Accepted Manuscript

Probabilistic optimization of engineering system with prescribed target design in a reduced parameter space

A. Kundu, H.G. Matthies, M.I. Friswell

PII: S0045-7825(18)30165-8
DOI: https://doi.org/10.1016/j.cma.2018.03.041
Reference: CMA 11847

To appear in: *Comput. Methods Appl. Mech. Engrg.*

Received date : 5 August 2017
Revised date : 3 February 2018
Accepted date : 26 March 2018

Please cite this article as: A. Kundu, H.G. Matthies, M.I. Friswell, Probabilistic optimization of engineering system with prescribed target design in a reduced parameter space, *Comput. Methods Appl. Mech. Engrg.* (2018), https://doi.org/10.1016/j.cma.2018.03.041

Probabilistic optimization of engineering system with prescribed target design in a reduced parameter space

A. Kundu[a,*], H. G. Matthies[b], M. I. Friswell[c]

[a]*Applied and Computational Mechanics, Cardiff University, Queen's Buildings, The Parade, Cardiff, CF24 3AA, UK*
[b]*Institute of Scientific Computing, TU Braunschweig, Hans-Sommer Staße 65, Braunschweig 38106, Germany*
[c]*College of Engineering, Swansea University, Bay Campus Fabian Way, Crymlyn Burrows, Swansea SA1 8EN, UK*

Abstract

A novel probabilistic robust design optimization framework is presented here using a Bayesian inference framework. The objective of the proposed study is to obtain probabilistic descriptors of the system parameters conditioned on the user-prescribed target probability distributions of the output quantities of interest or figures of merit of a system. A criterion-based identification of a reduced important parameter space is performed from the typically high number of parameters modelling the stochastically parametrized physical system. The criterion can be based on sensitivity indices, design constraints or expert opinion or a combination of these. The posterior probabilities on the reduced or important parameters conditioned on prescribed target distributions of the output quantities of interest is derived using the Bayesian inference framework. The probabilistic optimal design proposed here offers the distinct advantage of prescribing probability bounds of the system performance functions around the optimal design points such that robust operation is ensured. The proposed method has been demonstrated with two numerical examples including the optimal design of a structural dynamic system based on user-prescribed target distribution for the resonance frequency of the system.

Keywords: Bayesian inference; robust design; probabilistic optimization; uncertainty propagation; stochastic structural dynamics; sensitivity analysis;

1. Introduction

Optimal design of engineering systems under uncertainty is important to ensure their fail-safe and robust performance. It is in this context that uncertainty quantification (UQ) of engineering systems has received significant attention in the past decade due to its effectiveness and applicability in handling parameteric uncertainty in engineering systems. The uncertainty may stem from the lack of knowledge of the parameter values or randomness inherent in them or error (noise or bias) associated with model predictions. Hence design optimization under uncertainty requires identifying optimal posterior distributions on the design parameters which satisfy the target design as opposed to working with point estimates.

The forward UQ problem forms an inner loop, and is often the first step, in the problem of robust optimization under uncertainty. It considers propagation of input parametric uncertainty using efficient computational methods to create a high-dimensional stochastic response surface in the parameter space. Some of the most efficient methods which fall under this category are stochastic Galerkin method using polynomial chaos basis functions [1–5], stochastic collocation techniques [6, 7], Monte-Carlo sampling

*Corresponding author. Tel: +44 (0)29 2087 5953
Email address: kundua2@cardiff.ac.uk (A. Kundu)

based methods (and its various improvements) [8–10] with significant effort devoted to constructing reduced order approximations of the true solution .

For robust optimization, different classes of methods have been developed, often aimed at specific engineering applications, to probe the search space of parametric models for optimum solutions. Genetic algorithm [11, 12], and its various improvements based on novel crossover and fuzzy algorithms, is amongst one of the most widely used search algorithm in optimization in various applications (such as shortest path, scheduling, amongst others). This falls under the umbrella of evolutionary algorithm and belongs to the category of global search heuristics where probabilistic strategies are employed to successively refine the solution based on a fitness criterion. Such heuristic approaches focus on randomization to ensure completeness in probing the search space in an attempt to find the global optimum. The heuristic (and metaheuristic optimization) techniques demonstrate their effectiveness by virtue of their success in finding the global optimum.

