changes in the prior distribution.

6.1 Paper V – 'An Interacting Wasserstein Gradient Flow Strategy to Robust Bayesian Inference'

This paper has not yet been published, but has been submitted to the open-access journal Data-Centric Engineering and is available on arXiv.

https://doi.org/10.48550/arXiv.2401.11607

An Interacting Wasserstein Gradient Flow Strategy to Robust Bayesian Inference

Felipe Igea[1]* and Alice Cicirello[1,2]

[1] Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

[2] Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK

*Corresponding author: felipe.igea@eng.ox.ac.uk

Abstract

Model Updating is frequently used in Structural Health Monitoring to determine structures' operating conditions and whether maintenance is required. Data collected by sensors are used to update the values of some initially unknown physics-based model's parameters. Bayesian Inference techniques for model updating require the assumption of a prior distribution. This choice of prior may affect posterior predictions and subsequent decisions on maintenance requirements, specially under the typical case in engineering applications of little informative data. Therefore, understanding how the choice of prior may affect the posterior prediction is of great interest.

In this paper, a Robust Bayesian Inference technique evaluates the optimal and worst-case prior in the vicinity of a chosen nominal prior, and their corresponding posteriors. This technique employs an interacting Wasserstein Gradient Flow formulation.

Two numerical case studies are used to showcase the proposed algorithm: a double-banana-posterior and a double beam structure. Optimal and worst-case prior are modelled by specifying an ambiguity set containing any distribution at a statistical distance to the nominal prior, less or equal to the radius. Examples show how particles flow from an initial assumed Gaussian distribution to the optimal worst-case prior distribution that lies inside the defined ambiguity set, and the resulting particles from the approximation to the posterior.

The resulting posteriors may be used to yield the lower and upper bounds on subsequent calculations used for decision-making. If the metric used for decision-making is not sensitive to the resulting posteriors, it may be assumed that decisions taken are robust to prior uncertainty.

Impact Statement

Bayesian Inference may be significantly sensitive to assumptions about the priors chosen for the latent parameters of the structure, especially if due to some restrictions such as time constraints and cost, the number of observations available is limited. In these cases, the selection of priors may affect the resulting posteriors, and as a consequence the decisions about reliability, useful lifetime and maintenance of the structure are influenced. To address these limitations a Robust Bayesian Inference approach based on interacting Wasserstein Gradient Flows has been developed in this paper. The method estimates the optimal and worst cases of priors, and calculates their corresponding approximations to the posterior that may be used as lower and upper bounds on subsequent calculations used for decision-making. If the resulting metric used for decision-making does not change significantly from using the lower to using the upper bound, it may be assumed that the decisions taken are robust to prior uncertainty.

Keywords: Robust Bayesian Inference; Interacting Wasserstein Gradient Flows; Ambiguity sets.

1. Introduction

Numerical simulations are used in a multitude of scientific and engineering areas to study the behaviour of complex systems under various conditions. However, the resulting numerical 'observations' from those numerical simulations have to be cautiously used during the inference process when compared with the corresponding experimental observations, as otherwise unfaithful posterior estimates of the uncertain parameters will be produced (Hermans et al., 2021). It should also be mentioned that the observations employed in the inference should be informative, providing new information about the system under analysis. In this paper, a Bayesian Inference approach is developed with the purpose of assessing the sensitivity of the posterior predictions with respect to (w.r.t.) uncertain priors.

When a model is chosen to generate data, the statistical model is just an approximation, and some errors are unavoidable (Dewaskar et al., 2023): data noise, assumption of model normality, incorrect assumptions about parameters, etc. For the inference of some model's parameters, significant impact may be produced by only small errors on the specification of the statistical model (Dewaskar et al., 2023). The traditional approach has been the use of a statistical model flexible enough to explicitly include the intricacies of the actual data (noises, outliers, etc.). As models characterising systems have grown in complexity, methodologies able to perform inference in these cases with intricate likelihood functions have been developed (Frazier, 2020). Unreliable approximations of the system's behaviour are mainly produced by not accounting for all plausible values of the observations that may be obtained from experiments. In engineering, this is of particular interest, as the number of experiments that may be run is limited due to the high cost incurred and time constraints. In these cases where the complexities of the likelihood are increased to improve the models' accuracy, and therefore, the reliability of the inferences, techniques such as: mixture models (Diakonikolas et al., 2020), nonparametric or semiparametric models (Lyddon et al., 2018), and models with heavier tails (Gonçalves et al., 2015) have been used as likelihood functions (Dewaskar et al., 2023). Nonetheless, the introduction of those methodologies to define the likelihood functions frequently lead to a set of

new issues: higher numerical cost, definition of parameters, and harder interpretability. Although these techniques to define complex likelihood functions may improve the model's specification, some amount of inaccuracy is unavoidable (Dewaskar et al., 2023).

Current literature shows two main methodologies for robustifying inferences in the presence of model misspecification by focusing on the likelihood, and those can be broadly grouped into: (i) modified likelihood functions (Ghosh & Basu, 2016; Hooker & Vidyashankar, 2011), and (ii) distance-based estimation (Chérief-Abdellatif & Alquier, 2019; Matsubara et al., 2021). In the paper (Ghosh & Basu, 2016) an algorithm designed to produce robust Bayes estimators through modified likelihood functions, using the density power divergence is described. The method is designed to overcome the problems to manage outliers that may arise when the usual Bayesian estimator based on the ordinary posterior density is used. Another paper that uses the concept of modified likelihood functions is (Hooker & Vidyashankar, 2011). The method proposed is based on disparity theory, and it produces efficient and robust Bayesian inferences. Substituting the log likelihood by a suitably scaled disparity, the authors, using several examples, illustrate that robust inferences are obtained. The concept of distance-based estimation is used in the paper (Chérief-Abdellatif & Alquier, 2019) to produce a robust pseudo-posterior that is called the MMD-Bayes. MMD stands for Maximum Mean Discrepancy. The authors show that even for cases where noisy data and outliers are present, the MMD-Bayes posterior shows robustness to model misspecification. In their methodology, robustness is introduced through the substitution of the likelihood by the exponential of the MMD. In other words, this algorithm introduces in the Bayes equation the MMD-Bayes as a substitute loss function. Another example where robustness is sought through distance-based estimation is shown in (Matsubara et al., 2021). In (Matsubara et al., 2021), the loss function used in the Bayesian Inference scheme is a Stein discrepancy. This method produces a robust posterior for cases where the likelihood is intractable/misspecified, and it generates a tractable posterior for Markov Chain Monte Carlo (MCMC).

However, when little data is available, Bayesian Inference may be significantly sensitive to the

assumptions of the prior for the latent parameters. In these cases where a limited number of observations is available, the choice of the prior may substantially affect the posterior obtained. In engineering, this may affect subsequent decisions such as those used to assess the reliability of a structure, its remaining useful lifetime, and whether a structure requires predictive maintenance. Therefore, a method able to quantify the robustness of the posterior prediction when the assumptions of the prior are changed, is of great interest. More specifically, if the optimal prior and the worst-case prior with respect to a metric could be obtained, the resulting posteriors may be used as lower and upper bounds on subsequent calculations used for decision-making. If the difference between the upper bound and lower bound found using the method is low for the metric used to support a decision, then it may be confirmed that the decision taken is robust to prior uncertainty. This is the focus of the present paper.

More specifically, it might not be possible to define exactly the prior for the latent variables. This type of situation could arise in the presence of conflicting opinions from experts. For those cases, we would like to explore how the approximation to the posterior might be affected by distributions that are in the neighbourhood of an assumed nominal prior, as this might have consequences on subsequent calculations. Therefore, it would be useful to develop a method that could determine the worst or most optimal distribution inside that neighbourhood of distributions in terms of a particular functional of interest.

In this paper, the problem of robustness to prior uncertainty in Bayesian Inference is dealt with by developing an interacting Wasserstein Gradient Flow combined with a so-called ambiguity set. An interacting Wasserstein Gradient Flow is derived to find: a) the best approximation to the posterior by minimising the Kullback-Leibler divergence (KL divergence) between the posterior and the approximation to the posterior, where the posterior is subject to change (due to the prior also changing). b) the (optimal or worst-case) prior that either minimises or maximises the KL divergence between the posterior and the approximation to the posterior. The proposed approach calculates the resulting optimal or worst-case prior by constraining the space of distributions to be explored using

an ambiguity set. This ambiguity set is defined by a nominal distribution and all the distributions that lie within a specified value of a statistical distance, where both are assumed to be known. The robustness of the method is derived from this distance metric. A useful property of the Wasserstein distance is that distributions that do not share the same support may be investigated inside the ambiguity set (Kuhn et al., 2019). A particle-based interacting Wasserstein Gradient Flow Wasserstein-2 space algorithm is developed and the results from two numerical case studies are presented.

2. Robust Bayesian Inference framework

The proposed Robust Bayesian Inference approach is based on the Wasserstein Gradient Flow formulation (Santambrogio, 2016). This method has been developed to deal with situations where the prior is uncertain, but it can be described by an ambiguity set (Bayraksan & Love, 2015). This is useful, as in some Bayesian Inference problems significant difficulties arise to define the priors of the latent parameters to be inferred. For example, when the suggestions of different experts about which priors should be used significantly differ. For the cases where the amount of observed data is limited, significant changes of posterior may be found for different choices of prior, and therefore, decisions to be taken for predictive maintenance may be affected. In these situations, an ambiguity set defined by a nominal prior, a statistical distance and a radius may be assumed, and the posterior resulting from identifying the optimal and the worst-case prior can be investigated by limiting the distribution space to priors within a statistical distance ε of the nominal prior. In the next subsection, the concept of ambiguity set is defined.

2.1. Ambiguity set

An ambiguity set is a set of distributions close to a reference distribution $p(\boldsymbol{\theta})$ with respect to some statistical distance r (Bayraksan & Love, 2015). An ambiguity set is defined by the nominal distribution $p(\boldsymbol{\theta})$, a statistical distance r and a radius ε. The ambiguity set is used to restrict the space of distributions that the prior could in theory take to solve the optimisation of the chosen

functional. Figure 1, shows an ambiguity set that is centered at a nominal distribution $p(\boldsymbol{\theta})$ and contains any distribution p^* within a statistical distance r less or equal to ε, this may be expressed as:

$$A(\varepsilon, p) = \left\{ p^* : r\left(p^*(\theta) \,\|\, p(\theta) \right) \leq \varepsilon \right\} \tag{1}$$

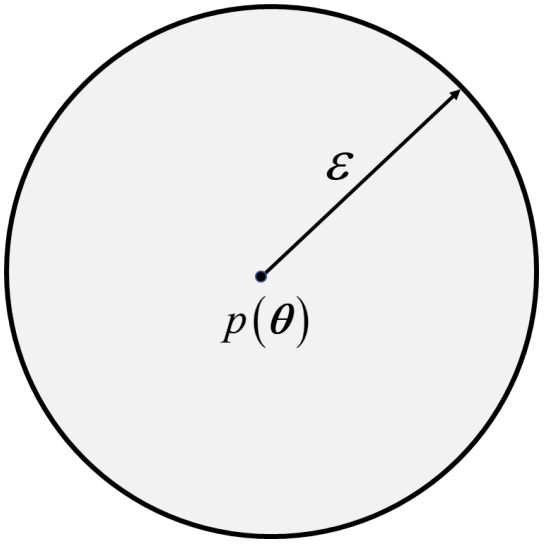


Fig. 1. Ambiguity set centered at $p(\boldsymbol{\theta})$ with radius ε.

When the ambiguity set is defined, two conditions must be met (Go & Isaac, 2022): $\int_{\Omega} p^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1$ and $r\left(p^*(\theta) \,\|\, p(\theta) \right) \leq \varepsilon$.

Any statistical distance may be used to define the ambiguity set, but care should be taken in choosing this distance, as which distributions lie inside that ambiguity set are defined by that statistical distance's properties. For example, if a phi divergence is used as the statistical distance in the ambiguity set, all distributions inside the ambiguity set must be absolutely continuous w.r.t. the nominal distribution (van Parys et al., 2017). However, if the 2-Wasserstein distance is used to define the ambiguity set, then the distributions that lie within the ambiguity set do not need to be absolutely continuous w.r.t. the nominal distribution (Kuhn et al., 2019). The use of the 2-Wasserstein distance also means that distributions that lie within the ambiguity set do not need to share the same support

(Kuhn et al., 2019).

Depending on the information that the practitioner has available, the nominal distribution may be given by either an empirical distribution or a parametric distribution (e.g., Gaussian distribution) as shown in fig. 2. In fig. 2, $\hat{p}$ is a possible nominal distribution, N is the number of data points, δ is the Kronecker delta function, ξ is the parameter of the data, $\mathcal{N}\left(\hat{\mu}_N, \hat{\Sigma}_N\right)$ is a Gaussian distribution with sample mean $\hat{\mu}_N$ and sample covariance $\hat{\Sigma}_N$, obtained from N data points.

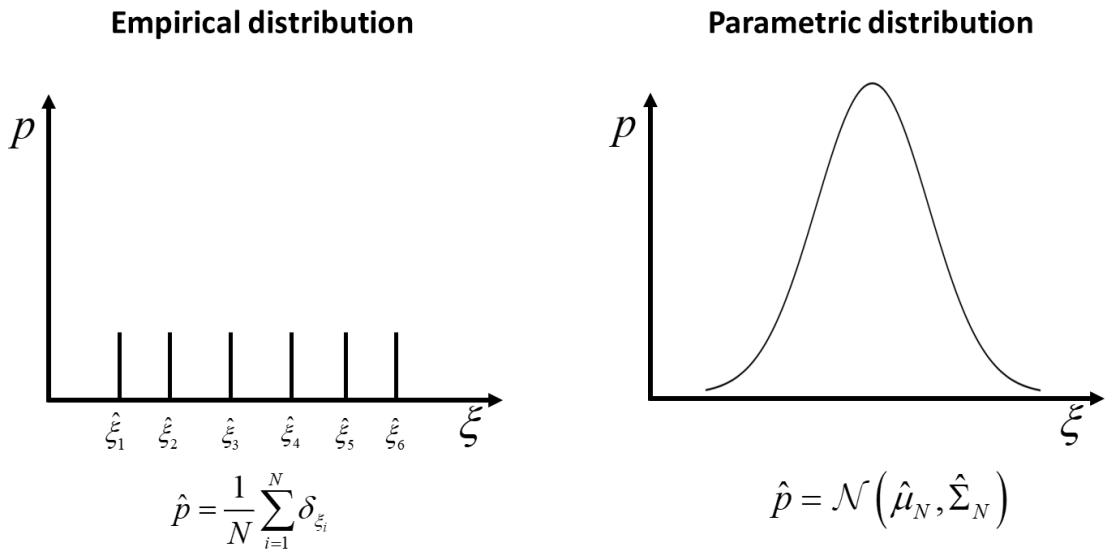


Fig. 2. Nominal distributions: empirical vs. parametric distribution.

The chosen statistical distance for the ambiguity set is the 2-Wasserstein distance, as previously mentioned, this allows us to explore distributions that do not need to share the same support as the nominal distribution.

2.2. Simultaneous optimization of approximated posterior and optimal or worst-case prior

In this paper, we explore the problem of the simultaneous optimization of the approximation to the posterior and the optimal or worst-case prior by using an interacting Wasserstein Gradient Flow (WGF) scheme. The proposed approach differs from current Bayesian Inference WGFs based approaches, as it formulates a new problem that requires interacting WGFs for the simultaneous

optimization of the chosen functional. This interacting WGF simultaneously obtains the best approximation to the posterior, and the optimal or worst-case prior that either minimises or maximises a certain functional. The ambiguity set is used to restrict the space of distributions that the prior could in theory take to solve the optimisation of the chosen functional $E\left(\rho(\theta), p(\theta)\right)$. In this present paper, the min-max (or min-min) formulation problem that needs to be solved is:

$$\min_{\rho(\theta)\in\mathcal{P}(\Omega)} \min \text{or} \max_{p^*(\theta)\in\mathcal{W}\left(p(\theta),p^*(\theta)\right)\leq\varepsilon} E\left(\rho(\theta), p(\theta)\right)$$

$$\text{where } E\left(\rho(\theta), p(\theta)\right) := \int \rho(\theta)\log\left(\frac{\rho(\theta)}{p(\theta)\,p(\boldsymbol{y}_{obs}\,|\,\theta)}\right) \tag{2}$$

The distribution $\rho(\theta)$ is the approximation to the posterior, the likelihood distribution is $p(\boldsymbol{y}_{obs}|\theta)$, the density $p(\theta)$ is the prior, and $\mathcal{W}$ is the 2-Wasserstein distance chosen to define the ambiguity set. The chosen functional $E\left(\rho(\theta), p(\theta)\right)$ is the KL divergence between the unnormalized posterior $p(\theta, \boldsymbol{y}_{obs})$ and the approximation to the posterior $\rho(\theta)$. This functional is chosen as it recently has been used to derive a WGF for Bayesian Inference (Y. Chen et al., 2023; Gao & Liu, 2020; Wang et al., 2022).

By using the properties of the logarithm, the functional in equation (2) can be rewritten as:

$$E\left(\rho(\theta), p(\theta)\right) := \int \rho(\theta)\log\left(\rho(\theta)\right)d\theta - \int \rho(\theta)\log\left(p(\theta)\right)d\theta - \int \rho(\theta)\log\left(p(\boldsymbol{y}\,|\,\theta)\right)d\theta$$

The first term of the equation corresponds to the definition of entropy $\mathcal{H}$ w.r.t. the approximation to the posterior, therefore, the functional can be further expressed as:

$$E\left(\rho(\theta), p(\theta)\right) := -\mathcal{H}\left(\rho(\theta)\right) - \int \rho(\theta)\log\left(p(\theta)\right)d\theta - \int \rho(\theta)\log\left(p(\boldsymbol{y}\,|\,\theta)\right)d\theta \tag{3}$$

The purpose of deriving an interacting Wasserstein Gradient Flow is to locate the pair of probability distributions $\left(\rho^*, p^*\right)$ that balances the simultaneous minimisation and maximisation (or minimisation) of the functional in equation (3). In other words, we are interested in finding simultaneously the distribution $\rho(\theta)$ that minimizes the KL divergence between the unnormalized

posterior $p(\boldsymbol{\theta}, \boldsymbol{y}_{obs})$ and the approximation to the posterior $\rho(\boldsymbol{\theta})$, and the prior/s that minimizes/maximises the KL divergence between the unnormalized posterior $p(\boldsymbol{\theta}, \boldsymbol{y}_{obs})$ and the approximation to the posterior $\rho(\boldsymbol{\theta})$.

In numerous occasions, efforts have been made to prove the convergence of algorithms with interacting Wasserstein Gradient Flows to their global solution (Chizat & Bach, 2018; Mei et al., 2018), but these attempts generally require entropy regularization. The entropy regularisation is already included in the formulation of the functional shown in equation (3), where the first term regularises the partial differential equation of the WGF that minimises the KL divergence between the posterior and the approximation to the posterior. The second term in equation (3) serves as a regulariser of the WGF that minimises/maximises the KL divergence between the posterior and the approximation to the posterior to obtain the optimal or worst-case prior respectively. In this paper, it is assumed that the regularisers allow convergence to the pair of probability distributions that are sought. Proving convergence to this pair of probability distributions is still a problem currently under investigation, and not attempted to be solved in the current paper, the reader is referred to the Mixed Nash Equilibria literature for more details (Ding et al., 2023; Lin et al., 2019; Y. Lu, 2022).

2.3. Proposed algorithm and workflow

The proposed approach is schematically summarised in fig. 3, and it is composed of three main parts, the inputs, the simultaneous functional optimization and the outputs. The physics-based model (analytical, numerical or equivalent surrogate model) of the engineering system of interest, a nominal prior distribution on the unknown latent parameters with a specified radius, a statistical distance to define the ambiguity set, an assumed likelihood and measurements taken from the engineering system are needed as inputs. The main outputs, as shown in fig. 3, are the optimal or worst-case prior distribution and, consequently, the approximation to the posterior.

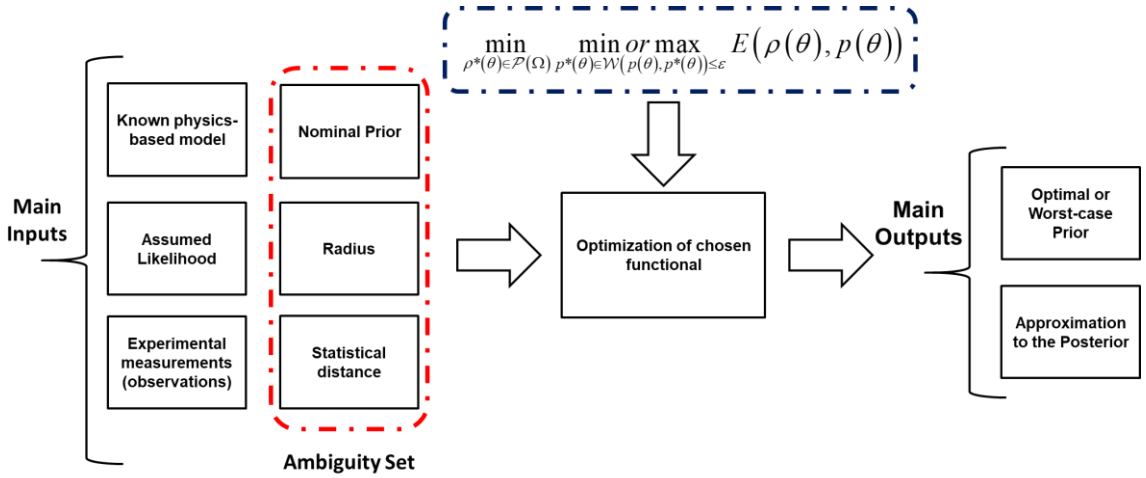


Fig. 3. Main inputs, functional optimization and main outputs of the proposed approach.

In fig. 4, the elements of the optimization block shown in fig. 3 of the proposed approach are described. At first, we allow the approximation to the posterior to minimize the KL divergence between the posterior and the approximation to the posterior without changing the prior distribution. This is done by making the step size τ equal to zero of the WGF that results from either maximising or minimising the functional in equation (3) with respect to the prior for a prescribed amount of iterations N_a. The optimization to find the best approximation to the posterior is performed as follows. For the first iteration, N_0 initial particles are chosen at random (usually drawn from the nominal prior), and the same set of particles are used for both the initial prior $p_0(\boldsymbol{\theta})$ and the initial approximation to the posterior $\rho_0(\boldsymbol{\theta})$. At each iteration $i < N_a$, the physics-based model $PM(\boldsymbol{\theta})$ is run at the corresponding particle positions Θ_i^N of the approximation to the posterior. These numerical simulations at the particle positions Θ_i^N are then used to calculate the gradient of the logarithm of the likelihood $\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta})$ at those respective locations. The gradient of the logarithm of the prior $\nabla_{\boldsymbol{\theta}} \log\left(p(\boldsymbol{\theta})\right)$, and the gradient of the logarithm to the approximation to the posterior $\nabla \log \rho_t(\boldsymbol{\theta})$, are approximated using a kernel density estimation approach, where the bandwidth is chosen using the median approach (Q. Liu & Wang, 2016). Using the equation (17), a new set of N particles $\Theta_{i+1}^N \sim \rho_{i+1}(\boldsymbol{\theta})$ are obtained. This process is repeated until the iteration number reaches $i = N_a$, this

is done to ensure that the approximation to the posterior has converged to the true posterior.

Once the prescribed number of iterations has been reached, the step size τ is allowed to be non-zero and positive, such that at every iteration $i \geq N_a$, a new set of prior particles $\boldsymbol{\theta}^N_{prior,i+1}$ are obtained using the second equation (40). At this stage, both the approximation to the posterior and prior are updated using equation (40), such that the resulting new set of particles corresponds to independently and identically distributed samples from the distributions $\rho_{i+1}(\boldsymbol{\theta})$ and $p_{i+1}(\boldsymbol{\theta})$ of the next iteration number $i+1$.

Additionally, with the purpose of constraining the distribution to be optimised inside the ambiguity set, the 2-Wasserstein distance from the nominal prior to the prior at iteration i is calculated at each iteration of the proposed method. If the distribution lies outside the ambiguity set, the distribution is discarded, and the size of the step in the particle flow algorithm is reduced until the distribution lies within the ambiguity set. In this way, the step size is controlled to restrict the prior within the radius of the Wasserstein ambiguity set. This is based on the assumption that the distribution that maximises or minimises the KL divergence between the actual posterior and the approximation to the posterior lies at the radius of the ambiguity set. Moreover, if a preset number N_b of distributions are discarded when determining whether a distribution belongs in the ambiguity set, the distribution at iteration $i+1$ is reset to the distribution from an earlier iteration $i-N_c$ to avoid the optimisation getting trapped at one of the local optima. Convergence of the prior to the optimal or worst-case prior is assumed if the previously mentioned resetting occurs N_{reset} times. In this case, the prior is no longer reset, and in a manner similar to the one defined at the beginning of the algorithm, an additional number of iterations N_a are allowed, so the approximation to the posterior can converge. At this stage, the algorithm checks if the stopping criteria have been fulfilled, if it has not, a new iteration $i+1$ is started. The stopping criteria are set as: (i) the maximum number of allowed iterations $N_{\max}$ is reached; (ii) a maximum number of prior distributions N_{reset} are reset to the distribution from an

earlier iteration, and an additional number of iterations N_a are allowed for the approximation to the posterior to converge.

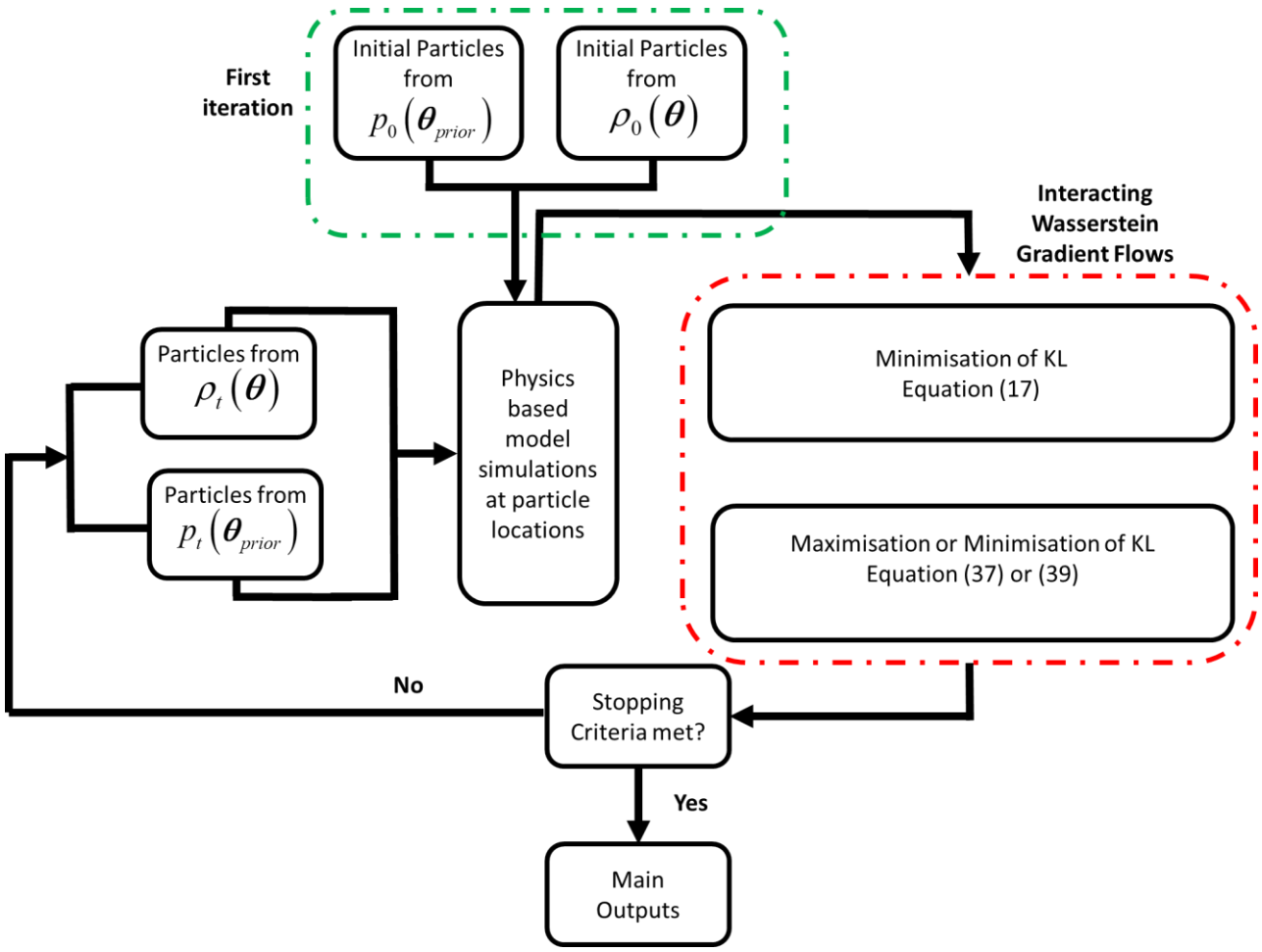


Fig. 4. Pictorial description of simultaneous optimization of chosen functional.

A summary of the steps to be run for the proposed method are given below:

1. Calculate approximation to the posterior for the initial prior (by setting $\tau_t = 0$)
    a. Obtain N_0 initial particles from the prior, and approximation to the posterior.
    b. Calculate next set of particles of the approximation to the posterior using equation (17).
    c. Repeat from 1a) until iteration number reaches $i = N_a$.
2. Simultaneous optimization of equation (2) to calculate the approximation of the posterior and optimal or worst-case prior (allow $\tau_t > 0$):

a. Calculate the next set of particles of the approximation to the posterior and prior using equations (17) and (37) or (39).

b. Check if the prior lies outside the defined ambiguity set:

   i. if false, continue to 2c).

   ii. if true:

      1. reduce the time step τ_t until it is inside ambiguity set.

      2. check if the number of discarded distributions is less than N_b.

         a. if true, continue.

         b. if false, reset current prior particles to prior particles from iteration $i - N_c$.

      3. check if the number of times prior distributions have been reset is less than N_{reset}.

         a. if true, continue.

         b. if false, skip to step 3.

c. Repeat from 2a) until iterations reach $N_{\max}$ and stop running the algorithm.

3. Calculate the approximation to the posterior for the final prior (by setting τ_t equal to zero, and allowing an additional number of iterations N_a):

   a. Calculate next set of particles of the approximation to the posterior using equation (17).

   b. Repeat from 3a) until iterations reach $N_{\max}$ or the additional number of iterations is reached.

The system illustrated in fig. 5 is used to show the main results that would be obtained by using the proposed algorithm. A 1D mass-spring system with mass $m = 1$ [kg], stiffness $k = 1$ [N/m], and angular frequency $\omega = \sqrt{\frac{k}{m}}$ [rad/s] is studied. In this example, a Gaussian observational error

$\sigma = \sqrt{0.1}$ is assumed when obtaining a numerical observation of the angular frequency $\omega_{obs} = \sqrt{\frac{k}{m}} + \zeta$, where $\zeta \sim \mathcal{N}(0, \sigma)$. It is also assumed, that the uncertain parameter is the spring stiffness $k = \theta$ [N/m]. The initial Gaussian prior (which is the same as the nominal prior of the ambiguity set) is assumed to be $p(\theta) = \mathcal{N}(1, 0.1)$. Two different runs to obtain the optimal prior and the worst-case prior (and their corresponding approximation to the posterior) w.r.t. the chosen functional in equation (2) are shown in fig. 6. An ambiguity set with a radius $\varepsilon = 0.005$ is chosen. For the both optimal and worst-case prior, the step sizes in the interacting particle flow WGF algorithm cases are $\alpha = 3*10^{-3}$ and $\tau = 3*10^{-4}$. The values of N_a, N_b, N_c, N_{reset} and $N_{\max}$ used to run the algorithm are the same as the two numerical examples shown in Section 6 and they can be found in the introduction of that section. Also, in this example, the number of initial particles $N_0 = 100$ is chosen. The obtained probability density functions plotted in fig. 6 are calculated using the `kde` function in (*MATLAB*, 2020) with the standard options, and using 100 samples from their respective distributions.

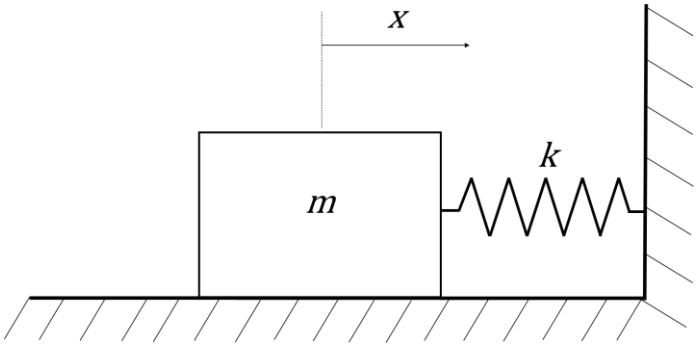


Fig. 5. 1-Degree of freedom mass-spring system.

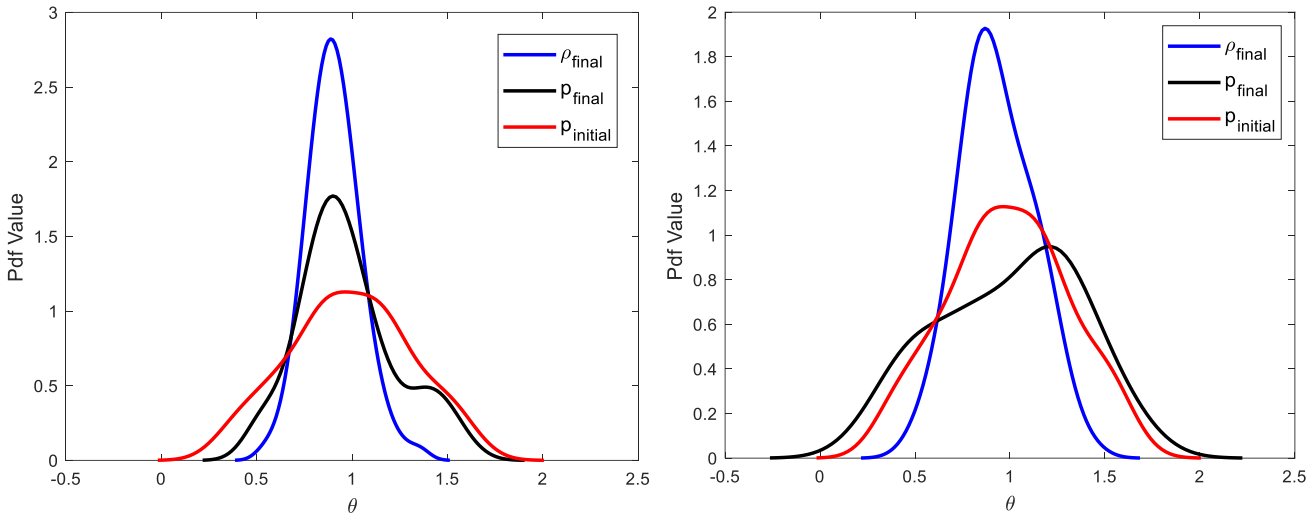


Fig. 6. Kernel density estimates of the distributions for a 1D mass-spring system given an initial/nominal prior (red – initial prior; blue – final approximation to the posterior; black – final prior): (a) Optimal prior case (b) Worst-case prior case.

As expected, it can be seen on fig. 6, that the optimal prior case assigns higher probability density at regions of high posterior density, while the worst-case prior moves prior density away from regions of high posterior density. The optimal prior has its support reduced w.r.t. the initial prior, whilst for the worst-case prior its support is increased. A slight multimodality can be seen for the optimal prior, with its main mode at the same location as the only mode found for the approximation to the posterior. Both the optimal and worst-case prior distributions are non-Gaussian and non-symmetric, even though the initial prior was Gaussian, and therefore symmetric.

The following sections of this paper build upon the knowledge needed to understand the main concepts and algorithmic approximations required for the proposed approach.

3. Wasserstein Gradient Flow

In this paper, to be able to consider the optimisation of functionals with respect to probability measures, the Wasserstein Gradient Flow (WGF) (Santambrogio, 2015) concept is introduced. The WFG applies on a probability measures space where a 2-Wasserstein metric has been defined.

Let us first consider the functional $E(\rho)$, where $E:\mathcal{P}(\Omega)\to\mathbb{R}$ maps a probability measure to a real value, where the $\mathcal{P}(\Omega)$ is the space of probability measures on $\Omega\subset\mathbb{R}^D$, and D is the number of dimensions.

To investigate the optimisation of the functional $E(\rho)$ as a Wasserstein Gradient Flow, the Jordan Kinderleher Otto (JKO) scheme (Ambrosio et al., 2005; Santambrogio, 2016) is used. The JKO scheme solves the variational problem by defining the time discretization of the diffusion process, for this discretization the approximate probability density, ρ_{i+1}^{τ} at the $i+1$ timestep is calculated:

$$\rho_{i+1}^{\tau}=\arg\min_{\rho}\left\{E(\rho)+\frac{\mathcal{W}_2^2\left(\rho,\rho_i^{\tau}\right)}{2\tau}\right\} \tag{4}$$

Where $\mathcal{W}_2$ is the 2-Wasserstein distance, $\tau>0$ is the size of the timestep, and as the size of the timestep approaches zero $\tau\to 0$, the expression above converges to the exact Wasserstein Gradient Flow. The 2-Wasserstein distance (curve length between two distributions) is defined as (Santambrogio, 2015):

$$\mathcal{W}_2^2(\mu,\nu)=\inf_{\gamma\in\Gamma(\mu,\nu)}\int_{\Omega\times\Omega}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\right\|_2^2\gamma\left(d\boldsymbol{\theta},d\boldsymbol{\theta}^*\right) \tag{5}$$

Where, γ is the deterministic coupling that minimises equation (5) and γ is inside the set of all possible couplings or joint distributions $\Gamma(\mu,\nu)$ over $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, where μ and ν are the marginal distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$. In the context of transport optimization, the calculation of the 2-Wasserstein distance can be interpreted as the transformation of elements in the domain μ to the domain ν at a minimum cost. Then, from this perspective, in the equation (5) of the 2-Wasserstein distance, $\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\right\|_2^2$ is the transportation cost of $\boldsymbol{\theta}$ in μ, to $\boldsymbol{\theta}^*$ in ν (Santambrogio, 2015). By defining the cost function c as $\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\right\|_2^2$, the equation (5) can be rewritten to:

$$\mathcal{W}_2^2(\mu,\nu):=\inf_{\mathcal{T}}\int_{\Omega}c\left(\boldsymbol{\theta},\mathcal{T}(\boldsymbol{\theta})\right)d\mu(\boldsymbol{\theta}) \tag{6}$$

For the cases where there is a unique solution for the problem of minimum transportation cost from

$\boldsymbol{\theta}$ in μ, to $\boldsymbol{\theta}^*$ in ν, the unique solution can also be expressed as a mapping $\mathcal{T}:\mathbb{R}^D \to \mathbb{R}^D$, that pushes elements $\boldsymbol{\theta}$ of the domain μ to the domain ν (C. Chen et al., 2018). The solution is unique when the marginal distribution of probability μ is absolutely continuous w.r.t. the Lebesgue measure (C. Chen et al., 2018).

If $\{\mu_t\}_{t\in[0,1]}$ is an absolutely continuous curve with finite second-order moments in the probabilistic space $\mathcal{P}(\Omega)$, then the changes of μ_t in that curve will be defined through investigation of $\mathcal{W}_2^2(\mu_t, \mu_{t+\tau})$. Studying the changes of μ_t, is related to the original JKO problem (Ambrosio et al., 2005; Santambrogio, 2016) of the minimisation of the functional shown in equation (4). These changes can be described using a velocity field given by: $\boldsymbol{v}_t(\boldsymbol{\theta}) := \lim_{\tau\to 0} \frac{\mathcal{T}(\boldsymbol{\theta}_t) - \boldsymbol{\theta}_t}{\tau}$. This velocity field $\boldsymbol{v}_t(\boldsymbol{\theta})$ defines in $\mathcal{P}(\Omega)$ the gradient flow (Ambrosio et al., 2005):

$$\partial_t \mu_t + \nabla \cdot (\boldsymbol{v}_t \mu_t) = 0 \tag{7}$$

Solving the equation (7) requires finding a velocity field $\boldsymbol{v}(t)$ such that its flow agrees with $\lim_{\tau\to 0}(\boldsymbol{\theta}_\tau(t))$. The WGF can be shown to have a velocity field $\boldsymbol{v}(t)$ that minimises the functional $E(\rho)$, with the following form $\boldsymbol{v}(t) = -\nabla \frac{\partial E(\rho)}{\partial \rho}$ (Ambrosio et al., 2005), where $\frac{\partial E(\rho)}{\partial \rho}$ is called the first variation of $E(\rho)$ at ρ. Based on this, the WGF may be expressed as:

$$\partial_t \rho_t = -\nabla \cdot (\boldsymbol{v}_t \rho_t) = \nabla \cdot \left(\rho_t \nabla \frac{\partial E(\rho_t)}{\partial \rho_t} \right) \tag{8}$$

Therefore, to derive the WGF for the optimisation of the functional $E(\rho)$ the following requirements are introduced:

1. The first variation of the functional $E(\rho)$ with respect to the density $\frac{\partial E(\rho)}{\partial \rho}$ needs to be calculated.

2. A perturbation that follows the formal definition of a derivative in the probability space has to be introduced.

3. The probability ρ is a probability density $\rho \in \mathcal{P}(\Omega)$ that has to be perturbed to $\rho + \varepsilon\chi$ which is also another probability density such that it also lies in the probability space $\mathcal{P}(\Omega)$, in this way $E(\rho + \varepsilon\chi)$ is well defined.

4. For all small $\varepsilon > 0$, both, the perturbed probability density is defined in the probability space $\rho + \varepsilon\chi \in \mathcal{P}(\Omega)$, and $\sigma = \rho + \chi \in \mathcal{P}(\Omega)$.

This can also be rewritten as $\rho + \varepsilon\chi = \rho + \varepsilon(\sigma - \rho) = \rho(1-\varepsilon) + \varepsilon\sigma$, where $\rho(1-\varepsilon) + \varepsilon\sigma \in \mathcal{P}(\Omega)$, as long as ρ and σ are also probabilities densities.

Now that the requirements have been introduced, the first variation of $E(\rho)$, $\frac{\partial E(\rho)}{\partial \rho}$ can be found, and it is given as (Ambrosio et al., 2005; Santambrogio, 2016):

$$\frac{\partial}{\partial \varepsilon} E(\rho + \varepsilon\chi)\Big|_{\varepsilon=0} = \int_{\Omega} \frac{\partial E(\rho)}{\partial \rho} \chi(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{9}$$

for all $\chi = \sigma - \rho$. If a constant C is added, $\int_{\Omega} \left(\frac{\partial E(\rho)}{\partial \rho} + C \right) \chi(\boldsymbol{\theta}) d\boldsymbol{\theta}$, it can be found that the first variation may be defined uniquely only up to additive constants, as that second integral $\int_{\Omega} \chi(\boldsymbol{\theta}) d\boldsymbol{\theta}$ includes the difference of 2 probability densities $\chi = \sigma - \rho$.

4. Wasserstein Gradient Flow for Bayesian Inference

Approximations to the posterior can be obtained using many different methods. Recently, methods based on Variational Inference (VI) have been gaining popularity (Blei et al., 2017). These methods are based on the minimisation of the KL divergence between the posterior $p(\boldsymbol{y}_{obs}|\boldsymbol{\theta})$ and a probability density (usually parametric) defined inside a family of distributions $\mathcal{Q}$, to quantify the degree of

dissimilarity between two distributions over the same domain:

$$\rho^{*} = \arg\min_{\rho \in \mathcal{Q}} KL(\rho \| p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})) \tag{10}$$

where the KL divergence is defined as:

$$KL\left(\rho \mid p\left(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs}\right)\right) = \int_{\Omega} \rho \log\left(\frac{\rho}{p\left(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs}\right)}\right) d\boldsymbol{\theta} \tag{11}$$

The approximation to the posterior is obtained by finding the member of the family and its respective hyperparameters that best minimise the KL divergence (Blei et al., 2017).

An alternative to VI would be to use the Wasserstein Gradient Flow to define an iterative procedure that uses the set of data $\boldsymbol{y}_{obs}$ to update a chain of $\rho_n(\boldsymbol{\theta})$ with the purpose of approximating $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs})$ given the minimisation of a suitable functional $E(\rho)$. In WGF, the optimisation of the functional can be solved by using equation (8). To be able to solve this equation, the velocity field $\boldsymbol{v}(t)$ given the chosen functional is required. In a first analysis, it may be thought that as the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs})$ is not known in advance, (because of the presence of the normalization constant $p(\boldsymbol{y}_{obs})$), then the functional of equation (11) cannot be used to derive a WGF. But as the first variation of the functional is only uniquely defined up to additive constants, a simpler functional $E(\rho)$ where the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs})$ is replaced for the unnormalized posterior $p(\boldsymbol{\theta}, \boldsymbol{y}_{obs})$ may be used (Gao & Liu, 2020). Therefore, the velocity field that results from replacing the posterior with the unnormalized posterior would be the same as the velocity field as in the functional in equation (11).