Bayesian methods for optimization can refer to different things in the context of parametric optimization. This includes methods which aim to minimize the expected deviation of a function from its extremum [13]. On the other hand, Gaussian process based studies as a means of approximating the behavior of function over the probe space (as indicated by prior distributions on hyperparameters modelling the process) is often a first step where few evaluations are used to learn the function behavior in the Bayesian sense followed by local optimization algorithms which utilize the posterior distribution of estimated functions to obtain the extremum [14]. In contrast to this sequential approach, there might also be adaptive resample and remodel strategies that aim to improve the response surface accuracy as the design space is probed in the optimization loop [15–18]. This includes a direct sampling and optimization process where the latter is realized as a direct consequence of the former required for generating the metamodel [19, 20]. However, this essentially constitutes a method of approximating a response surface using surrogate models rather than a probabilistic optimization based on adhoc target design criterion. Most uncertainty-based optimization problems focus on using probabilistic descriptors for input uncertainties and using black-box algorithms to probe the search space for acceptable solutions. In contrast probabilistic model informed optimizers aim to populate the solution set based on an initial guess of a set of random sample solutions and estimating (or updating) the new solution set based on acceptance criterion [21].

The present study elicits a novel approach whereby the Bayesian inference approach is utilized to perform robust optimization of an engineering system under uncertainty such that optimal probability distributions of the system parameters are identified which would meet the desired optimal design criterion specified (again) by full distributions on figures of merit of the system. Thus we are using the framework of inverse system identification to optimize the parameter distribution of the model problem. This can be regarded as an application of the Bayesian regularization method where the regularized functional includes the chosen prior and posterior distributions of the system performance functions. The benefits of this Bayesian regularization based optimization become clear when considering the fact that the optimal parameter distributions which produces the desired distributions of the system performance functions ensure their robust operation around the optimal points. This is facilitated by going beyond point estimates typical in deterministic analysis which do not represent the confidence in the optimal design or that of meeting the target design distribution [22]. The impact of this in real engineering applications is significant since it allows for reliable risk-informed decision-making [23].

The paper is organized as follows. Section 2 gives the problem definition using a step by step development of the problem setting starting from forward problem of uncertainty propagation to Bayesian system identification based on which the inverse optimization problem is defined. The function space elicitation of the problem definition and the solution is presented in section 3 where the inverse optimization problem is defined in terms of the theoretical framework of Kolmogorov's conditional expectation. Section 4 deals with the proposed methodology applied to the robust optimization of a structural dynamic

system with random elastic and geometric parameters as well as for a non-linear six-dimensional Hartmann function. The main conclusions of the study along with the potential directions for future work are given in section 5.

2. Problem setting

In this section we define the main components necessary to lay the probabilistic framework of Bayesian optimization under uncertainty. We begin by defining the forward problem, giving an input-output mapping of the solution to the set of input parameters. Following this, the layout the inverse or what is known as the system-identification problem under uncertainty is given. Based on these the Bayesian optimization methodology is established which is defined in the context of the forward and inverse problem.

2.1. Forward problem

In order to present a generic discussion, we start with the forward problem defined on a d-dimensional domain $\mathcal{D} \subset \mathbb{R}^d$ by an operator $\mathscr{L}$ as

$$\mathscr{L}(u; q) = f \tag{1}$$

where the system is parametrized by a set q. The system solution is represented by u under external forcing function f. In real-life engineering applications, uncertainties exist around parameter values as well as in forcing functions which are incorporated in the parameter set q with probabilistic descriptors. For the sake of simplicity, we consider the forcing function to be deterministic which does not limit the scope of applicability of the proposed method because forcing uncertainty is classified as additive uncertainty compared to the multiplicative uncertainty contained in parameters q in Eqn. (1). The forward problem consists of propagating the uncertainty in q to the system response u. This relies on numerical approximation techniques to represent the solution in the stochastic parameter space and when used in conjunction with finite element methods are collectively referred to as stochastic finite element methods [1, 24].

The elements of the parameter set q in Eqn. (1) are taken to be square integrable (having a finite variance) and is modeled with a set of independent random variables $\boldsymbol{\xi}(\theta) = \{\xi_1, \ldots, \xi_n\}$ in the sample space Θ such that $q : \boldsymbol{\xi} \rightarrow q$ and $\boldsymbol{\xi}(\theta) \in L^2(\Theta, \mathcal{F}, \mathbb{P})$. The probability space triplet $(\Theta, \mathcal{F}, \mathbb{P})$ consists of elementary events θ in the sample space Θ as $\theta \in \Theta$, with the associated σ-algebra $\mathcal{F}$. The pair $(\Theta, \mathcal{F})$ is a measurable space with $\mathbb{P}$ being the probability measure attached to it.

Thus u can be approximated in the tensor product space as $\mathscr{U} = \mathscr{X} \otimes \mathscr{S}$ where $\mathscr{X} \subset L^2(\mathcal{D})$ denotes the subspace spanned by the spatial basis functions, while $\mathscr{S} \subset L^2(\Theta)$ is a subspace spanned by finite order stochastic basis functions. The solution is approximated as

$$u(\mathbf{r}, \boldsymbol{\xi}(\theta)) = \sum_{i,j} u_{ij} \mathscr{H}_i(\boldsymbol{\xi}(\theta)) \mathscr{N}_j(\mathbf{r}) \quad \text{where} \quad \mathscr{H}_i \in \mathscr{S}, \ \mathscr{N}_j \in \mathscr{X} \quad \forall i, j \tag{2}$$

in the most general case where $\theta \in \Theta$ and $\mathbf{r} \in \mathcal{D}$. Based on the differential operator and the associated probability measurable, the subspace in which the solution is sought can be defined to have certain degree of continuity (as is the case for diffusion operator, where the solution is sought in a Sobolev space with at least first order continuity and a finite L^2 norm).

The stochastic Galerkin method for approximating the system solution relies on approximating the response u in a finite dimensional space of basis functions. We assume that $u \in \mathfrak{U}$ which is a Hilbert space endowed with an inner product structure to be defined later in this paper. The problem is assumed to be well posed in the Hadamard sense.

The weak form of the parametrized stochastic pde can then be expressed in terms of the bilinear ($\mathcal{A}$) and linear ($\mathcal{B}$) forms as

$$\mathcal{A}(u, v; \theta) = \mathcal{B}(v; \theta) \quad \forall v \in \mathscr{X} \otimes \mathscr{S} \tag{3}$$

$$\text{where} \quad \mathcal{A}(u, v) = \mathbb{E}\left[a(u(\theta), v(\theta); \theta)\right]; \quad \mathcal{B}(v) = \mathbb{E}\left[b(v(\theta); \theta)\right]; \quad \text{with} \quad \mathbb{E}\left[\cdot; \theta\right] = \int_\theta [\cdot; \theta]\, d\mathbb{P}(\theta)$$

The Galerkin method ensures that the approximate solution $u \in \mathscr{X} \otimes \mathscr{S}$ minimizes the error in the sense

$$\|u_{\text{true}} - u\|_e \leq \|u_{\text{true}} - v\|_e \quad \forall v \in \mathscr{X} \otimes \mathscr{S} \quad \text{where} \quad \|\cdot\|_e = [\mathcal{A}(\cdot, \cdot)]^{1/2} \tag{4}$$

The above constitutes the forward uncertainty propagation problem which provides a stochastic mapping between the input parameter space and the output response variables.

2.2. Sensitivity analysis: statistically important predictors

The various sources of uncertainty captured in the parameter $\boldsymbol{\xi}(\theta)$ lead to a high-dimensional or over-parametrized system and it is essential to identify the statistically important predictors which significantly affect the model outputs. Sensitivity indices [25] are the most commonly used measures to identify and rank the parameters in their order of importance. The sensitivity estimation strategies can be primarily categorized into local (concerned with output variability around a nominal parameter value) and global (overall variability in model output over the entire parameter range) sensitivity measures. An exhaustive discussion on the various sensitivity analysis tools can be found in the literature [26, 27]. In this article we have focussed on and worked with the global sensitivity approach based on Sobol' indices. This section gives a brief description of the method. It is to be noted however that the applicability of the robust optimization approach proposed in this work is not limited by a particular choice of sensitivity analysis method.

The output $u(:, \boldsymbol{\xi}(\theta))$ is a function of the independent input parameters $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_n\}$ in the n-dimensional hyperspace. Hence, a quantitative estimate of the sensitivity of the model output $u(:, \boldsymbol{\xi}(\theta))$ to the input parameters is provided by the Sobol's sensitivity indices [28]. These provide an estimate of the relative importance of the input variables (or a group of input variables) on the output quantities of interest (qoi). If a stochastic response quantity u expressed as a function of the input uncertainty contained in $\boldsymbol{\xi}(\theta)$ is expanded as

$$u(\boldsymbol{\xi}) = u_0 + \sum_{i=1}^{n} \bar{u}_i(\xi_i) + \sum_{1 \leq i < j \leq n} \bar{u}_{ij}(\xi_i, \xi_j) + \ldots + \bar{u}_{1\ldots n}(\xi_1, \ldots, \xi_n) \tag{5}$$