By obtaining a WGF of the functional $E(\rho)$, the partial differential equation can be solved to flow the approximation to the posterior $\rho_t(\boldsymbol{\theta})$ to its equilibrium $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs})$ for the observed data. The dynamic system is defined by an initial density $\rho_0(\boldsymbol{\theta})$ that is given by the prior $p(\boldsymbol{\theta})$, and $\rho_\infty(\boldsymbol{\theta})$ tends to the posterior distribution, (Gao & Liu, 2020). In a more rigorous manner, in a manifold M in the parameter space, a pushforward density $\rho_t(\boldsymbol{\theta}) = \mathcal{T}_t \# p(\theta) \in M$ is considered, where $\#$ is the

push forward operator, and the best curve (under certain restrictions) ρ_t, that drives ρ_0 to ρ_∞ has to be found (Gao & Liu, 2020).

The WGF of the chosen functional $E(\rho)$, may be performed by first calculating the first variation (where the bounds are omitted for clarity):

$$\frac{\partial}{\partial\varepsilon}E(\rho+\varepsilon\chi)\Big|_{\varepsilon=0}=\frac{\partial}{\partial\varepsilon}\Big[\int(\rho+\chi\varepsilon)\log\big((\rho+\chi\varepsilon)\big)d\boldsymbol{\theta}-\int(\rho+\chi\varepsilon)\log\big(p(\boldsymbol{\theta},\boldsymbol{y}_{obs})\big)d\boldsymbol{\theta}\Big]\Big\|_{\varepsilon=0}=$$

$$=\Bigg[\int\chi\log(\rho+\chi\varepsilon)d\boldsymbol{\theta}+\int(\rho+\chi\varepsilon)\frac{\chi}{(\rho+\chi\varepsilon)}d\boldsymbol{\theta}-\int(\rho+\chi\varepsilon)\log\big(p(\boldsymbol{\theta},\boldsymbol{y}_{obs})\big)d\boldsymbol{\theta}\Bigg]\Bigg\|_{\varepsilon=0}=$$

$$=\int\chi\log(\rho)d\boldsymbol{\theta}+\int\chi d\boldsymbol{\theta}-\int\chi\log\big(p(\boldsymbol{\theta},\boldsymbol{y}_{obs})\big)d\boldsymbol{\theta}=\int\Big(\log(\rho)+1-\log\big(p(\boldsymbol{\theta},\boldsymbol{y}_{obs})\big)\Big)\chi d\boldsymbol{\theta} \tag{12}$$

The first variation of the functional $E(\rho)$ with respect to the density ρ is then given by:

$$\frac{\partial E(\rho)}{\partial\rho}=\log(\rho)+1-\log\big(p(\boldsymbol{\theta},\boldsymbol{y}_{obs})\big) \tag{13}$$

and the velocity field is:

$$\boldsymbol{v}(t)=-\nabla\frac{\partial E(\rho)}{\partial\rho}=\nabla\Big(\log\big(p(\boldsymbol{\theta},\boldsymbol{y}_{obs})\big)-\log(\rho)-1\Big)=\nabla\log\big(p(\boldsymbol{\theta},\boldsymbol{y}_{obs})\big)-\nabla\log(\rho) \tag{14}$$

If the first variation of the functional $E(\rho)$ is introduced into the continuity equation, the following equation is obtained (Wang et al., 2022):

$$\partial_t\rho_t=\nabla\cdot\Big(\rho_t\big(\nabla\log\big(p(\boldsymbol{\theta},\boldsymbol{y}_{obs})\big)-\nabla\log(\rho_t)\big)\Big) \tag{15}$$

The KL WGF is an approximation in continuous time of the deterministic mean-field particle system called mean-field Wasserstein dynamics (Wang et al., 2022):

$$d\boldsymbol{\theta}_t=\Big[\nabla\log\big(p(\boldsymbol{\theta},\boldsymbol{y}_{obs})\big)-\nabla\log(\rho_t)\Big]dt \tag{16}$$

The mean-field term is derived from the fact that the dynamics' evolution varies with the current density function ρ_t. The deterministic particle descent WGF may be obtained from the mean-field Wasserstein dynamics (Wang et al., 2022):

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha_t \left(\nabla \log\left(p\left(\boldsymbol{\theta}, \boldsymbol{y}_{obs}\right)\right) - \nabla \log\left(\rho_t\right)\right) \tag{17}$$

The equation (17) represents one of the two particle discretisation WGFs equations needed for the simultaneous optimization of the chosen functional in equation (3) shown in Section 2. In equation (17), an approximation of $\nabla \log\left(\rho_t\right)$ is required, as no analytical expression is available. Many different methods may be used to obtain an approximation. In this paper, a kernel density estimate (KDE) approach is chosen and explained in subsection 5.1. It should be noted that the WGD follows a deterministic rule for the updating, and therefore the initial positions of the system determine the particle interactions and randomness.

5. Approximations in Wasserstein Gradient Flow for Robust Bayesian Inference

This section provides a more detailed explanation of some of the mathematical tools required for the application of the algorithm described in fig. 4.

5.1. Approximation to $\nabla_{\boldsymbol{\theta}} \log(\rho)$ from samples

When the velocity field $\boldsymbol{v}(t)$ is to be approximated, one of the difficulties that arises is the estimation of $\nabla \log \rho(\boldsymbol{\theta})$ (C. Liu et al., 2018). Only a finite set of samples $\left\{\theta^{(i)}\right\}_{i=1}^{N}$ of $\rho(\boldsymbol{\theta})$ is known. However, a direct approximation of $\rho(\boldsymbol{\theta})$ using the empirical distribution $\hat{\rho}(\boldsymbol{\theta}) := \frac{1}{N}\sum_{i=1}^{N} \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(i)}\right)$, where δ is the Dirac delta function is not possible. The reason why that direct approximation cannot be performed is because the WGF of the KL divergence at $\hat{\rho}(\boldsymbol{\theta})$ is not defined, consequence of $\hat{\rho}(\boldsymbol{\theta})$ not been absolutely continuous. Using the absolutely continuous approximated expression $\tilde{\rho}(\boldsymbol{\theta}) := (\hat{\rho} * K)(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} K\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i)}\right)$ ("$*$" symbolizes convolution), the velocity field $\boldsymbol{v}(t)$ can be well-defined by smoothing $\hat{\rho}(\boldsymbol{\theta})$ through a smooth kernel K on $\boldsymbol{\theta}$.

In this paper, the approximation of $\rho(\boldsymbol{\theta})$ is produced using the KDE $\tilde{\rho}(\boldsymbol{\theta})$, where

$K\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(i)}\right):\mathbb{R}^D\times\mathbb{R}^D\to\mathbb{R}$ is a given positive and differentiable kernel function, and the Gaussian kernel is used:

$$K\left(\boldsymbol{\theta},\boldsymbol{\theta}^{*}\right)=\left(2\pi h\right)^{-\frac{N}{2}}\exp\left(-\frac{\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^{*}\right\|_2^2}{2h}\right) \tag{18}$$

N is the number of samples used to define the kernel function $K\left(\boldsymbol{\theta},\boldsymbol{\theta}^{*}\right)$, h is the bandwidth and it is defined by $h=\text{med}^2/\log(\text{N})$, and med represents the median of distances of the samples (Q. Liu & Wang, 2016).

When the KDE is used as an approximation of $\rho(\boldsymbol{\theta})$, the following expression may be used to calculate an approximation of $\nabla\log\rho(\boldsymbol{\theta})$ (Wang et al., 2022):

$$\nabla\log\tilde{\rho}(\boldsymbol{\theta})=\frac{\nabla\tilde{\rho}(\boldsymbol{\theta})}{\tilde{\rho}(\boldsymbol{\theta})}=\frac{\sum_{i=1}^{N}\nabla_{\boldsymbol{\theta}}K\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(i)}\right)}{\sum_{i=1}^{N}K\left(\boldsymbol{\theta},\boldsymbol{\theta}^{(i)}\right)} \tag{19}$$

The kernel chosen does not affect the solution of the Gradient Flow if the size of the ensemble tends to infinity (J. Lu et al., 2019). Nonetheless, the distribution of particles for a finite number of them, may not be unique. An alternative manner to explain this is that for different given kernels, that is, with different particle flows, different results (final positions in the state space of the particles) are obtained. However, for those kernels as their number of particles increases, the representation of the posterior probability density function (pdf) becomes more accurate.

5.2. Approximation to $\nabla_{\boldsymbol{\theta}}\log\left(p\left(\boldsymbol{\theta},\boldsymbol{y}_{obs}\right)\right)$

In this paper, two different ways to estimate the gradient of log likelihood are considered. The first one uses local estimations of that Jacobian matrix of the model's ensemble, whereas the second one uses gaussian processes. The first approach is only able to obtain estimates of the gradient of log likelihood at particle positions where the model has been run previously. However, the Gaussian

Process approach is able to obtain estimates of the gradient of log likelihood at particle positions that have not been evaluated by leveraging on the prior assumptions and previous model runs. The choice of approach is usually based on the computational cost of dealing with the physics-based model involved.

In general $\nabla_{\boldsymbol{\theta}} \log\left(p(\boldsymbol{\theta})\right)$ can be calculated analytically, as most of the $\log\left(p(\boldsymbol{\theta})\right)$ chosen in Bayesian Inference are differentiable. However, if an analytical expression is not available, $\nabla_{\boldsymbol{\theta}} \log\left(p(\boldsymbol{\theta})\right)$ may be approximated using equation (19) as long as samples from the prior are available.

5.2.1. Gradient of log likelihood using ensemble method

Assuming a multivariate Gaussian likelihood, $p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta})$ can be written as:

$$p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp\left(-\frac{1}{2}\left(\boldsymbol{y}_{\text{obs}} - \boldsymbol{y}_{\text{model}}\right)^T \Sigma^{-1} \left(\boldsymbol{y}_{\text{obs}} - \boldsymbol{y}_{\text{model}}\right)\right) \tag{20}$$

In the above expression, d refers to the dimensionality of the observation space (the number of observations), $\boldsymbol{y}_{\text{model}}$ and $\boldsymbol{y}_{\text{obs}}$ respectively are the $n \times 1$ vectors of simulated and observed states, and the inverse of the $n \times n$ error covariance matrix Σ is denoted by Σ^{-1}.

By taking the gradient of the logarithm of the multivariate Gaussian likelihood, the below expression is obtained:

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta}) = \frac{1}{2} \nabla_{\boldsymbol{\theta}} \boldsymbol{y}_{\text{model}}{}^T \Sigma^{-1} \left(\boldsymbol{y}_{\text{obs}} - \boldsymbol{y}_{\text{model}}\right) \tag{21}$$

In equation (21), $\nabla_{\boldsymbol{\theta}} \boldsymbol{y}_{\text{model}}$ is a matrix of dimensions $n \times D$. The number of model parameters is denoted by D. The elements of the $\nabla_{\boldsymbol{\theta}} \boldsymbol{y}_{\text{model}}$ matrix are the partial derivatives of each simulated state (associated to rows 1, …, n) w.r.t. each parameter (associated to columns 1, …, D). The states are simulated introducing input parameters $\boldsymbol{\theta}$ into a computational model:

$$\boldsymbol{y}_{\text{model}} = PM(\boldsymbol{x}, \boldsymbol{\theta}) \tag{22}$$

The expression above assumes that the observed states are directly simulated by the model. If the Jacobian is defined as $J(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} PM(\boldsymbol{x}, \boldsymbol{\theta})^T$, the matrix of dimensions $(D \times n)$, equation (21) can be rewritten as:

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta}) = \frac{1}{2} J(\boldsymbol{\theta})^T \Sigma^{-1} (\boldsymbol{y}_{\text{obs}} - \boldsymbol{y}_{\text{model}}) \tag{23}$$

As a result, using the equation (23), the log likelihood gradient may be evaluated using local estimations of that Jacobian matrix. Computational difficulties arise during the evaluation of the Jacobian matrix $J(\boldsymbol{\theta})$ of dimensions $(n \times D)$, as the closed form of this matrix is frequently unavailable.

To solve the mentioned difficulty, an approach that consists of obtaining nonintrusive estimations of the Jacobian $J(\boldsymbol{\theta})$ may be taken (Ramgraber et al., 2021). The vector $\boldsymbol{\theta}$, that has the parameters as elements, is perturbed in a small increment in each of its D dimensions, and the Jacobian matrix $J(\boldsymbol{\theta})$ is approximated using the obtained two or three-points finite difference derivatives. This computational differentiation may produce very accurate results, but it becomes unpractical if the model has a high number of parameters D. If the ensemble size or number of particles is denoted N, and a set of local Jacobians is to be required, the model has to be run $(D+1)N$ times if two-points finite difference derivatives are used, or even more times $(2D+1)N$, if three-points finite difference derivatives are chosen (Ramgraber et al., 2021). Those numbers are well above the number of evaluations of the model that practitioners may consider affordable.

A technique that requires only N model evaluations $PM(\boldsymbol{x}, \boldsymbol{\theta})$, and it is able to produce the estimation of the Jacobian matrix $J(\boldsymbol{\theta})$, directly from the ensemble, may be found in (Ramgraber et al., 2021). This methodology makes use of the relative differences between particles:

$$\tilde{J}(\boldsymbol{\theta}_r) = \frac{P}{N} \sum_{r=1}^{N} \frac{PM(\boldsymbol{\theta}_r) - PM(\boldsymbol{\theta}_s)}{\left\| PM(\boldsymbol{\theta}_r) - PM(\boldsymbol{\theta}_s) \right\|} \cdot \frac{\left\| PM(\boldsymbol{\theta}_r) - PM(\boldsymbol{\theta}_s) \right\|}{\left\| \boldsymbol{\theta}_r - \boldsymbol{\theta}_s \right\|} \cdot \frac{\boldsymbol{\theta}_r^T - \boldsymbol{\theta}_s^T}{\left\| \boldsymbol{\theta}_r - \boldsymbol{\theta}_s \right\|} \tag{24}$$

In the equation (24) above, P is the rank expected for the Jacobian matrix $J(\boldsymbol{\theta})$, this expected rank is the smallest value between $N-1$ and D. Inside the summation symbol three fractions are found, in correlative order: the vector from particle $\boldsymbol{\theta}_r$ to the particle $\boldsymbol{\theta}_s$ (normalized), the scalar gradient between the observation and the parameter space, the normalized vector in parameter space. The equation (24) may be simplified as follows (Ramgraber et al., 2021):

$$\tilde{J}(\boldsymbol{\theta}_r) = \frac{P}{N}\sum_{r=1}^{N}\frac{\left(PM(\boldsymbol{\theta}_r) - PM(\boldsymbol{\theta}_s)\right)\left(\boldsymbol{\theta}_r^T - \boldsymbol{\theta}_s^T\right)}{\left\|\boldsymbol{\theta}_r - \boldsymbol{\theta}_s\right\|^2} \tag{25}$$

The factor $\frac{P}{N}$ external to the sum is made up of a correction factor P to consider that the maximum possible contribution of each vector to the rank of the Jacobian is one, and a factor $\frac{1}{N}$ to account for an arithmetical average. For $N \to \infty$ and an isotropic arrangement of particles, the Jacobian in equation (25) should converge against the correct one (Ramgraber et al., 2021).

5.2.2. Gradient of log likelihood using Gaussian process

For cases where the physics-based model is expensive-to-evaluate, an approximation of the gradient of the log likelihood may be produced using a Gaussian process. This methodology allows the estimation of the gradient at particle positions where the physics-based model has not been evaluated.

Assuming that the likelihood function is given by a multivariate Gaussian with zero error mean and covariance Σ, the log likelihood function is:

$$\log p(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{\theta}) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log(\det\Sigma) - \frac{1}{2}\left(\boldsymbol{y}_{\text{obs}} - \boldsymbol{y}_{\text{model}}\right)^T \Sigma^{-1}\left(\boldsymbol{y}_{\text{obs}} - \boldsymbol{y}_{\text{model}}\right) \tag{26}$$

Focus is placed on the last term of the equation (26), as the gradient of the log likelihood function w.r.t. the parameter $\boldsymbol{\theta}$ only depends on that term. Consequently, the partially observed potential is modelled as (Dunbar et al., 2022):

$$V_L(\theta) = \frac{1}{2}\left(\boldsymbol{y}_{\text{obs}} - \boldsymbol{y}_{\text{model}}\right)^T \Sigma^{-1}\left(\boldsymbol{y}_{\text{obs}} - \boldsymbol{y}_{\text{model}}\right) \tag{27}$$

Where $V_L(\theta)$ is a Gaussian process $f \sim GP(0,k)$, and k denotes a positive definite kernel on $\mathbb{R}^D$ that has been chosen according to the explanations below.

In this paper, k is a Gaussian radial basis function kernel that has the form $k\left(\boldsymbol{\theta},\boldsymbol{\theta}^*;\lambda,l\right)=\lambda\exp\left(-\frac{\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\right\|^2}{2l^2}\right)$. In this expression, $l>0$ denotes the kernel bandwidth, and $\lambda>0$ is the amplitude of the kernel. A function f is sought so that for some $\sigma>0$, and for some noisy evaluations of the potential at the ensemble of points $\Theta_t=\left(\Theta_t^1,\ldots,\Theta_t^N\right)\in\mathbb{R}^{N\times D}$, then (Dunbar et al., 2022):

$$V_L\left(\Theta_t^i\right)=f\left(\Theta_t^i\right)+\sigma\xi^i,\quad \xi^i=\left(\xi^1,\ldots,\xi^N\right)\sim\mathcal{N}\left(0,I\right) \tag{28}$$

The mean function of the associated Gaussian process posterior for f is (Rasmussen, 2003):

$$\mu\left(\theta^*\right)=k\left(\theta^*,\Theta\right)K\left(\Theta,\Theta\right)^{-1}V_L\left(\Theta\right) \tag{29}$$

and the expression for the variance function is (Rasmussen, 2003):

$$\sigma^2\left(\theta^*\right)=k\left(\theta^*,\theta^*\right)-k\left(\theta^*,\Theta\right)K\left(\Theta,\Theta\right)^{-1}k\left(\Theta,\theta^*\right) \tag{30}$$

Where $K\left(\Theta,\Theta\right)=diag\left(\sigma^2\right)+k\left(\Theta,\Theta\right)$. Equations (31) and (32) express the well-defined gradient of the posterior mean (Rasmussen, 2003):

$$\mathbb{E}\left[\frac{\partial V_L\left(\boldsymbol{\theta}^*\right)}{\partial\boldsymbol{\theta}_d^*}\right]=\frac{\partial\mathbb{E}\left[V_L\left(\boldsymbol{\theta}^*\right)\right]}{\partial\boldsymbol{\theta}_d^*}=\frac{\partial k\left(\theta^*,\Theta\right)}{\partial\boldsymbol{\theta}_d^*}K\left(\Theta,\Theta\right)^{-1}V_L\left(\Theta\right) \tag{31}$$

$$\nabla_{\boldsymbol{\theta}}V_L\left(\Theta\right)=\left.\frac{\partial k\left(\boldsymbol{\theta}^*,\Theta\right)}{\partial\boldsymbol{\theta}_d^*}\right|_{\theta^*=\Theta}K\left(\Theta,\Theta\right)^{-1}V_L\left(\Theta\right) \tag{32}$$

Both the energy term $V_L(\theta)$ and the hyperparameters (σ,λ,l) are updated at each iteration, and are calculated considering the new incoming data (Rasmussen, 2003).

5.3. Derivations of Wasserstein Gradient Flow equations for optimal or worst-case prior

In section 4, the WGF for the case when the approximation to the posterior is made to vary to minimize the KL divergence between the posterior, and the approximation to the posterior has been derived. Now, the WGF that either maximises or minimises the KL divergence between the posterior and the approximation to the posterior with respect to the prior needs to be calculated. Currently, the interacting WGF has the following form:

$$\begin{cases} \partial_t \rho_t = \nabla \cdot \left(\rho_t \left(\nabla \log\left(p\left(\boldsymbol{\theta}, \boldsymbol{y}_{obs}\right)\right) - \nabla \log\left(\rho_t\right)\right)\right) \\ \partial_t p_t\left(\boldsymbol{\theta}\right) = \eta\left(-\nabla \cdot \left(p_t\left(\boldsymbol{\theta}\right) \nabla \left(\frac{\partial E}{\partial p\left(\boldsymbol{\theta}\right)}\left(\rho, p\left(\boldsymbol{\theta}\right)\right)\right)\right)\right) \end{cases} \tag{33}$$

The first step is to calculate the first variation of the functional $E\left(\rho\left(\boldsymbol{\theta}\right), p\left(\boldsymbol{\theta}\right)\right)$ with respect to the prior $p\left(\boldsymbol{\theta}\right)$. When the optimal prior is of interest, this results in the minimization of equation (3), to obtain an expression of the first variation we first need to calculate the following:

$$\frac{\partial}{\partial \varepsilon} E\left(p\left(\boldsymbol{\theta}\right) + \varepsilon\chi\right)\Big|_{\varepsilon=0} = \frac{\partial}{\partial \varepsilon}\left[\int \rho \log \rho d\boldsymbol{\theta} - \int \rho \log\left(p\left(\boldsymbol{\theta}\right) + \varepsilon\chi\right) d\boldsymbol{\theta} - \int \rho \log\left(p\left(\boldsymbol{\theta} \mid \boldsymbol{y}_{obs}\right)\right) d\boldsymbol{\theta}\right]\Big|_{\varepsilon=0} =$$

$$= \left[-\int \rho \frac{\chi}{\left(p\left(\boldsymbol{\theta}\right) + \chi\varepsilon\right)} d\boldsymbol{\theta} \right]\Big|_{\varepsilon=0} = -\int \frac{\rho}{p\left(\boldsymbol{\theta}\right)} \chi d\boldsymbol{\theta} \tag{34}$$

Now an expression of the first variation of the functional to be optimized can be obtained and it is given by:

$$\frac{\partial E}{\partial p\left(\boldsymbol{\theta}\right)}\left(p\left(\boldsymbol{\theta}\right)\right) = -\frac{\rho}{p\left(\boldsymbol{\theta}\right)} \tag{35}$$

and the velocity field is:

$$\boldsymbol{v}\left(t\right) = -\nabla \frac{\partial E}{\partial \rho}\left(\rho\right) = \nabla\left(\frac{\rho}{p\left(\boldsymbol{\theta}\right)}\right) = \frac{p\left(\boldsymbol{\theta}\right)}{p\left(\boldsymbol{\theta}\right)} \nabla\left(\frac{\rho}{p\left(\boldsymbol{\theta}\right)}\right) = \frac{\rho}{p\left(\boldsymbol{\theta}\right)}\left(\nabla \log \rho - \nabla \log p\left(\boldsymbol{\theta}\right)\right) \tag{36}$$

The resulting particle-based Wasserstein Gradient Flow, using an Euler discretisation, is given as:

$$\boldsymbol{\theta}_{prior,t+1}^{N} = \boldsymbol{\theta}_{prior,t}^{N} + \tau_t \left(\frac{\rho_t\left(\boldsymbol{\theta}_{prior}\right)}{p_t\left(\boldsymbol{\theta}_{prior}\right)} \left(\nabla \log p_t\left(\boldsymbol{\theta}_{prior}\right) - \nabla \log \rho_t\left(\boldsymbol{\theta}_{prior}\right) \right) \right) \tag{37}$$

If the maximization of the KL divergence is sought instead, this requires the calculation of the worst-case prior, and the resulting velocity field is given as the negative of the previously calculated velocity field:

$$\boldsymbol{v}(t) = \frac{\rho}{p(\boldsymbol{\theta})} \left(\nabla \log p(\boldsymbol{\theta}) - \nabla \log \rho \right) \tag{38}$$

Therefore, the resulting particle-based Wasserstein Gradient Flow using an Euler discretisation is given as:

$$\boldsymbol{\theta}_{prior,t+1}^{N} = \boldsymbol{\theta}_{prior,t}^{N} - \tau_t \left(\frac{\rho_t\left(\boldsymbol{\theta}_{prior}\right)}{p_t\left(\boldsymbol{\theta}_{prior}\right)} \left(\nabla \log p_t\left(\boldsymbol{\theta}_{prior}\right) - \nabla \log \rho_t\left(\boldsymbol{\theta}_{prior}\right) \right) \right) \tag{39}$$

Now that the particle based WGF for the minimisation or maximisation of the functional with respect to the prior has been derived, an interacting particle based WGF can be defined as follows:

$$\begin{cases} \boldsymbol{\theta}_{t+1}^{N} = \boldsymbol{\theta}_{t}^{N} + \alpha_t \left(\nabla \log\left(p_t\left(\boldsymbol{\theta}, \boldsymbol{y}_{obs}\right) \right) - \nabla \log\left(\rho_t\right) \right) \\ \boldsymbol{\theta}_{prior,t+1}^{N} = \boldsymbol{\theta}_{prior,t}^{N} \pm \tau_t \left(\frac{\rho_t\left(\boldsymbol{\theta}_{prior}\right)}{p_t\left(\boldsymbol{\theta}_{prior}\right)} \left(\nabla \log p_t\left(\boldsymbol{\theta}_{prior}\right) - \nabla \log \rho_t\left(\boldsymbol{\theta}_{prior}\right) \right) \right) \end{cases} \tag{40}$$

The resulting simultaneous equations (40) are composed of: a) the top equation which is the particle discretisation of the WGF that results from the minimisation of the KL divergence between the posterior and the approximation to the posterior w.r.t. the approximation to the posterior, and b) the bottom equation that results from either minimising or maximising the KL divergence between the posterior and the approximation to the posterior w.r.t. the prior. These simultaneous equations may be used to obtain the prior that either maximises or minimises the functional, and their resulting approximations to the posterior. For the case where the step size τ_t of the bottom equation in the simultaneous equation (40) is zero, the original particle based WGF for Bayesian Inference would be recovered, as this would mean the prior is static (it does not change with time).

5.4. Density ratio estimation from samples

The equations (37) and (39), require the calculation of the pdf of the $\rho_t(\boldsymbol{\theta}_{prior})$ and the pdf of $p_t(\boldsymbol{\theta}_{prior})$. This may be done for example using kernel density estimates. In this paper, rather than doing the direct estimation of the densities, the density ratio is calculated directly:

$$g(\boldsymbol{\theta}_{prior}) = \frac{\rho_t(\boldsymbol{\theta}_{prior})}{p_t(\boldsymbol{\theta}_{prior})} \tag{41}$$

Numerous methods have been developed for the calculation of the density ratio in equation (41), the method chosen in this paper is the one called Relative unconstrained Least-Squares Importance Fitting (RuLSIF), and the interested reader can find it in (Yamada et al., 2011).

6. Data and numerical models

In this section, the proposed method is validated using two numerical examples. These two case studies have been selected to showcase the applicability of the proposed approach to deal with problems of different complexity, and an engineering case study is included. In the first example, the 2D double banana posterior problem (Detommaso et al., 2018) is used to show the resulting particles obtained from the optimal and worst-case prior, and also the resulting particles from the approximation to the posterior. In the second example, a double beam system is used to show the differences between the ensemble method and the Gaussian process to numerically estimate the gradient of the logarithm of the likelihood at the particle positions.

In both case studies, the number of initial samples is $N_0 = 100$ for the approximation to the posterior and also the prior, and those initial samples are picked from identically and independently distributed draws from the nominal prior. Each iteration of the algorithm uses the same number of particles $(N = 100)$ and corresponds to evaluations of the physics-based model at the positions of those particles. As described in section 2, in the beginning of the method τ is set to zero until the number of iterations reaches $N_a = 50$.

The Gaussian kernel in equation (18), is used to produce the estimations of $\nabla \log \rho$ and $\nabla \log p(\boldsymbol{\theta})$, and the bandwidth is chosen using the median methodology.

If the distribution that is being optimised lies outside the ambiguity set, the size of the step in the particle flow algorithm is reduced to half until the distribution lies within the ambiguity set. Also, the distribution at iteration $i+1$ may be reset to a distribution from an earlier iteration $i-N_c$, where $N_c = 10$, if a preset number of distributions $(N_b = 5)$ are discarded when determining whether a distribution belongs to the ambiguity set. The maximum number of prior distributions that are allowed to be reset is $N_{reset} = 2$, once this number is reached, an additional number of iterations N_a are allowed. The total maximum allowed number of iterations $N_{\max} = 400$.

6.1. Double banana posterior example

This first example is based on the paper (Detommaso et al., 2018) that results in a two-dimensional double-banana-shaped posterior distribution. The equation that defines the model used is given by the logarithmic Rosenbrock function used in (Detommaso et al., 2018):

$$PM(\boldsymbol{\theta}) = \log\left((1-\theta_1)^2 + 100\left(\theta_2 - \theta_1^2\right)^2\right) \tag{42}$$

The initial prior chosen is a standard multivariate Gaussian, $\mathcal{N}(0, I)$. The numerical observation used to update the prior knowledge is obtained by $y_{obs} = PM(\boldsymbol{\theta}_{true}) + \zeta$, where $\boldsymbol{\theta}_{true}$ is a random variable drawn from the assumed prior, the standard deviation of the observational error is σ = 0.3, and $\zeta \sim \mathcal{N}(0, \sigma^2 I)$.

For the ambiguity set, the nominal prior is chosen to be the same as the initial prior. The statistical distance used is the 2-Wasserstein distance, and a radius $\varepsilon = 0.05$ has been chosen.

Using the algorithm inputs described above, the interacting Wasserstein Gradient Flows are used to find the resulting distributions for two different cases: a) the optimal prior and its resulting approximation to the posterior and b) the worst-case prior and its resulting approximation to the

posterior. In this example, the ensemble method described in section 5.2.1 is used to calculate an approximation to the gradient of the log likelihood at the particle positions to be evaluated.

The step sizes in the interacting particle flow WGF algorithm for the optimal prior case are $\alpha = 3*10^{-3}$ and $\tau = 1.5*10^{-3}$. For the worst-case prior case, the step sizes in the interacting particle flow WGF algorithm are $\alpha = 3*10^{-3}$ and $\tau = 3*10^{-4}$.

In this numerical case two different subcases are run, fig. 7 to fig. 13 correspond to the situations when the optimal prior and its approximation to the posterior are calculated. Figures 14 to 16 correspond to the situations when the worst-case prior and its approximation to the posterior are calculated.

In fig. 7 and fig. 8 respectively is shown, for each iteration, the positions of the particles from the optimal prior and from the approximation to the posterior. In both plots of fig. 7, it may be seen that after iteration $i = N_a$ the inner particles of the prior initially tend inwards, i.e., to the direction of smaller absolute values of parameters θ_1 and θ_2, this is due to the fact that the prior tries to move the closest particles to the positions of the particles of the approximation to the posterior. It can also be seen that the prior particles after a number of approximately 150 iterations do not change much of position, this occurs because of the step size decrease performed with the purpose of constraining the prior inside the defined ambiguity set. Figure 8 illustrates how the particle positions of the approximation to the posterior start moving into the regions of higher probability density. After iteration $i = N_a$, the particles' positions of the approximation to the posterior concentrate even more into regions of high probability density due to the prior having a greater effect on the positions of the particles.

Figures 9 and 10 show respectively, for each iteration, the values of the gradient of the logarithm of the prior and the gradient of the logarithm of the likelihood at the particles Θ_i^N positions. Figure 9 shows how the values of the gradient of the logarithm of the prior at the particle positions of the approximation to the posterior start to decrease as the prior particles start concentrating around the

particles of the approximation to the posterior. As the iterations progress, the values of the gradient of the logarithm of the prior at the approximation of the posterior particle positions start decreasing, this is because the particles of the prior become closer to the particles of the approximation to the posterior. This also means that the particles of the approximation to the posterior are becoming closer to regions of high prior density as the iterations progress. Figure 10 shows that during the initial iterations, high absolute values of the gradient of the logarithm of the likelihood may be found. This happens because during the initial stages of the algorithm there are particles that are still distant from the regions of high likelihood density. After around twenty to thirty iterations the values concentrate in a more defined region, even though some occasional extreme values can still be found.

In fig. 11, the initial particle positions (where the prior and approximation to the posterior particles are the same, shown in red), the final particle positions of the prior (black) and the approximation to the posterior (blue) can be seen. As expected, the final positions of the particles from the approximation to the posterior are shown to resemble the double-banana posterior in (Detommaso et al., 2018). It can also be observed that most of the final positions of the particles from the prior (optimal prior) are near the particles of the approximation to the posterior, and a smaller number of particles lie close to the initial prior particles. This means that the optimal prior assigns a high probability to the region close to the approximation to the posterior, and a lower probability to the outer particles far from the approximation to the posterior density.

Figure 12 shows a quiver plot, also known as vector plot, that is produced by the generic function `quiver` in (*MATLAB*, 2020). The scaling of the quiver function's default setting is chosen to prevent arrow length overlap. In this plot, the gradients of the logarithm of the prior, and the logarithm of the likelihood at the particle's positions from the approximation to the posterior in the final iteration are plotted. The gradients of the logarithm of the prior at the final prior particle positions are also shown.

Figure 13 illustrates the 2-Wasserstein distances at each iteration. Three plots can be found. The following distances at each iteration are plotted: the first is from the initial prior to the approximation

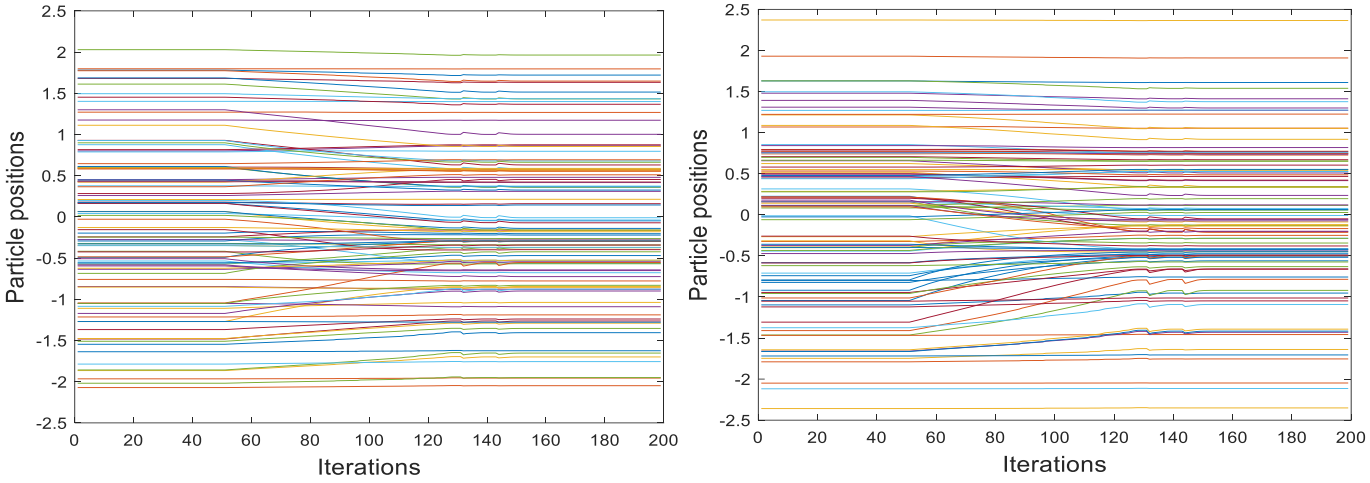
of the posterior; the second from the approximation to the posterior and the prior; and the third from the initial prior to the prior.

In fig. 14 and fig. 15 respectively, the positions of the particles from the worst-case prior, and from the approximation to the posterior, are shown for each iteration. In fig. 14 (a and b), it may be seen that the inner particles of the worst-case prior tend to move outwards to the direction of higher absolute values of parameters θ_1 and θ_2 as more iterations occur. Figure 14 also illustrates that after approximately 200 iterations the prior particles do not change much their positions. This is due to the decreasing size of the time step that is introduced with the purpose of limiting the prior inside the ambiguity set. In a similar manner to what occurs for the optimal prior case, in fig. 15 it can be observed that the particle positions of the approximation to the posterior also move to areas of higher probability density as iterations advance.

The final particle positions of the worst-case prior (black), initial particle positions (where the prior and approximation to the posterior particles are the same, shown in red), and its approximation to the posterior (blue) can be seen in fig. 16. As anticipated, the layout of the final positions of the particles from the approximation to the posterior takes a shape similar to the one shown by the double-banana posterior in (Detommaso et al., 2018). The worst-case prior assigns a higher density to areas of a low posterior density and vice versa. In a manner consistent with the previous statement, fig. 16 also shows that most of the final positions of the particles from the prior (worst-case prior), are positioned away from the final positions of the approximation to the posterior.

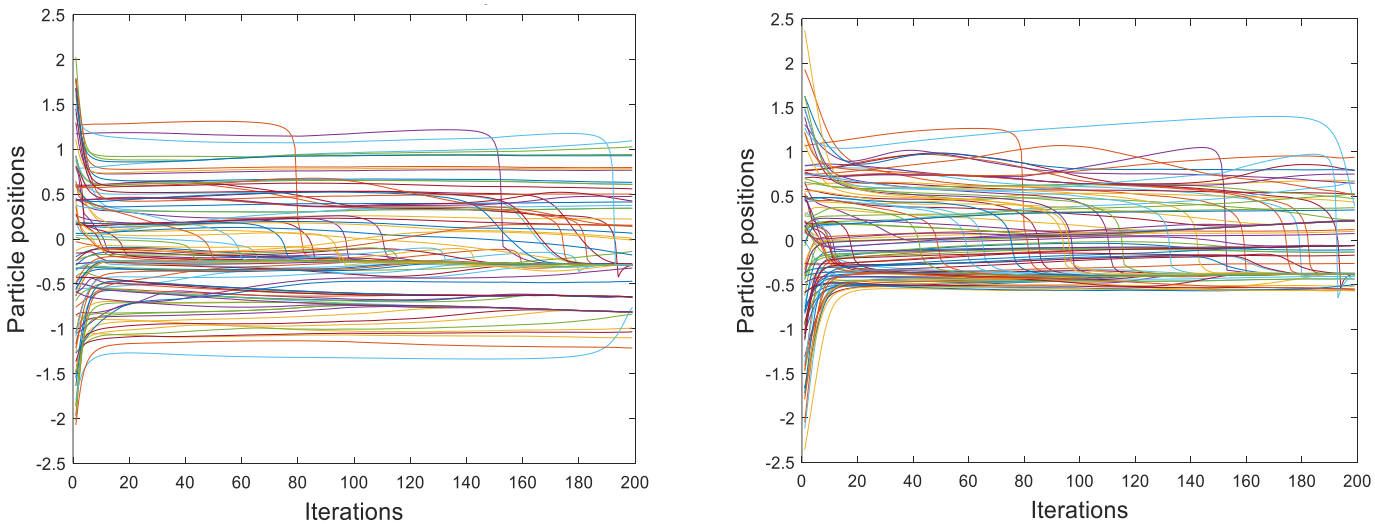
A direct comparison of the optimal prior and worst-case prior in the form of scatter plots and histograms of the latent variables is found on fig. 17. Figure 17 has been produced using the `plotmatrix` function from (*MATLAB*, 2020). It can be clearly seen that the optimal and worst-case prior differ from the initial prior and are no longer Gaussian. A very similar support can be seen of the optimal prior w.r.t. worst-case prior.

Scatter plots and histograms can also be found on fig. 18. For the cases where the optimal prior and worst-case prior have been estimated, the scatter plots and the histograms of the latent variables of the resulting approximation to the posterior are plotted. Very small differences are found when comparing between the resulting approximations to the posteriors. This is a consequence of the small sensitivity of the posterior to changes of the considered uncertain prior.



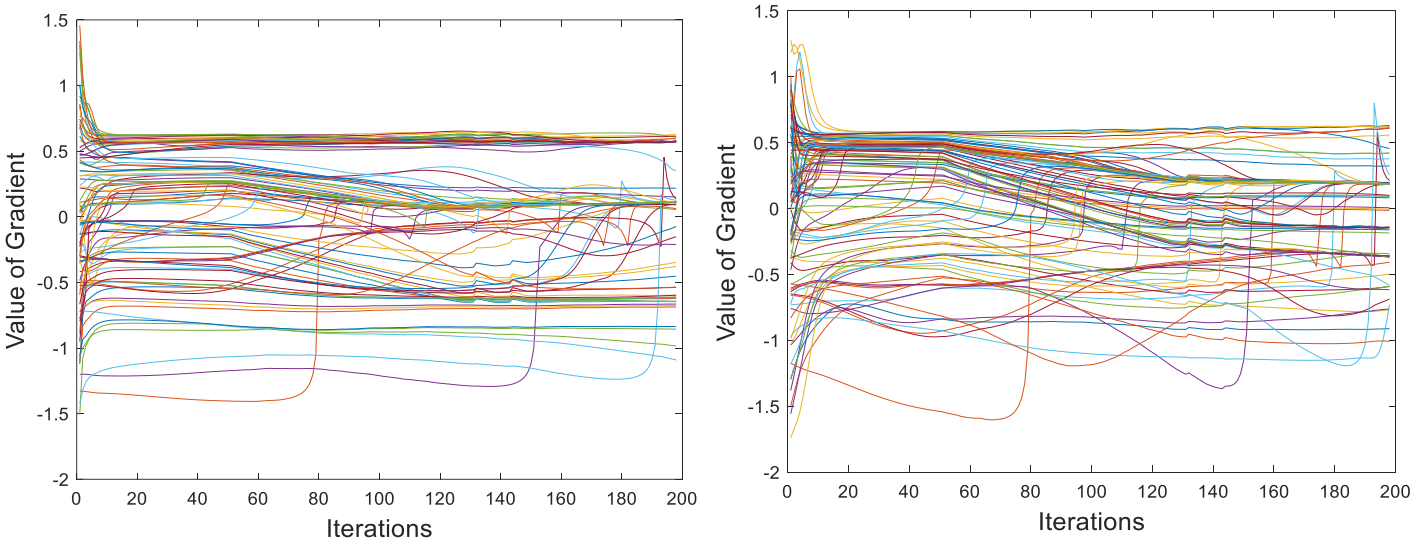
(a) (b)

Fig. 7. Optimal prior particle positions at different iterations: a) particle positions at θ_1; b) particle positions at θ_2.



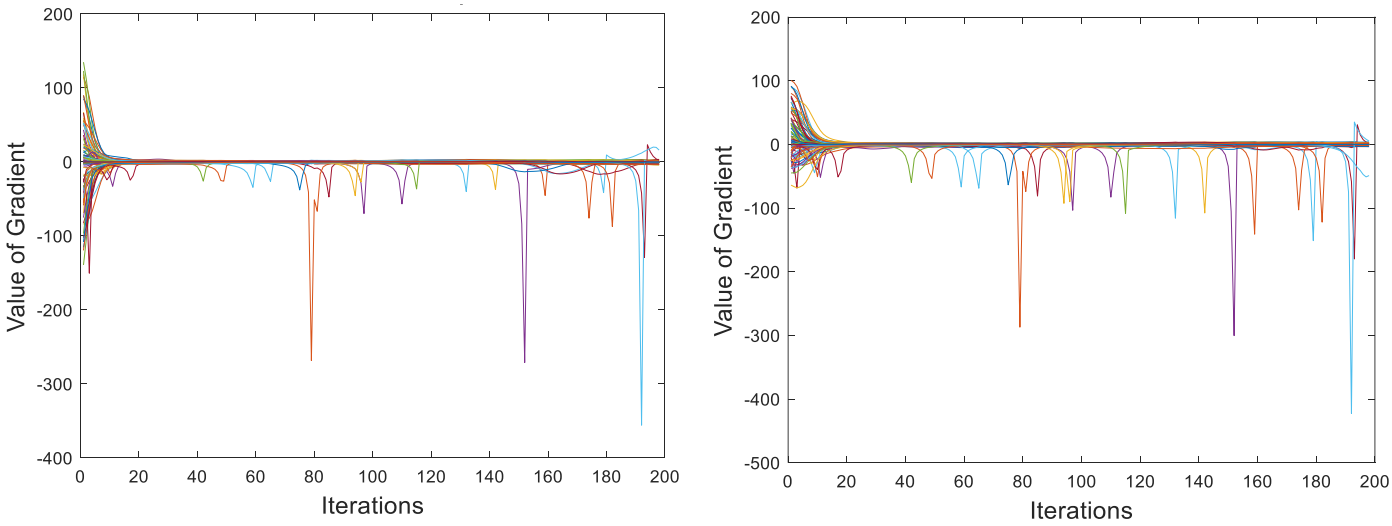
(a) (b)

Fig. 8. Approximation to posterior particle positions at different iterations: a) particle positions at θ_1; b) particle positions at θ_2.



(a) (b)

Fig. 9. Gradient of log prior at different iterations and at particle positions Θ_i^N w.r.t.: a) latent parameter θ_1; b) latent parameter θ_2 .



(a) (b)

Fig. 10. Gradient of log likelihood at several iterations and at particle positions Θ_i^N w.r.t.: a) latent parameter θ_1; b) latent parameter θ_2 .