then it is seen that the expansion terms successively gives the dependence of the stochastic functions on the individual random variables as well as their interaction with other variables. Here each term $\bar{u}_{i_1 \ldots i_s}(\xi_{i_1}, \ldots, \xi_{i_s})$ denotes a sum of the collection of all terms involving random variables $\{\xi_{i_1}, \ldots, \xi_{i_s}\}$. There are a total of $2^n - 1$ terms in the above equation and the set of indices in the expansion is contained in $\mathscr{I}$ (with cardinal $2^n - 1$). Equation (5) is termed as the ANOVA (Analysis of Variance) decomposition [25] of the function $u(\boldsymbol{\xi})$. The term u_0 is the mean value of the function while each term of the summand has zero expectation and are orthogonal to each other, i.e.

$$\int_\Theta \bar{u}_I \, \mathrm{d}\mathbb{P}_{\boldsymbol{\xi}} = 0 \quad \text{and} \quad \int_\Theta \bar{u}_I \bar{u}_J \, \mathrm{d}\mathbb{P}_{\boldsymbol{\xi}} = 0 \quad \forall \; I, J \in \mathscr{I}, \;\; I \neq J \tag{6}$$

where I, J denote specific terms in expansion of $u(\boldsymbol{\xi})$ in Eqn. (5) and $\mathrm{d}\mathbb{P}_{\boldsymbol{\xi}}$ is the joint distribution on parameter set $\boldsymbol{\xi}$. Assuming $\boldsymbol{\xi}$ are independent, then the simple relation $\mathbb{P}_{\boldsymbol{\xi}} = \prod_i \mathbb{P}_{\xi_i}$ holds. When the random variables are correlated, a decomposition of the correlation matrix is performed to identify a denumerable set of independent random variables to model the variability. Extensive research in this specific area

has been reported which include methods ranging from non-standard transformations which goes beyond simple mixing to obtain uncorrelated variables to arbitrary polynomial chaos expansion method [29–31]. Going back to Eqn. (5), the successive terms of the expansion maybe derived analytically and is unique when the functions $u(\boldsymbol{\xi}(\theta))$ is square-integrable in Θ. When the input random variables are independent of each other the total variance of $u(\boldsymbol{\xi}(\theta))$ i.e. $\Gamma = \mathrm{Var}\left[u(\boldsymbol{\xi}(\theta))\right]$ can be written as

$$\Gamma = \sum_{i=1}^{n} \Gamma_i + \sum_{1\le i<j\le n} \Gamma_{ij} + \ldots + \Gamma_{1,\ldots,n} = \int_{\Theta} u^2(\boldsymbol{\xi})\, \mathrm{d}\mathbb{P}_{\boldsymbol{\xi}} - u_0^2 \quad \text{where} \quad \Gamma_I = \int_{\Theta} \bar{u}_I^2\, \mathrm{d}\mathbb{P}_{\boldsymbol{\xi}} \tag{7}$$

for all $I \in \mathscr{I}$. Here $\Gamma_{i_1,\ldots,i_k}$ are the *partial variances* in the expression. The expression for partial variances follows from Eqns. (5)–(6). The Sobol' indices are defined as

$$\mathcal{S} = \sum_{I\in\mathscr{I}} \mathcal{S}_I = \sum_{i=1}^{n} \mathcal{S}_i + \sum_{1\le i<j\le n} \mathcal{S}_{ij} + \ldots + \mathcal{S}_{1,\ldots,n} = 1 \quad \text{where} \quad \mathcal{S}_I = \frac{\Gamma_I}{\Gamma} \quad \forall\, I \in \mathscr{I} \tag{8}$$