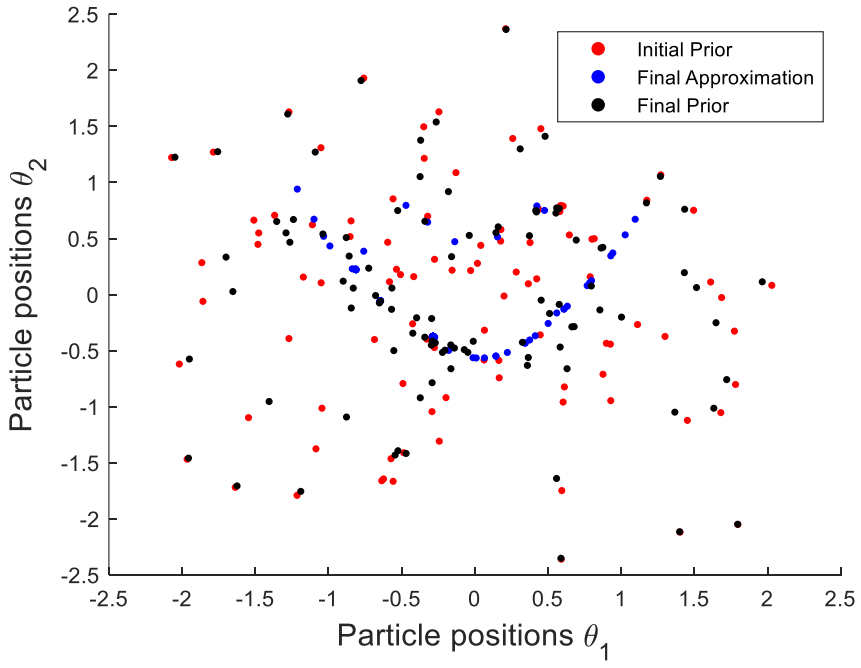


Fig. 11. Initial prior, final approximation to the posterior and final prior particle positions.

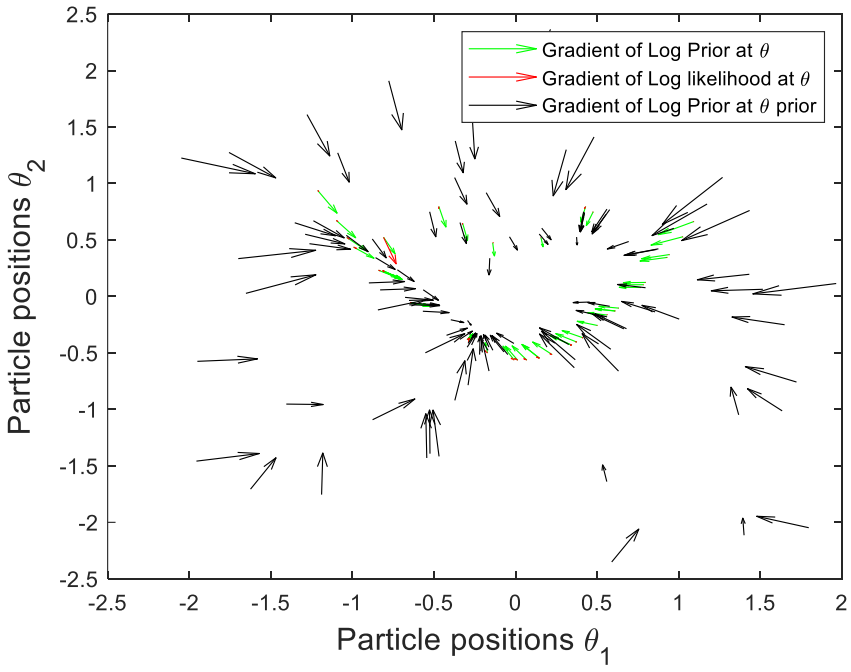


Fig. 12. Gradient/Quiver plot of log prior and log likelihood.

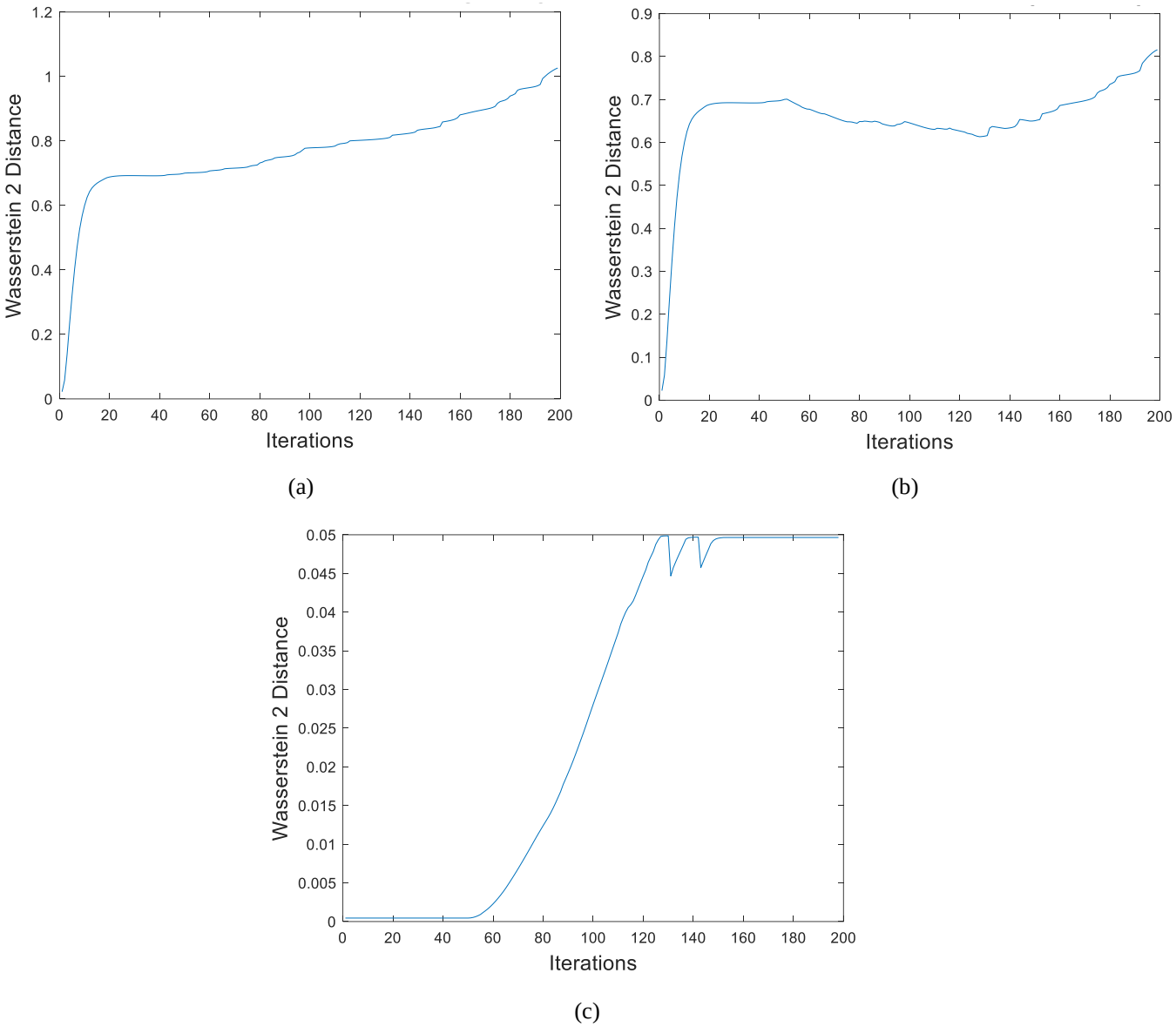
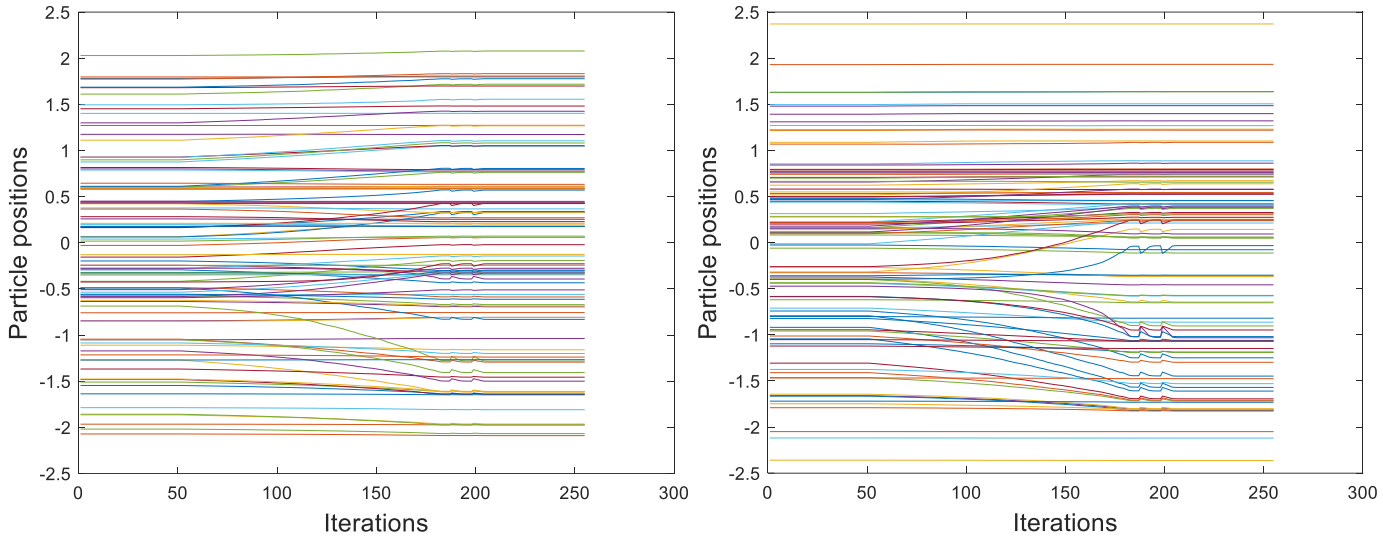
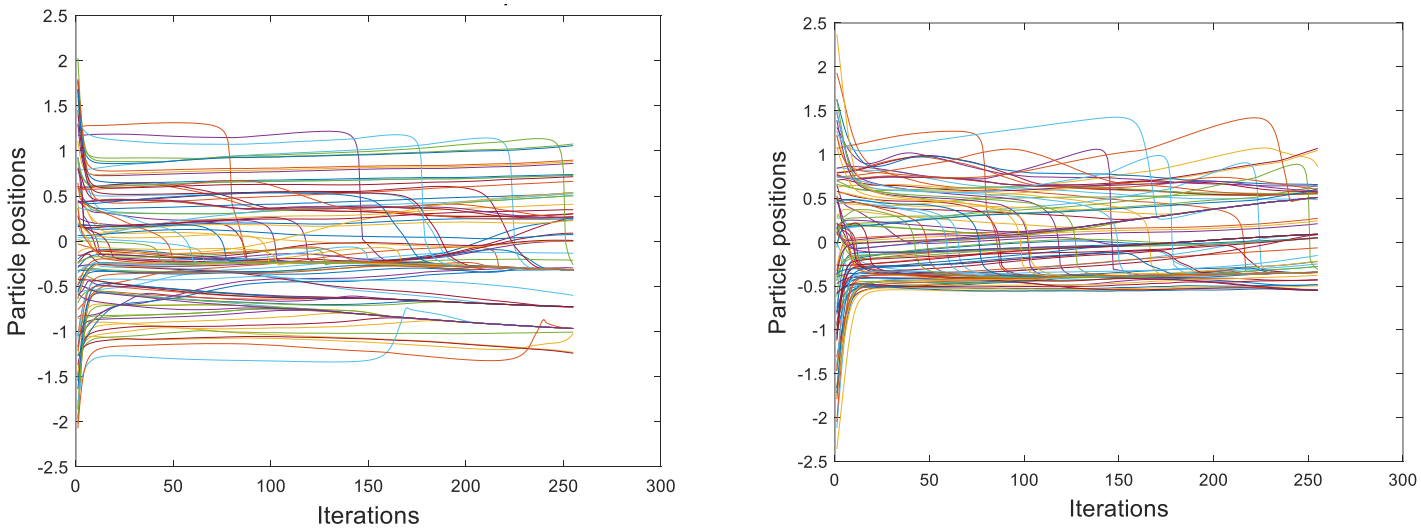


Fig. 13. 2-Wasserstein distance at different iterations i between: a) initial prior and approximation to posterior; b) approximation to posterior and prior; c) initial prior and prior.



(a) (b)

Fig. 14. Worst-case prior particle positions at different iterations: a) particle positions at θ_1; b) particle positions at θ_2.



(a) (b)

Fig. 15. Approximation to posterior particle positions at different iterations: a) particle positions at θ_1; b) particle positions at θ_2.

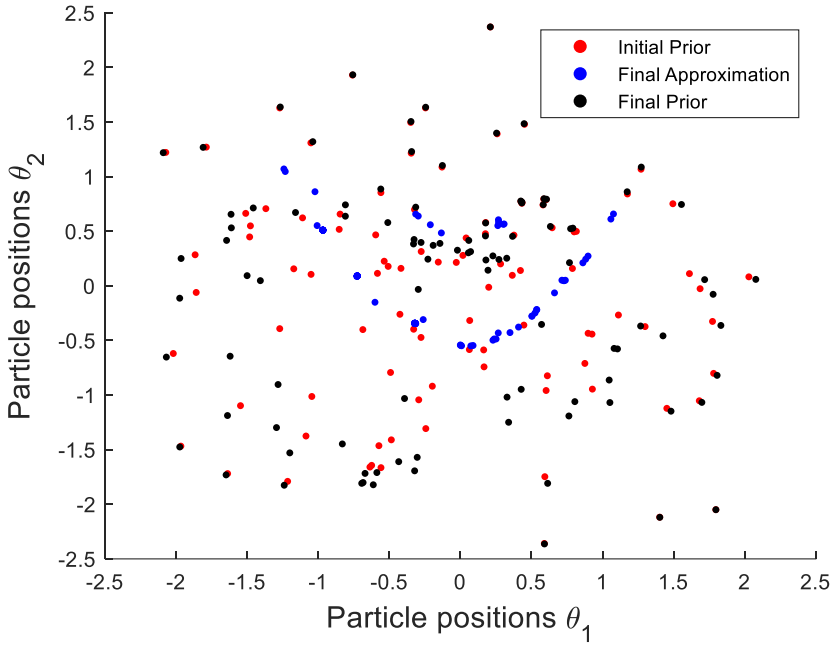


Fig. 16. Initial prior, final approximation to the posterior and final worst-case prior particle positions.

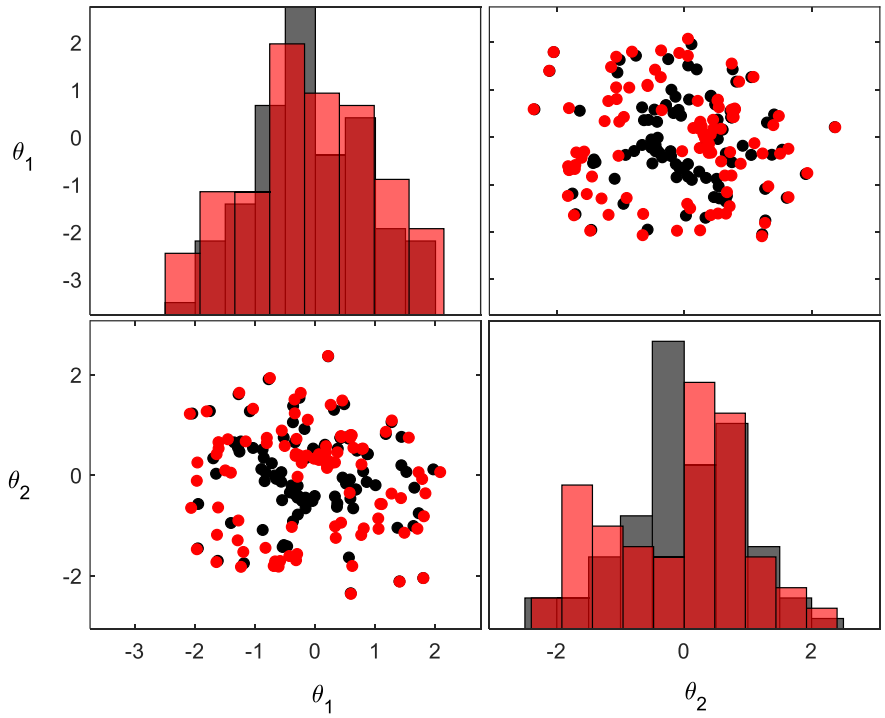


Fig. 17. Scatterplots and histograms show the prior, black – optimal prior case; red – worst-case prior.

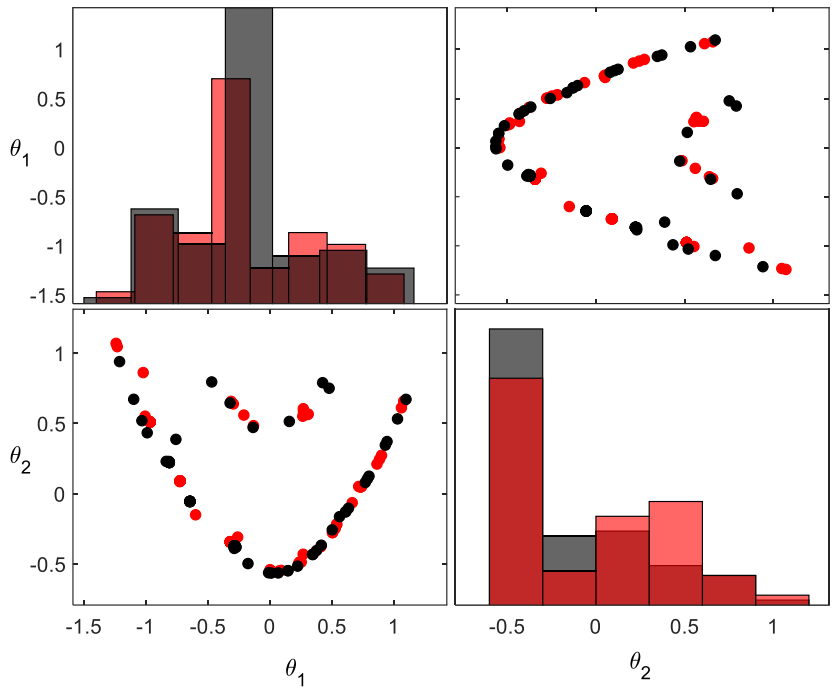


Fig. 18. Scatterplots and histograms show the approximation to the posterior, black – optimal prior case; red – worst-case prior.

6.2. Double beam structure example

The structure is shown on fig. 19, two connecting fixtures composed of three springs each: one translational, one shear, and one rotational, link two beams. This example shows practical interest, as it can be used to depict structural conditions where the attaching ensembles between elements show uncertainty. The causes of such uncertainty can be derived from boundary conditions and manufacturing variability. More specifically, the four uncertain parameters chosen are the spring stiffnesses and the young's modulus of both beams: the rotational springs $k_2 = 500\theta_1$ [Nm/rad], the shear springs $k_3 = 10^7 \theta_2$ [N/m], the translational springs $k_1 = 10^{10}\theta_3$ [N/m] and the Young's modulus of both beams $E_1 = E_2 = 210*10^9\theta_3$ [Pa]. For those four uncertain parameters, the initial prior

distribution is a multivariate gaussian prior chosen as $\mathcal{N}(I, 0.03I)$.

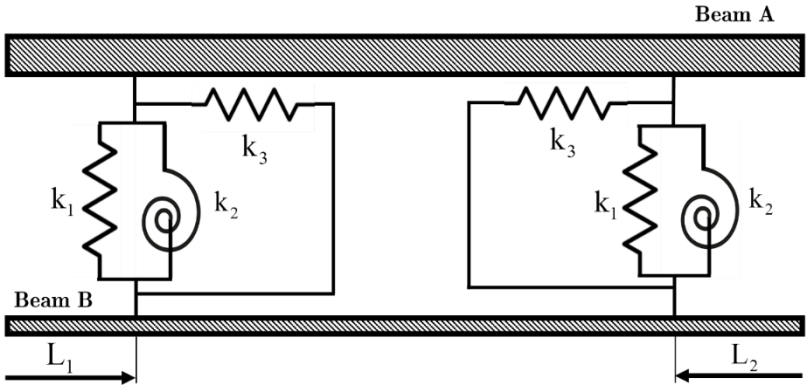


Fig. 19. Theoretical model of a coupled beam structure.

Dimensions and mechanical characteristics of the double beam model may be found on Table 1.

Table 1: Coupled beam dimensions, distances from edges to connections, and mechanical characteristics.

	Thickness [mm]	Width [mm]	Length [mm]	L_1	L_2	Young's modulus [GPa]	Density [Kg/m3]
Beam A	6	25	600	20	20	210	7800
Beam B	3						

	k_1	k_3	k_2
Springs	[MN/m]		[Nm/rad]
	100	10	500

Using the data on Table 1, the first eight natural frequencies of the model were assessed and introduced on Table 2.

Table 2: Coupled beam structure natural frequencies [Hz].

f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
16.0	50.2	92.8	134.6	245.3	260.7	428.0	478.6

The numerical frequencies obtained on Table 2 were produced using a Finite Element (FE) code. The code assumes a 2-dimensional Euler-Bernoulli beam model. Uniform discretization with two hundred FEs for each beam was used. Each FE has two nodes and each node has two degrees of freedom.

The likelihood function is assumed to be a multivariate gaussian distribution, the mean is given by the deterministic value of the eight natural frequencies in Table 2, and the covariance is assumed to be diagonal covariance matrix that has standard deviations of 2% of their deterministic values $\left(\sigma_i = 0.02\, f_i\right)$.

In this example, for the definition of the ambiguity set, the statistical distance used is also the 2-Wassertein distance, where the radius is $\varepsilon = 0.04$, and a nominal prior equal to the initial prior is selected.

The values described above are used as inputs of the algorithm, and the interacting Wasserstein Gradient Flows are used to find the resulting distributions for two different cases: a) the optimal prior and its resulting approximation to the posterior, and b) the worst-case prior and its resulting approximation to the posterior. In this example, the gaussian process method described in section 5.2.2 is used to calculate an approximation to the gradient of the log likelihood at the particle positions evaluated.

The values of step size used in the interacting particle flow WGF algorithm for the optimal prior case are $\alpha = 5*10^{-5}$ and $\tau = 2.5*10^{-3}$. The values used for the worst-case prior case are $\alpha = 5*10^{-5}$ and $\tau = 5*10^{-5}$.

Figures 20 and 21 respectively illustrate the positions of the particles from the optimal prior and from

the approximation to the posterior for each iteration. In fig. 20, after iteration $i = N_a$, it can be seen that for θ_1, θ_2 and θ_4 the prior particle positions start concentrating at values close to one. It can also be seen that θ_4 has the most rapid change out of all the latent variables, this is probably due to being the latent variable which most affects the model output. However, the opposite effect can be observed for θ_3, this is most likely due to the low sensitivity of the model output to changes of the latent variable θ_3. Figure 21 shows how the particles of the approximation to the posterior also concentrate to values closer to one as the number of iterations progresses for all the latent variables except for θ_3.

Figure 22 show scatter plots and histogram produced by the `plotmatrix` function of (*MATLAB*, 2020), of the initial particles from the prior/approximation to the posterior (red), the final particles from the optimal prior (black) and the final particles from approximation to the posterior (blue). It can be seen that the particles positions from the optimal prior and the approximation to the posterior are quite similar for all latent variables except for θ_3. From the histogram, it can be also seen that for the latent variable θ_3, the optimal prior has a bigger support than the initial prior.

The positions of the particles from the worst-case prior, and from the approximation to the posterior for each iteration are shown on fig. 23 and fig. 24 respectively. Figure 23 shows that after iteration $i = N_a$ for θ_1, θ_2 and θ_4 the prior particle positions part from values close to one. This is the opposite of what occurs for the optimal prior case. In a manner similar to what happens for the optimal case, the latent variable θ_4 experiments the fastest change of all the uncertain parameters. This is most likely due to the higher sensitivity of the model output to the changes of this latent variable. Figure 24 illustrates how as the number of iterations progresses the particles of the approximation to the posterior depart from values close to one. However, in this case, the change in the positions of the particles of the approximation to the posterior is not as significant as in the case for the optimal prior.

Histograms and scatter plots produced by the `plotmatrix` function of (*MATLAB*, 2020), can be found on fig. 25. The graphs illustrate the positions of the particles. In blue, the final particles from approximation to the posterior. In black, the final particles from the optimal prior. In red, the initial particles from the prior/approximation to the posterior. From the histograms, it can be deduced that in the worst-case prior, the supports of the graphs of all latent variables are bigger compared to the ones of the initial prior. The scatterplots show that worst-case prior particles have moved in such a manner that most of its particles lie in regions of lower posterior density.

Figure 26 directly compares the optimal prior and worst-case prior using the `plotmatrix` function from (*MATLAB*, 2020) by plotting scatter plots and the histograms of the latent variables. In general, for most of the latent variables it can be seen that the support of the worst-case prior is bigger than the optimal prior.

Figure 27 also has scatter plots and the histograms of the latent variables of the resulting approximation to the posterior when the optimal prior and worst-case prior has been calculated. When comparing the resulting approximations to the posteriors, it can be seen that for the case with the optimal prior the resulting approximation to the posterior is more concentrated compared to the approximation to the posterior that results from the worst-case prior. In this example, it is seen that the posterior is slightly sensitive to the considered uncertain prior.

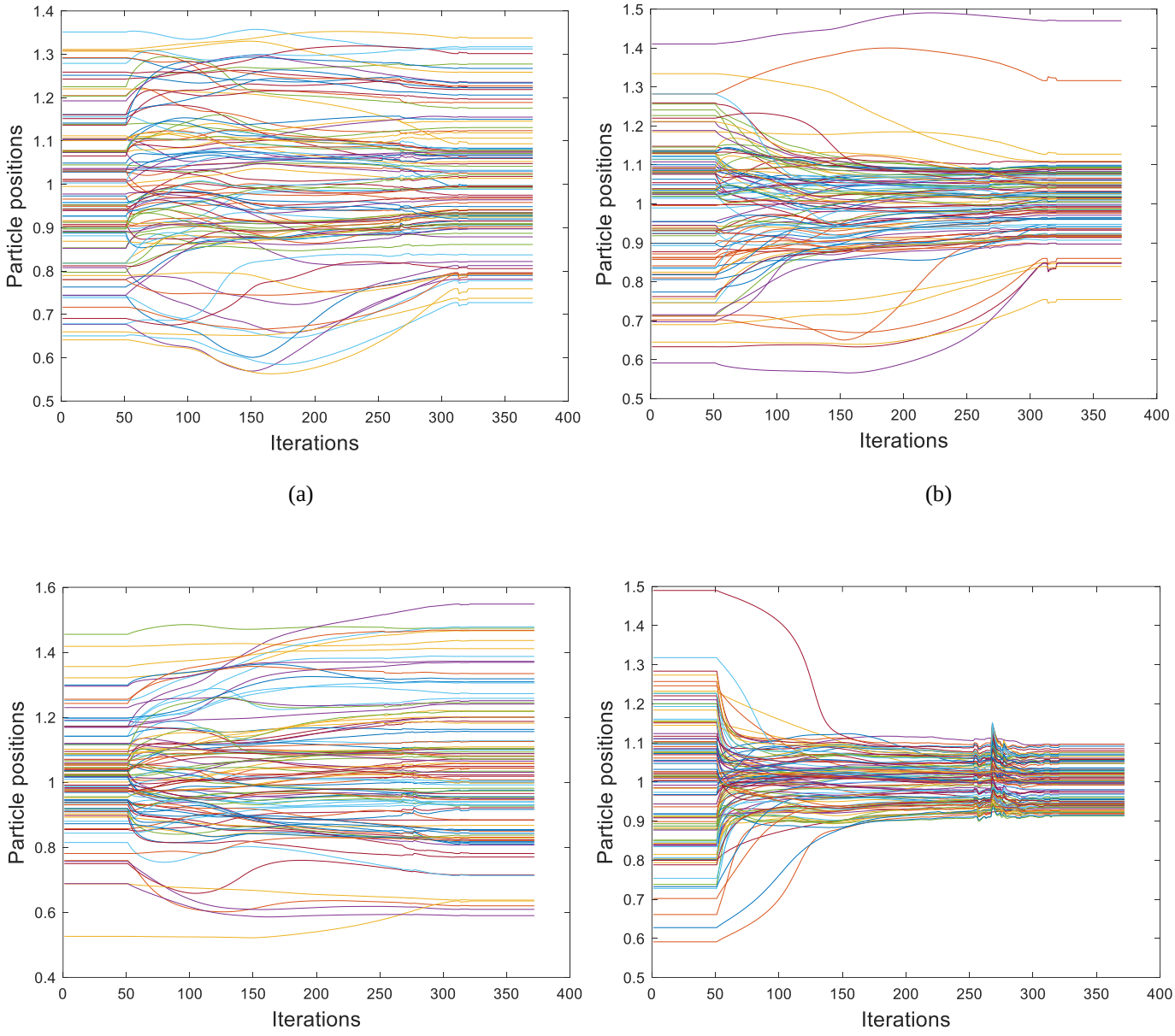
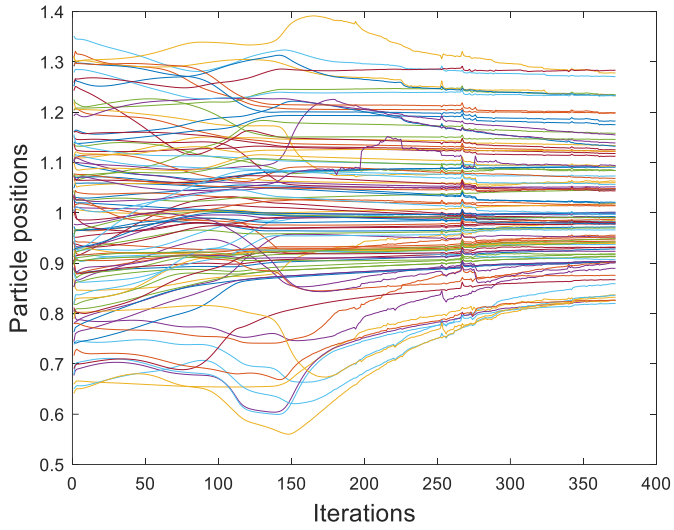
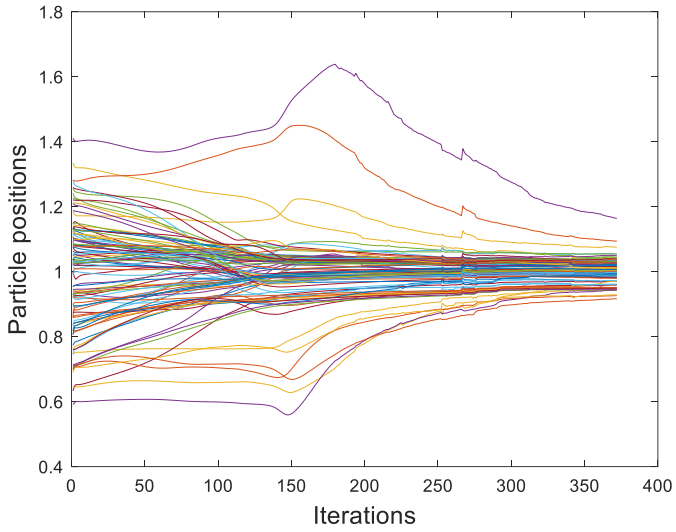


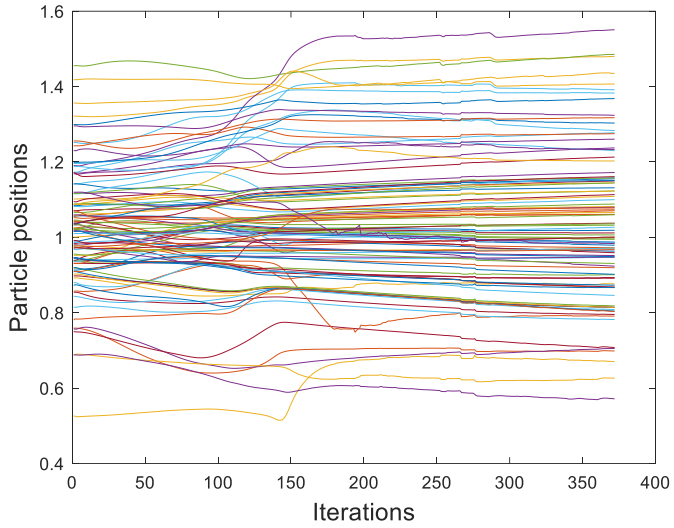
Fig. 20. Optimal prior particle positions at different iterations for different latent parameters: a) θ_1; b) θ_2; c) θ_3; d) θ_4.



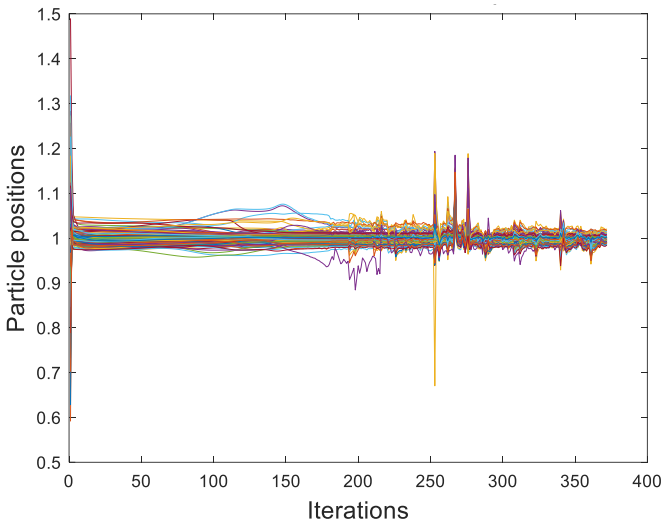
(a)



(b)



(c)



(d)

Fig. 21. Approximation to posterior particle positions at different iterations for different latent parameters: : a) θ_1; b) θ_2; c) θ_3; d) θ_4.

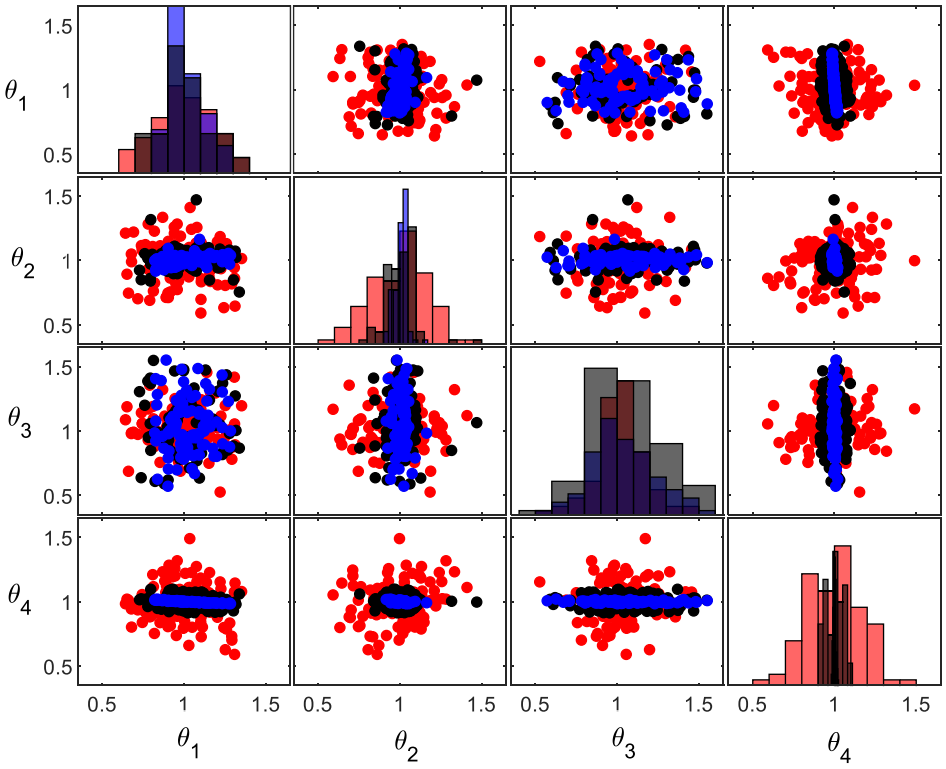


Fig. 22. Scatterplots and histograms show: red – initial prior; blue – final approximation to the posterior; black – optimal prior.

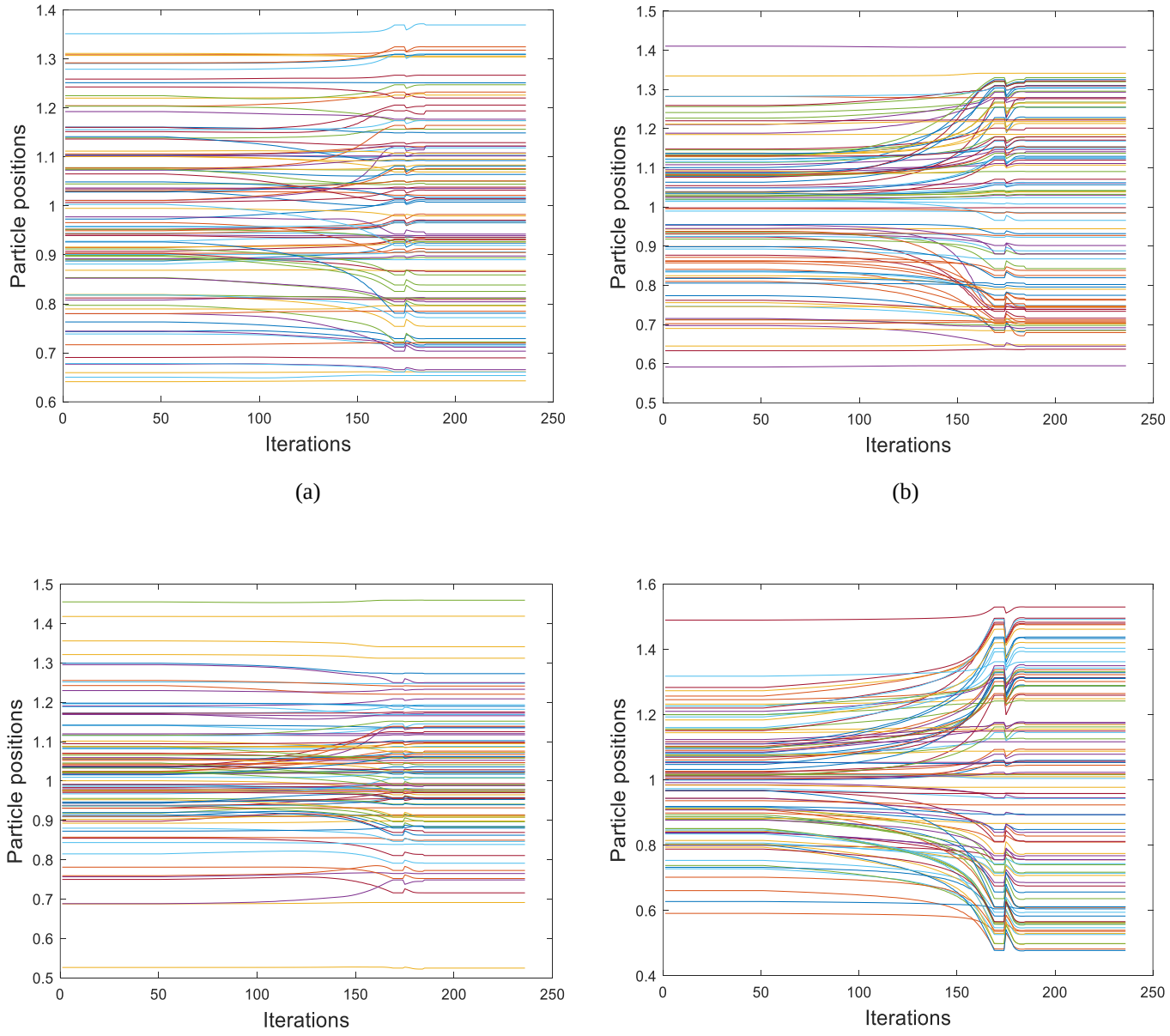


Fig. 23. Worst-case prior particle positions at different iterations for different latent parameters: a) θ_1; b) θ_2; c) θ_3; d) θ_4.

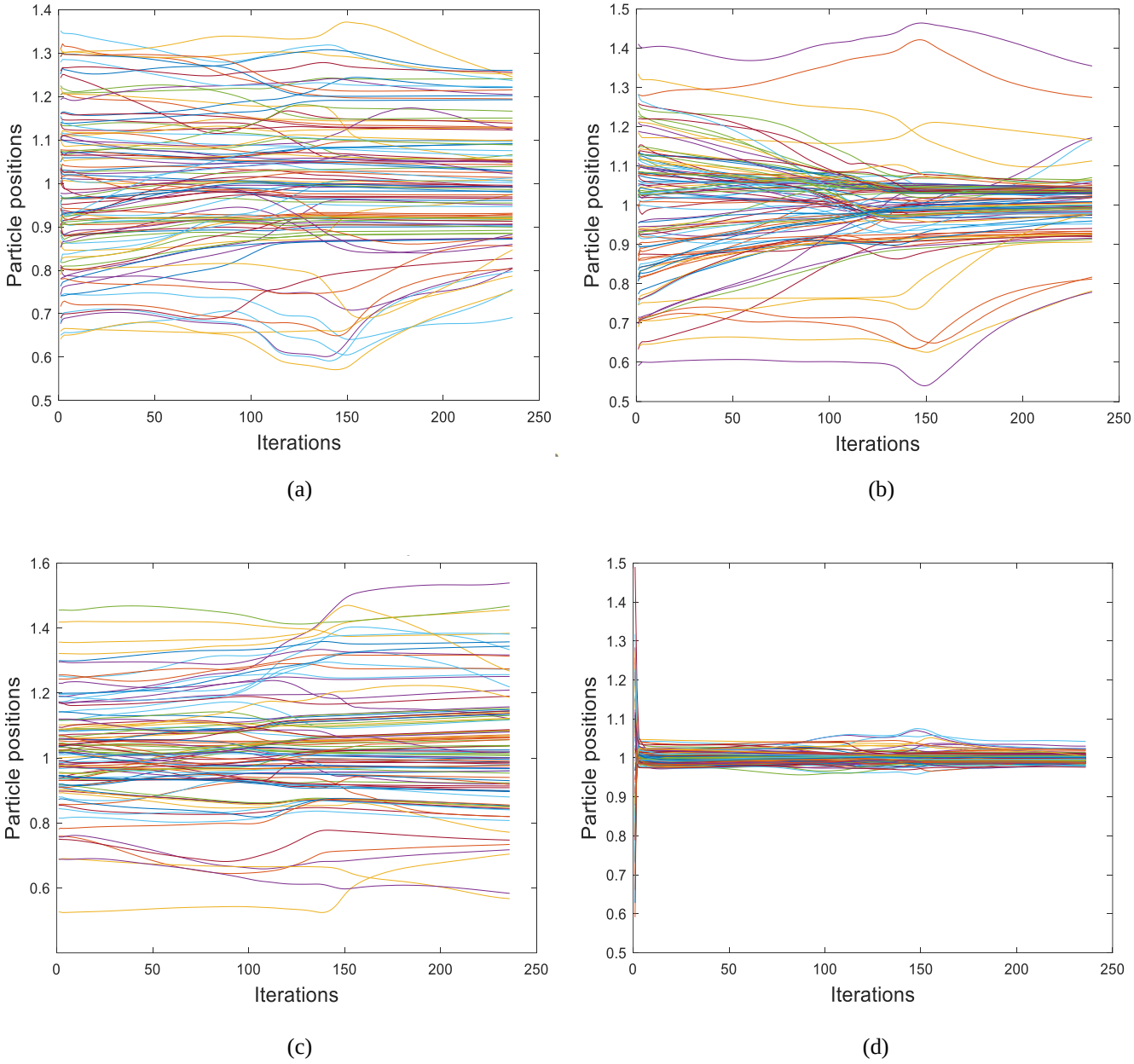


Fig. 24. Approximation to posterior particle positions at different iterations for different latent parameters: a) θ_1; b) θ_2; c) θ_3; d) θ_4.

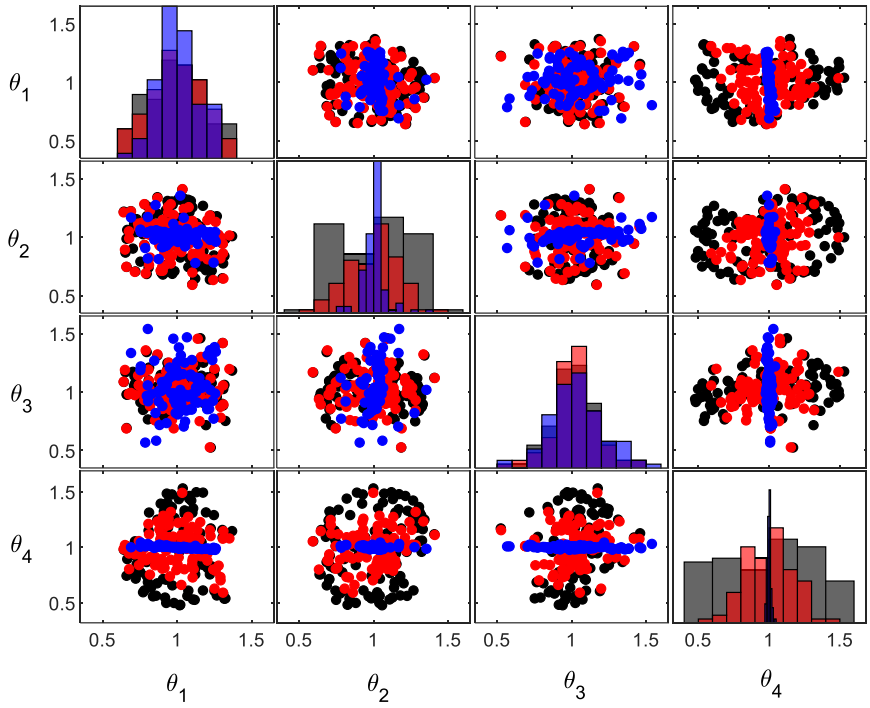


Fig. 25. Scatterplots and histograms show: red – initial prior; blue – final approximation to the posterior; black – worst-case prior.

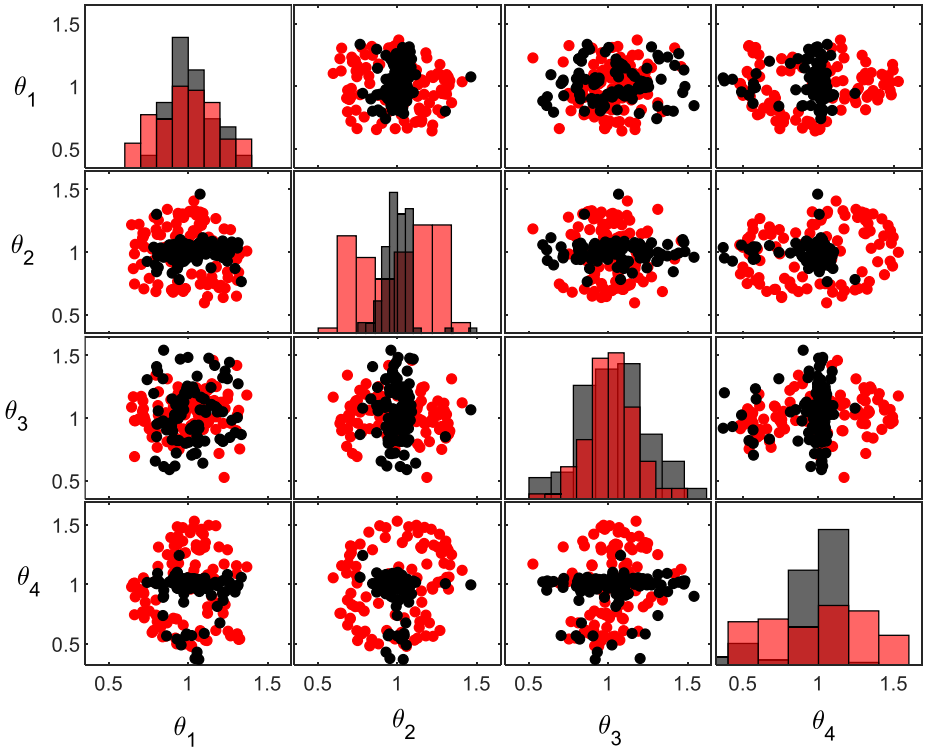


Fig. 26. Scatterplots and histograms show the prior, black – optimal prior case; red – worst-case prior.

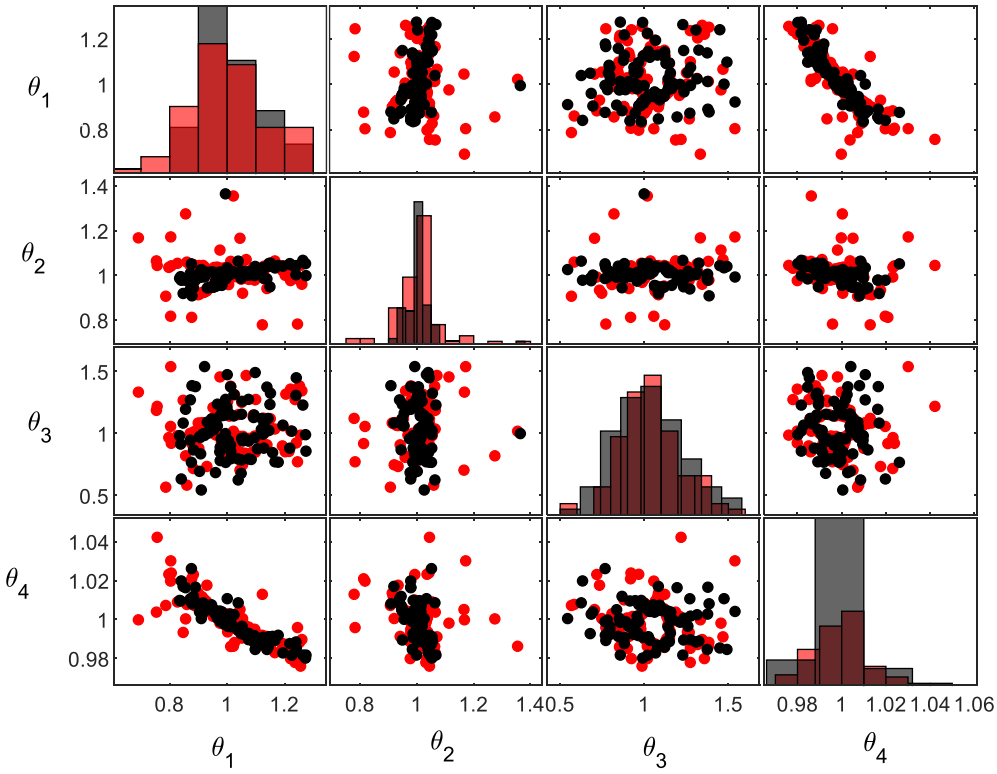


Fig. 27. Scatterplots and histograms show the approximation to the posterior, black – optimal prior case; red – worst-case prior.

7. Conclusion

In this paper, a Robust Bayesian Inference approach, based on Wasserstein Gradient Flows for the robust estimation of the latent parameters' posterior of a physics-based model, given observed data, has been proposed.

The proposed approach yields an estimation of the posterior distribution of the identified parameters by finding the optimal and worst-case prior distribution. This estimation is produced by an algorithm that combines an interacting Wasserstein Gradient Flow formulation with an ambiguity set. The ambiguity set is defined by a nominal distribution, a statistical distance and a radius. In this paper, the 2-Wasserstein distance is used as the statistical distance. Due to the properties of the 2-Wasserstein distance, this means that the distributions that lie inside the prescribed radius do not need to have the same support. The ambiguity set may be used to explore the sensitivity of the posterior prediction of the system to uncertainty in the prior. This investigation may be of particular interest for cases where the opinions of different experts are conflicting.

The interacting Wasserstein Gradient Flow formulation is derived from first principles, obtaining particle discretisation equations for the calculation of the optimal and worst-case prior. The derivation of the interacting WGFs, allows the development of the proposed method, which may reduce the computational cost incurred if all the possible prior distributions that lie inside the ambiguity set were to be tested directly. A kernel density estimator is used to obtain estimates of the gradient of the logarithm of the prior, and of the gradient of the logarithm of the approximations to the posterior with respect to the particle positions.

The paper illustrates how the gradient of the logarithm of the likelihood may be estimated either using an ensemble method, or a gaussian process regression method. Two numerical examples have been used to show both the optimal and worst-case prior and their resulting approximation to the posterior. In these examples, it is shown that for the optimal prior case the particles' positions tend to be near the particles of the approximation to the posterior, this means the optimal prior assigns a higher prior

density close to regions of high posterior density. For the worst-case prior, the opposite behaviour may be seen, the particles tend to move to positions far from the particles of the approximation to the posterior density. As a consequence, the worst-case prior has a bigger support than the initial prior.

The proposed approach may be of application to areas outside the scope of structural engineering, with particular interest to the commonly occurring engineering cases where some of the latent parameters may exhibit higher uncertainty than others.

Future work may focus on the convergence properties of the proposed approach and the dynamic selection of step sizes, as this would allow the proposed approach to become more computationally efficient. The method would also benefit from the development of a sample-efficient strategy, in which the reuse of samples from previous iterations may be integrated into the proposed methodology, reducing the number of simulations further. Another potential direction of interest is the development of a principled approach for the selection of the nominal prior and its radius, as at this stage is assumed to be known. These topics are currently under investigation.

References

Ambrosio, Luigi., Gigli, Nicola., & Savaré, Giuseppe. (2005). *Gradient flows : in metric spaces and in the space of probability measures*. Birkhäuser.

Bayraksan, G., & Love, D. K. (2015). Data-Driven Stochastic Programming Using Phi-Divergences. *INFORMS Tutorials in Operations Research*, 1–19. https://doi.org/10.1287/EDUC.2015.0134

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational Inference: A Review for Statisticians. In *Journal of the American Statistical Association* (Vol. 112, Issue 518, pp. 859–877). American Statistical Association. https://doi.org/10.1080/01621459.2017.1285773

Chen, C., Zhang, R., Wang, W., Li, B., & Chen, L. (2018). A Unified Particle-Optimization Framework for Scalable Bayesian Sampling. *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018*, *2*, 746–755. https://arxiv.org/abs/1805.11659v2

Chen, Y., Huang, D. Z., Huang, J., Reich, S., & Stuart, A. M. (2023). *Gradient Flows for Sampling: Mean-Field Models, Gaussian Approximations and Affine Invariance*. https://arxiv.org/abs/2302.11024v3

Chérief-Abdellatif, B.-E., & Alquier, P. (2019). *MMD-Bayes: Robust Bayesian Estimation via Maximum Mean Discrepancy*. https://arxiv.org/abs/1909.13339v2

Chizat, L., & Bach, F. (2018). On the Global Convergence of Gradient Descent for Over-parameterized Models using Optimal Transport. *Advances in Neural Information Processing Systems*, *2018-December*, 3036–3046. https://arxiv.org/abs/1805.09545v2

Detommaso, G., Cui, T., Spantini, A., Marzouk, Y., & Scheichl, R. (2018). A Stein variational Newton method. *Advances in Neural Information Processing Systems*, *2018-December*, 9169–9179. https://arxiv.org/abs/1806.03085v2

Dewaskar, M., Tosh, C., Knoblauch, J., & Dunson, D. B. (2023). *Robustifying likelihoods by optimistically re-weighting data*. https://arxiv.org/abs/2303.10525v1

Diakonikolas, I., Hopkins, S. B., Kane, D., & Karmalkar, S. (2020). *Robustly Learning any Clusterable Mixture of Gaussians*. https://arxiv.org/abs/2005.06417v1

Ding, S., Dong, H., Fang, C., Lin, Z., & Zhang, T. (2023). *Provable Particle-based Primal-Dual Algorithm for Mixed Nash Equilibrium*. https://arxiv.org/abs/2303.00970v1

Dunbar, O. R. A., Duncan, A. B., Stuart, A. M., & Wolfram, M. T. (2022). Ensemble Inference Methods for Models With Noisy and Expensive Likelihoods. *Https://Doi.Org/10.1137/21M1410853*, *21*(2), 1539–1572. https://doi.org/10.1137/21M1410853

Frazier, D. T. (2020). *Robust and Efficient Approximate Bayesian Computation: A Minimum Distance Approach*. https://arxiv.org/abs/2006.14126v1

Gao, Y., & Liu, J.-G. (2020). A note on parametric Bayesian inference via gradient flows. *Annals of Mathematical Sciences and Applications*, *5*(2), 261–282.

Ghosh, A., & Basu, A. (2016). Robust Bayes estimation using the density power divergence. *Annals of the Institute of Statistical Mathematics*, *68*(2), 413–437. https://doi.org/10.1007/S10463-014-0499-0/FIGURES/10

Go, J., & Isaac, T. (2022). Robust Expected Information Gain for Optimal Bayesian Experimental Design Using Ambiguity Sets. *Proceedings of the 38th Conference on Uncertainty in Artificial Intelligence, UAI 2022*, 728–737. https://arxiv.org/abs/2205.09914v1

Gonçalves, F. B., Prates, M. O., & Lachos, V. H. (2015). Robust Bayesian model selection for heavy-tailed linear regression using finite mixtures. *Brazilian Journal of Probability and Statistics*, *34*(1), 51–70. https://doi.org/10.1214/18-BJPS417

Hermans, J., Delaunoy, A., Rozet, F., Wehenkel, A., Begy, V., & Louppe, G. (2021). *A Trust Crisis In Simulation-Based Inference? Your Posterior Approximations Can Be Unfaithful*. https://arxiv.org/abs/2110.06581v3

Hooker, G., & Vidyashankar, A. N. (2011). Bayesian Model Robustness via Disparities. *Test*, *23*(3),

556–584. https://doi.org/10.1007/s11749-014-0360-z

Igea, F., & Cicirello, A. (2022). Cyclical Variational Bayes Monte Carlo for Efficient Multi-Modal Posterior Distributions Evaluation. In *Mechanical Systems and Signal Processing* (Vol. 186). Academic Press. http://arxiv.org/abs/2202.11645v1

Kuhn, D., Esfahani, P. M., Nguyen, V. A., & Shafieezadeh-Abadeh, S. (2019). Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning. *Operations Research & Management Science in the Age of Analytics*, 130–166. https://doi.org/10.1287/educ.2019.0198

Lin, T., Jin, C., & Jordan, M. I. (2019). On Gradient Descent Ascent for Nonconvex-Concave Minimax Problems. *37th International Conference on Machine Learning, ICML 2020, PartF168147-8*, 6039–6049. https://arxiv.org/abs/1906.00331v8

Liu, C., Cheng, P., Zhang, R., Zhuo, J., Zhu, J., & Carin, L. (2018). *Accelerated First-order Methods on the Wasserstein Space for Bayesian Inference*. https://www.researchgate.net/publication/326222852

Liu, Q., & Wang, D. (2016). Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm. *Advances in Neural Information Processing Systems*, 2378–2386. https://arxiv.org/abs/1608.04471v3

Lu, J., Lu, Y., & Nolen, J. (2019). Scaling Limit of the Stein Variational Gradient Descent: The Mean Field Regime. *Https://Doi.Org/10.1137/18M1187611*, *51*(2), 648–671. https://doi.org/10.1137/18M1187611

Lu, Y. (2022). *Two-Scale Gradient Descent Ascent Dynamics Finds Mixed Nash Equilibria of Continuous Games: A Mean-Field Perspective*. https://arxiv.org/abs/2212.08791v2

Lyddon, S., Walker, S., & Holmes, C. (2018). Nonparametric learning from Bayesian models with randomized objective functions. *Advances in Neural Information Processing Systems, 2018-*

December, 2071–2081. https://arxiv.org/abs/1806.11544v2

MATLAB (a). (2020).

Matsubara, T., Knoblauch, J., Briol, F. X., & Oates, C. J. (2021). Robust Generalised Bayesian Inference for Intractable Likelihoods. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *84*(3), 997–1022. https://doi.org/10.1111/rssb.12500

Mei, S., Montanari, A., & Nguyen, P. M. (2018). A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(33), E7665–E7671. https://doi.org/10.1073/PNAS.1806579115/SUPPL_FILE/PNAS.1806579115.SAPP.PDF

Ramgraber, M., Weatherl, R., Blumensaat, F., & Schirmer, M. (2021). Non-Gaussian Parameter Inference for Hydrogeological Models Using Stein Variational Gradient Descent. *Water Resources Research*, *57*(4), e2020WR029339. https://doi.org/10.1029/2020WR029339

Rasmussen, C. E. (2003). Gaussian Processes to Speed up Hybrid Monte Carlo for Expensive Bayesian Integrals. *BAYESIAN STATISTICS*, *7*, 651–659.

Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians* (Vol. 87). Springer International Publishing. https://doi.org/10.1007/978-3-319-20828-2

Santambrogio, F. (2016). { Euclidean, Metric, and Wasserstein } Gradient Flows: an overview. *Bulletin of Mathematical Sciences*, *7*(1), 87–154. https://doi.org/10.1007/s13373-017-0101-1

van Parys, B. P. G., Esfahani, P. M., & Kuhn, D. (2017). From Data to Decisions: Distributionally Robust Optimization is Optimal. *Management Science*, *67*(6), 3387–3402. https://doi.org/10.1287/mnsc.2020.3678

Wang, Y., Chen, P., & Li, W. (2022). Projected Wasserstein Gradient Descent for High-Dimensional Bayesian Inference. *Https://Doi.Org/10.1137/21M1454018*, *10*(4), 1513–1532. https://doi.org/10.1137/21M1454018

Yamada, M., Suzuki, T., Kanamori, T., Hachiya, H., & Sugiyama, M. (2011). Relative Density-Ratio Estimation for Robust Distribution Comparison. *Neural Computation*, *25*(5), 1324–1370. https://doi.org/10.1162/NECO_a_00442

6.2 Conclusions

The paper V attached develops an approach based on the combination of an ambiguity set defined on a nominal prior, and interacting Wasserstein Gradient Flows to evaluate the worst-case and optimal priors in the vicinity of a chosen nominal prior and their associated posteriors. The worst-case and optimal priors are obtained w.r.t. a metric illustrated in the paper. The worst-case and optimal prior are obtained by exploring a reduced probability space limited by an ambiguity set, defined by a statistical distance and a nominal distribution. This ambiguity set contains all the distributions in the vicinity of the nominal distribution at a lower distance than a specified value. The 2-Wasserstein distance is used as the statistical distance in the ambiguity set, therefore, the distributions that lie inside the prescribed radius do not need to have the same support [77].

Two numerical examples demonstrate how, in the case of the optimal prior, the particle positions tend to be close to those of the approximation to the posterior. This indicates that the optimal prior places a greater prior density close to areas with a high posterior density. The opposite behaviour for the worst-case prior can also be observed as the particles prefer to shift to positions far from the particles of the approximation to the posterior density. As a result, the worst-case prior has a bigger support than the initial prior.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	An interacting Wasserstein Gradient Flow strategy to Robust Bayesian Inference
Publication Status	□Published □ Accepted for Publication ☒Submitted for Publication □Unpublished and unsubmitted work written in a manuscript style
Publication Details	N/A

Student Confirmation

Student Name:	Felipe Igea		
Contribution to the Paper	Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing.		
Signature	Felipe Igea	Date	16/09/23

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title:	Dr Alice Cicirello		
Supervisor comments The student made a substantial contribution to this draft manuscript, and the description above is accurate			
Signature		Date	18/09/23

Chapter 7 - Conclusions and future work

7.1 Conclusions

At the design stage, before the structure is built, and no access to a prototype is available, a physics-based model of the structure can be used to infer predictions of the expected observations. Based on the data simulated using the physics-based model, optimal sensor placement studies may be performed. This is important as the collection of data, through sensors, is used to infer information about the condition of the structure, which can be subsequently used to support maintenance decision-making. After measurements are obtained from the sensors, and the condition of the structure is inferred given this new information (using for example Bayesian Inference), an optimal action given the inferred condition can be taken. Some of the possible actions may be repairment, further observation, etc.

In this thesis, in the context of limited data availability, two grand challenges in structural dynamics are tackled, the design of complex engineering structures, and the updating of information on already existing critical structures by integrating physics-based models. The motivation to complete these challenges is derived from the need to prevent unforeseen failures that could result in significant losses, such as lives or high repair expenses. To this end, in this thesis, these two grand challenges were split into four subchallenges: a) quantifying uncertainties at the design stage; b) where should measurements be taken?; c) updating the model using measurements; d) handling prior uncertainty.

Having established these four subchallenges, the main outcomes from this thesis can be summarised as follows:

In chapter 3, a non-parametric approach is developed for the evaluation of the pdfs of the natural frequencies, damping ratios, and modal shapes of an ensemble of nominally identical linear time-invariant systems. These distributions are important for the performance assessment of the designed structures, as this would enable the selection of designs that are robust to these uncertainties, avoiding extensive modifications of the manufactured product. The developed approach employs a non-

parametric model of uncertainty by exploiting Random Matrix Theory (RMT) results [6,69]. This is useful, as nonparametric methods [69] avoid the need to specify the uncertainties' sources, and the description of the model's parameters uncertainty, which are often hard to determine. The main advantage of this approach is that it does not require to specify the different sources of uncertainties in advance. The uncertainty of the parameters does not have to be propagated through the equations of motion, becoming of great advantage when the model of the system is highly complex or unknown, a situation that may introduce modelling errors that affect the overall results. In these circumstances, it may be more convenient to build the ensemble from the measurements of a prototype structure.

The single structure, represented either as a mathematical model, or a prototype, is used to obtain simulated data or measurements, respectively, which in turn are employed to build a discrete time state-space model description. This model is then used to efficiently assess the effects on the modal properties of different levels of uncertainties, represented through suitably chosen dispersion parameters. Numerical studies to evaluate the pdfs of the natural frequencies and the damping ratios on a two mass-spring-damper (discrete system), and a cantilever system (continuous system) for two levels of dispersion parameter values were conducted. For high values of dispersion parameter value, it is observed on the pdf of the modal parameters a non-physical behaviour. The presence of negative values in the support of the pdf of the first damping ratio in the mass-spring-damper system is an example of the non-physical behavior brought about by the approach.

During the investigation of the Bayesian framework for optimal sensor placement in chapter 4, to obtain the best location where to take measurements, it was found that most of the computational cost is due to the high number of runs required for the inference of the parameters of the physics-based model. As a high requirement of numerical resources became apparent, an alternative to the traditional sampling-based approaches that are typically used for Bayesian Inference was required. Three different utility functions were used to investigate the optimal sensor location for the identification of two physical parameters in the case of a beam connected to the ground by two supports, each made of one translational and one rotational spring. In this particular study, it is shown

that for the three different utility functions, the same location is chosen as the optimal. It was also confirmed that the values of the utility function at locations close to the nodal points are low. This occurs because the modal properties estimated at those locations are less accurate due to the lower information content of the observations taken near nodal points compared to readings performed in other points of the beam.

In chapter 5, VI is found to be an excellent alternative to the traditional sampling-based approaches for updating the model using measurements under a Bayesian framework, as reliable and efficient approximations of the posterior distributions of the latent parameters of interest were obtained. In the first paper attached in this chapter (paper III), an application of the VI methodology is investigated. A multivariate Gaussian variational is chosen as a posterior distribution of the parameters of interest, and its hyperparameters are optimised by maximising the evidence lower bound (ELBO). The second paper attached to this chapter (paper IV) deals with the VI of multimodal uncertain parameters in physical models characterised by numerical approximations that require significant computational resources when evaluated. There are several reasons that may create multimodality of the uncertain parameters in structural dynamics. One frequent reason for multimodality to occur may be environmental changes (e.g., temperature changes may imply stiffness changes [21]). Multimodal and unimodal posteriors of physics-based models parameters were chosen to compare the results of the proposed method (Cyclical VBMC), with the advanced Transitional Ensemble Markov Chain Monte Carlo (TEMCMC) [78], the monotonic VBMC, and the standard VBMC [53,54]. It was found both that the Cyclical VBMC reduces the number of evaluations by at least a factor of twenty compared to TEMCMC, and that it was better equipped to deal with multimodality compared to the monotonic VBMC and the standard VBMC.

Lastly, in chapter 6, the robustness of the posterior prediction to changes in the prior distribution in Bayesian Inference is addressed. An approach that uses a combination of the concepts of ambiguity set and interacting Wasserstein Gradient Flows is used to obtain the worst-case and optimal prior. A metric defined in the paper is used to calculate the optimal and worst-case priors. The priors are inside

an ambiguity set, this reduces the probability space needed to be explored. The ambiguity set is the set that contains a nominal distribution, and all the distributions at a lower statistical distance to the nominal distribution. As illustrated by the two numerical examples, the optimal prior case assigns a higher prior density in regions of high posterior density, and the placements of the particles of the prior tend to be close to the approximation to the posterior. In the worst-case prior, the behavior is the opposite, the prior's particles have a tendency to go to locations that are far from the approximation's particles to the posterior density. Therefore, compared to the initial prior, the worst-case prior exhibits a larger support.

The physics-based models used throughout the papers, have been used to showcase the applicability of the methods derived and investigated in this thesis. The examples were fairly simple and with a low number of uncertain parameters. However, the methods can be extended to more complex physics-based models, and applied for a higher number of uncertain parameters. To deal with a higher number of uncertain parameters, dimensionality reduction, or sensitivity analyses methods may be used with the purpose of reducing the overall number of uncertain parameters.

It should be noted that although the main goal of this thesis was to tackle the main pressing challenges in structural dynamics, most of the methods developed can be used in other fields of engineering where uncertainty is present.

7.2 Suggestions for further work

The following challenges can be classified as possible future work, and natural progression of the work described in this thesis:

1) In the paper attached in chapter 3, the Eigensystem Realization Algorithm limits the use of this method to free vibration. However, this assumption may be relaxed by using other system identification techniques, for example the Natural Excitation Technique - Eigensystem Realization Algorithm. The extension of the combination of Random Matrix Theory with other system identification methods would allow it to be used in different operational conditions.

2) In chapter 4, the main challenge faced was that to produce a Bayesian Optimal Design technique for sensor placement, methods able to obtain accurate predictions of the posterior pdfs of the latent parameters under conditions of limited computational budget are required. Therefore, variational approximations on the posterior would result in a significant reduction in the overall computational cost. Another potential direction of research would be the investigation and development of a robust optimal sensor placement framework. The robustness of the sensor configuration would be understood with respect to the choice of the prior distribution and/or likelihood, this could be implemented by extending the method developed in chapter 6 by obtaining interacting Wasserstein Gradient Flows with respect to a metric of interest, such as the expected information gain. This study would allow the understanding of how sensitive the optimal sensor configuration is with respect to our prior assumptions, under limited data availability.

3) In the first paper in chapter 5 (paper III), VI was first explored as a possible alternative to the common sampling-based Bayesian Inference methods, and it was confirmed to be a good option to reduce computational cost. Natural extensions to the approach used in paper III are: a) the use of Gaussian mixtures for the postulated posterior, as they possess the capability of modelling intricate smooth posteriors, as well as posterior distributions with multiple modes; b) the creation of an algorithm that efficiently selects samples from areas close to the higher values of unnormalized posterior ; c) the development of a technique for sample reusing, so that the numerical resources required would be further reduced. These three extensions were tackled in the development of the Cyclical VBMC approach (paper IV) that significantly reduces the number of model evaluations compared to state-of-the-art sampling-based methods such as TEMCMC.

4) As mentioned in point 2, the proposed Cyclical VBMC approach found in ‘Cyclical Variational Bayes Monte Carlo for Efficient Multi-Modal Posterior Distributions Evaluation’ (paper IV) may be used to significantly reduce the computational cost required for the statistical model updating involved in optimal sensor placement information-based methods. However, this approach is only suitable for applications with less than 20 uncertain parameters, as the use of Gaussian processes

implies a limitation to the number of uncertain parameters. Therefore, applicable methods in cases with a high number of uncertain parameters may be explored. For example, the Cyclical VBMC could be extended to cases with a high number of uncertain parameters by using a likelihood informed subspace method [79], thereby reducing the number of uncertain parameters to a lower amount that could be handled by the Gaussian process.

5) Subsequent research to enhance the computational efficiency of the method suggested in paper V ‘An Interacting Wasserstein Gradient Flow Strategy to Robust Bayesian Inference’ may centre on investigating the step size selection. For example, accounting for the magnitude of the gradients of the logarithm of the likelihood at the sample positions may be used to dynamically select the step size. Another option would be to restrict how much the distribution changes at each time step by allowing a maximum Wasserstein distance between two distributions obtained at consecutive time steps. A further improvement to the approach would be the introduction of a methodology in which samples of preceding iterations could be reused, increasing the numerical efficiency through the reduction of the number of simulations. For example, Gaussian Processes may be used to obtain a surrogate model of the gradient of the likelihood, however, one should ensure that the samples used to create this surrogate model are appropriate to estimate the gradients at samples where the physics-based model has not been evaluated.

References

[1] S. Adhikari, A. Sarkar, Uncertainty in structural dynamics: Experimental validation of a Wishart random matrix model, J Sound Vib 323 (2009) 802–825. https://doi.org/10.1016/j.jsv.2009.01.030.

[2] E. Simoen, G. De Roeck, G. Lombaert, Dealing with uncertainty in model updating for damage assessment: A review, Mech Syst Signal Process 56 (2015) 123–149. https://doi.org/10.1016/j.ymssp.2014.11.001.

[3] J. Zhang, G. Shen, Y. Du, P. Hu, Modal analysis of a lightweight engine hood design considering stamping effects, in: Applied Mechanics and Materials, Trans Tech Publications Ltd, 2013: pp. 364–369. https://doi.org/10.4028/www.scientific.net/AMM.281.364.

[4] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech Syst Signal Process 21 (2007) 2560–2574. https://doi.org/10.1016/j.ymssp.2006.12.007.

[5] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A.K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, Mech Syst Signal Process 138 (2020) 106587. https://doi.org/10.1016/j.ymssp.2019.106587.

[6] C. Soize, Uncertainty Quantification -An Accelerated Course with Advanced Applications in Computational Engineering, Springer International Publishing, Cham, 2017. https://doi.org/10.1007/978-3-319-54339-0.

[7] M. Wright, R. Weave, New directions in linear acoustics and vibration: Quantum chaos, random matrix theory, and complexity, Cambridge University Press, 2010. https://doi.org/10.1017/CBO9780511781520.

[8] M.S. Kompella, R.J. Bernhard, Measurement of the statistical variation of structural-acoustic characteristics of automotive vehicles, in: SAE Technical Papers, SAE International, 1993. https://doi.org/10.4271/931272.

[9] J.-F. Durand, C. Soize, L. Gagliardini, Structural-acoustic modeling of automotive vehicles in presence of uncertainties and experimental identification and validation, J Acoust Soc Am 124 (2008) 1513–1525. https://doi.org/10.1121/1.2953316.

[10] R.S. Langley, Unified Approach to Probabilistic and Possibilistic Analysis of Uncertain Systems, J Eng Mech 126 (2000) 1163–1172. https://doi.org/10.1061/(asce)0733-9399(2000)126:11(1163).

[11] C. Soize, Stochastic modeling of uncertainties in computational structural dynamics - Recent theoretical advances, J Sound Vib 332 (2013) 2379–2395. https://doi.org/10.1016/j.jsv.2011.10.010.

[12] M. Faes, D. Moens, Recent Trends in the Modeling and Quantification of Non-probabilistic Uncertainty, Archives of Computational Methods in Engineering 27 (2020) 633–671. https://doi.org/10.1007/s11831-019-09327-x.

[13] D. Li, Herausgeber: Claus-Peter Fritzen Sensor Placement Methods and Evaluation Criteria in Structural Health Monitoring, 2011.

[14] J.L. Beck, L.S. Katafygiotis, Updating Models and Their Uncertainties. I: Bayesian Statistical Framework, J Eng Mech 124 (1998) 455–461. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:4(455).

[15] L.S. Katafygiotis, J.L. Beck, Updating Models and Their Uncertainties. II: Model Identifiability, J Eng Mech 124 (1998) 463–467. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:4(463).

[16] C. Papadimitriou, G. Lombaert, The effect of prediction error correlation on optimal sensor placement in structural dynamics, Mech Syst Signal Process 28 (2012) 105–127. https://doi.org/10.1016/j.ymssp.2011.05.019.

[17] W. Ostachowicz, R. Soman, P. Malinowski, Optimization of sensor placement for structural health monitoring: a review, Struct Health Monit 18 (2019) 963–988. https://doi.org/10.1177/1475921719825601.

[18] Y. Tan, L. Zhang, Computational methodologies for optimal sensor placement in structural health monitoring: A review, Struct Health Monit (2019) 1–22. https://doi.org/10.1177/1475921719877579.

[19] A. Lye, A. Cicirello, E. Patelli, Sampling methods for solving Bayesian model updating problems: A tutorial, Mech Syst Signal Process 159 (2021) 107760. https://doi.org/10.1016/j.ymssp.2021.107760.

[20] J.N. Bericht, Bayesian techniques for inverse uncertainty quantification, 2019. https://doi.org/10.3929/ethz-b-000371683.

[21] Y. Xia, B. Chen, S. Weng, Y.Q. Ni, Y.L. Xu, Temperature effect on vibration properties of civil structures: A literature review and case studies, J Civ Struct Health Monit 2 (2012) 29–46. https://doi.org/10.1007/S13349-011-0015-7/FIGURES/17.

[22] W. Yao, X. Chen, W. Luo, M. Van Tooren, J. Guo, Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles, Progress in Aerospace Sciences 47 (2011) 450–479. https://doi.org/10.1016/J.PAEROSCI.2011.05.001.

[23] J. Legault, R.S. Langley, J. Woodhouse, Physical consequences of a nonparametric uncertainty model in structural dynamics, J Sound Vib 331 (2012) 5469–5487. https://doi.org/10.1016/j.jsv.2012.07.017.

[24] B.M. Ayyub, G.J. Klir, Uncertainty modeling and analysis in engineering and the sciences, Uncertainty Modeling and Analysis in Engineering and the Sciences (2006) 1–378. https://doi.org/10.1201/9781420011456/UNCERTAINTY-MODELING-ANALYSIS-ENGINEERING-SCIENCES-BILAL-AYYUB-GEORGE-KLIR.

[25] M. Faes, D. Moens, Recent Trends in the Modeling and Quantification of Non-probabilistic Uncertainty, Archives of Computational Methods in Engineering 2019 27:3 27 (2019) 633–671. https://doi.org/10.1007/S11831-019-09327-X.

[26] V. Barnett, Comparative Statistical Inference: Third Edition, Comparative Statistical Inference: Third Edition (2008) 1–381. https://doi.org/10.1002/9780470316955.

[27] D.R. Cox, Principles of statistical inference, Principles of Statistical Inference (2006) 1–219. https://doi.org/10.1017/CBO9780511813559.

[28] F.J. Samaniego, A Comparison of the Bayesian and Frequentist Approaches to Estimation, Springer New York, New York, NY, 2010. https://doi.org/10.1007/978-1-4419-5941-6.

[29] S.S. Rao, L. Berke, Analysis of Uncertain Structural Systems Using Interval Analysis, Https://Doi.Org/10.2514/2.164 35 (1997) 727–735. https://doi.org/10.2514/2.164.

[30] I. Elishakoff, Essay on uncertainties in elastic and viscoelastic structures: From A. M. Freudenthal's criticisms to modern convex modeling, Comput Struct 56 (1995) 871–895. https://doi.org/10.1016/0045-7949(94)00499-S.

[31] D. Moens, D. Vandepitte, A fuzzy finite element procedure for the calculation of uncertain frequency-response functions of damped structures: Part 1—Procedure, J Sound Vib 288 (2005) 431–462. https://doi.org/10.1016/J.JSV.2005.07.001.

[32] S.K. Au, J. Ching, J.L. Beck, Application of subset simulation methods to reliability benchmark problems, Structural Safety 29 (2007) 183–193. https://doi.org/10.1016/J.STRUSAFE.2006.07.008.

[33] E. Jahani, M.A. Shayanfar, M.A. Barkhordari, A new adaptive importance sampling Monte Carlo method for structural reliability, KSCE Journal of Civil Engineering 17 (2013) 210–215. https://doi.org/10.1007/S12205-013-1779-6/METRICS.

[34] A.M.J. Olsson, G.E. Sandberg, Latin Hypercube Sampling for Stochastic Finite Element Analysis, J Eng Mech 128 (2002) 121–125. https://doi.org/10.1061/(ASCE)0733-9399(2002)128:1(121).

[35] M. Mehta, Random Matrices, 3rd ed., 2004.

[36] A. Cicirello, R.S. Langley, Probabilistic assessment of performance under uncertain information using a generalized maximum entropy principle, Probabilistic Engineering Mechanics 53 (2018) 143–153. https://doi.org/10.1016/J.PROBENGMECH.2017.07.006.

[37] C. Soize, Random matrix theory for modeling uncertainties in computational mechanics, Comput Methods Appl Mech Eng 194 (2005) 1333–1366. https://doi.org/10.1016/j.cma.2004.06.038.

[38] A. Gray, A. Wimbush, M. de Angelis, P.O. Hristov, D. Calleja, E. Miralles-Dolz, R. Rocchetta, From inference to design: A comprehensive framework for uncertainty quantification in engineering with limited information, Mech Syst Signal Process 165 (2022) 108210. https://doi.org/10.1016/J.YMSSP.2021.108210.

[39] F. Igea, A. Cicirello, An Interacting Wasserstein Gradient Flow Strategy to Robust Bayesian Inference, (2024). https://arxiv.org/abs/2401.11607v1 (accessed January 23, 2024).

[40] Logging-While-Drilling EcoScope* Tool - Division of Marine and Large Programs, (2016). https://mlp.ldeo.columbia.edu/research/technology/schlumberger-lwd-tools/logging-while-drilling-ecoscope-tool/ (accessed August 16, 2020).

[41] M. Petyt, Introduction to finite element vibration analysis, 2nd ed., Cambridge University Press, Cambridge ; New York, 2010.

[42] Meirovitch, Parker, Fundamentals of Vibrations, 2001. https://doi.org/10.1115/1.1421112.

[43] B. Jacobs, A channel-based perspective on conjugate priors, Mathematical Structures in Computer Science 30 (2020) 44–61. https://doi.org/10.1017/S0960129519000082.

[44] D. V. Lindley, Bayesian Statistics, Society for Industrial and Applied Mathematics, 1972. https://doi.org/10.1137/1.9781611970654.

[45] E.G. Ryan, C.C. Drovandi, J.M. Mcgree, A.N. Pettitt, A Review of Modern Computational Algorithms for Bayesian Optimal Design, International Statistical Review 84 (2016) 128–154. https://doi.org/10.1111/insr.12107.

[46] J. Beck, B.M. Dia, L.F.R. Espath, Q. Long, R. Tempone, Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain, Comput Methods Appl Mech Eng 334 (2018) 523–553. https://doi.org/10.1016/j.cma.2018.01.053.

[47] D.R. Cavagnaro, J.I. Myung, M.A. Pitt, J. V. Kujala, Adaptive design optimization: a mutual information-based approach to model discrimination in cognitive science, Neural Comput 22 (2010) 887–905. https://doi.org/10.1162/NECO.2009.02-09-959.

[48] A.R. Cook, G.J. Gibson, C.A. Gilligan, Optimal observation times in experimental epidemic processes, Biometrics 64 (2008) 860–868. https://doi.org/10.1111/J.1541-0420.2007.00931.X.

[49] P. Müller, B. Sansó, M. De Iorio, Optimal Bayesian Design by Inhomogeneous Markov Chain Simulation, Https://Doi.Org/10.1198/016214504000001123 99 (2012) 788–798. https://doi.org/10.1198/016214504000001123.

[50] X. Huan, Y.M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, J Comput Phys 232 (2013) 288–317. https://doi.org/10.1016/j.jcp.2012.08.013.

[51] Z. Xu, Q. Liao, Gaussian process based expected information gain computation for bayesian optimal design, Entropy 22 (2020) 258. https://doi.org/10.3390/e22020258.

[52] K.J. Ryan, Estimating Expected Information Gains for Experimental Designs with Application to the Random Fatigue-Limit Model, Journal of Computational and Graphical Statistics 12 (2003) 585–603. https://doi.org/10.1198/1061860032012.

[53] L. Acerbi, Variational Bayesian Monte Carlo, Adv Neural Inf Process Syst 2018-December (2018) 8213–8223. http://arxiv.org/abs/1810.05558 (accessed May 12, 2021).

[54] L. Acerbi, Variational Bayesian Monte Carlo with Noisy Likelihoods, Advances in Neural Information Processing Systems 34 (2020). http://arxiv.org/abs/2006.08655 (accessed May 12, 2021).

[55] D.M. Blei, A. Kucukelbir, J.D. McAuliffe, Variational Inference: A Review for Statisticians, J Am Stat Assoc 112 (2017) 859–877. https://doi.org/10.1080/01621459.2017.1285773.

[56] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[57] S. Kullback, R.A. Leibler, On Information and Sufficiency, The Annals of Mathematical Statistics 22 (1951) 79–86. https://doi.org/10.1214/aoms/1177729694.

[58] M.P. Wand, J.T. Ormerod, S.A. Padoan, R. Frührwirth, Mean field variational bayes for elaborate distributions, Bayesian Anal 6 (2011) 847–900. https://doi.org/10.1214/11-BA631.

[59] G. Consonni, J.M. Marin, Mean-field variational approximate Bayesian inference for latent variable models, Comput Stat Data Anal 52 (2007) 790–798. https://doi.org/10.1016/j.csda.2006.10.028.

[60] R. Ranganath, S. Gerrish, D.M. Blei, Black Box Variational Inference, Journal of Machine Learning Research 33 (2013) 814–822. http://arxiv.org/abs/1401.0118 (accessed May 20, 2021).

[61] T. Campbell, X. Li, Universal Boosting Variational Inference, ArXiv (2019). http://arxiv.org/abs/1906.01235 (accessed May 12, 2021).

[62] S. Gershman, M. Hoffman, D. Blei, Nonparametric variational inference, Proceedings of the 29th International Conference on Machine Learning, ICML 2012 1 (2012) 663–670. http://arxiv.org/abs/1206.4665 (accessed May 12, 2021).

[63] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, D.M. Blei, Automatic Differentiation Variational Inference, Journal of Machine Learning Research 18 (2016) 1–45. http://arxiv.org/abs/1603.00788 (accessed May 12, 2021).

[64] A.C. Miller, N. Foti, R.P. Adams, Variational Boosting: Iteratively Refining Posterior Approximations, 34th International Conference on Machine Learning, ICML 2017 5 (2016) 3732–3747. http://arxiv.org/abs/1611.06585 (accessed May 16, 2021).

[65] F. Santambrogio, { Euclidean, Metric, and Wasserstein } Gradient Flows: an overview, Bull Math Sci 7 (2016) 87–154. https://doi.org/10.1007/s13373-017-0101-1.

[66] Luigi. Ambrosio, Nicola. Gigli, Giuseppe. Savaré, Gradient flows : in metric spaces and in the space of probability measures, Birkhäuser, 2005.

[67] R. Jordan, D. Kinderlehrer, F. Otto, The Variational Formulation of the Fokker--Planck Equation, Https://Doi.Org/10.1137/S0036141096303359 29 (2006) 1–17. https://doi.org/10.1137/S0036141096303359.

[68] F. Santambrogio, Optimal Transport for Applied Mathematicians, Springer International Publishing, Cham, 2015. https://doi.org/10.1007/978-3-319-20828-2.

[69] C. Soize, A comprehensive overview of a non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics, in: J Sound Vib, Academic Press, 2005: pp. 623–652. https://doi.org/10.1016/j.jsv.2005.07.009.

[70] A. Foster, M. Jankowiak, E. Bingham, P. Horsfall, Y.W. Teh, T. Rainforth, N. Goodman, Variational Bayesian Optimal Experimental Design, Adv Neural Inf Process Syst 32 (2019). https://doi.org/10.48550/arxiv.1903.05480.

[71] A. Foster, M. Jankowiak, M. O', M.§ Yee, W. Teh, T. Rainforth, A Unified Stochastic Gradient Approach to Designing Bayesian-Optimal Experiments, (2019). https://doi.org/10.48550/arxiv.1911.00294.

[72] K. Wu, T. O'Leary-Roseberry, P. Chen, O. Ghattas, Large-Scale Bayesian Optimal Experimental Design with Derivative-Informed Projected Neural Network, J Sci Comput 95 (2023) 1–20. https://doi.org/10.1007/S10915-023-02145-1/FIGURES/10.

[73] D.P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, M. Welling, Improving Variational Inference with Inverse Autoregressive Flow, Adv Neural Inf Process Syst (2016) 4743–4751. http://arxiv.org/abs/1606.04934 (accessed May 12, 2021).

[74] W.K. Hastings, Monte carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109. https://doi.org/10.1093/biomet/57.1.97.

[75] W.R. Gilks, P. Wild, Adaptive Rejection Sampling for Gibbs Sampling, Appl Stat 41 (1992) 337. https://doi.org/10.2307/2347565.

[76] S. Chib, Chapter 57 Markov chain Monte Carlo methods: computation and inference, in: Handbook of Econometrics, Elsevier, 2001: pp. 3569–3649. https://doi.org/10.1016/S1573-4412(01)05010-3.

[77] D. Kuhn, P.M. Esfahani, V.A. Nguyen, S. Shafieezadeh-Abadeh, Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning, Operations Research & Management Science in the Age of Analytics (2019) 130–166. https://doi.org/10.1287/educ.2019.0198.

[78] A. Lye, A. Cicirello, E. Patelli, An efficient and robust sampler for Bayesian inference: Transitional Ensemble Markov Chain Monte Carlo, Mech Syst Signal Process 167 (2022) 108471. https://doi.org/10.1016/J.YMSSP.2021.108471.

[79] T. Cui, J. Martin, Y.M. Marzouk, A. Solonen, A. Spantini, Likelihood-informed dimension reduction for nonlinear inverse problems, Inverse Probl 30 (2014) 114015. https://doi.org/10.1088/0266-5611/30/11/114015.