The relative importance of the individual input random variables in the response quantities are provided by the values of the Sobol' indices. The first order Sobol' index associated with the i-th random variable is given by $\mathcal{S}_i = \Gamma_i/\Gamma,\ \forall i = 1,\ldots,n$ (note that the first order Sobol' index is defined with the subscript i and not I). This is equivalent to defining it as

$$\mathcal{S}_i = \frac{\mathrm{Var}\left[\mathbb{E}\left[\mathbf{u}(\boldsymbol{\xi})\,\middle|\,\xi_i\right]\right]}{\mathrm{Var}\left[\mathbf{u}(\boldsymbol{\xi})\right]} = \frac{\Gamma_i}{\Gamma}, \qquad i = 1,\ldots,n \tag{9}$$

which is useful for sample based uncertainty propagation schemes (such as Monte Carlo or collocation methods). The total order Sobol' index $\mathcal{S}_i^T$ of a particular input random variable ξ_i gives an estimation of the total contribution of a random input, including interaction terms, to the output variance and is calculated as

$$\mathcal{S}_i^T = \sum_{\mathscr{I}_i} \Gamma_{\mathscr{I}_i} \quad \text{where} \quad \mathscr{I}_i \subset \mathscr{I} \tag{10}$$

such that the set of multi-indices in $\mathscr{I}_i$ always include the random variable $\boldsymbol{\xi}_i$ (both independent $\mathcal{S}_i$ and interaction terms).

The Sobol' indices can be calculated analytically from a finite order chaos expansion [32] where the terms u_I in Eqn. (5) can be simply obtained by grouping the terms of a finite order polynomial chaos expansion. The variance measures Γ_I are obtained from the square integrable products of the chaos expansion functions $|\mathscr{H}_I(\boldsymbol{\xi}(\theta))|^2_{L^2(\theta)}$ (following the orthogonality property). Depending on the nature of parameter dependence of the problem both the first and total order Sobol' indices may be pertinent for identifying the important parameter space as will be seen later in this work.

2.3. Inverse problem

We consider an observation operator $y = \mathcal{Y}(q;u)$ which connects the output qoi y to a set of model forecast values and the measurements corresponding to these are denoted by $\bar{y} \in \bar{\mathbf{y}}$. It is commonly expressed as $\bar{\mathbf{y}} = y + \varepsilon$ where ε is a discrepancy term expressing the difference between the simulated output and the measured data. In general ε is assumed to be independent of q. The observation y is obtained by the mapping: $\mathcal{Y} : (q;u) \to y$ which gives

$$y = \bar{\mathbf{y}} - \varepsilon = \mathcal{Y}(q;u) \quad \text{where} \quad y \in \mathscr{Y} \tag{11}$$

where $\mathscr{Y}$ is the vector space containing y. The most common engineering problem is to infer about the statistical properties of the set of input parameters q given the simulated observations y and measurements $\bar{\mathbf{y}}$. For this the Bayesian system identification framework is utilized which is discussed in this section.

It is important here to comment on the prior probabilities imposed on the input parameter set which elicits the prior belief on the uncertainty of the parameters. Many works have proposed which highlight systematic methods for elicitation of prior uncertainty, including parametric and non-parametric models [33–37]. In presence of significant amount of data it is standard to use 'noninformative priors' without much risk, since the evidence would negate any effect of prior. However, data is limited/scarce, the priors play an important role and in such cases it is essential to have systematic and consistent ways of choosing priors. Methods ranging from expert elicitation with probabilistic descriptors [38] to entropy-based conditional elicitations [39] have been used. In the context of this work, we have used uniform distributions as prior for the input parameters which reflects our knowledge about the acceptable range of variation of these parameters and to avoiding weighing any area of the parameter space. It should also be mentioned that the proposed method can perform satisfactorily with different choices of priors (or hyperpriors) in the parameters since no limiting assumptions have been made on the priors for the successful working of this algorithm.

The problem being tackled here is defined as follows. Given the interval of variation and the prior probability $\mathbb{P}(\boldsymbol{\xi})$ on this interval for the n-dimensional input parameter vector $\boldsymbol{\xi} \in \mathbb{R}^n$, the system output due to the observation operator $\mathcal{Y}$ follows the prior distribution $\mathbb{P}(y)$. We assume that the target distribution for output y is specified by another random variable $\bar{y} : \Theta \to \bar{y}$. The problem is to find the optimal distribution on the input parameters $\boldsymbol{\xi}$ (or a subset of input parameters $\boldsymbol{\xi}_r \in \mathbb{R}^r,\ r < n$) such that when the uncertainty in $\boldsymbol{\xi}$ is propagated to $y(\boldsymbol{\xi})$, then y and $\bar{y}$ are as *close* as possible. The closeness is defined by the L^2 distance in the probability space as $d(y, \bar{y}) = \sqrt{\int_\Theta (y - \bar{y})^2 \, \mathrm{d}\mathbb{P}} = \|y - \bar{y}\|_{L^2(\Theta, \mathcal{F}, \mathbb{P})}$

Mathematically, the problem can be stated as: find a map $f : \boldsymbol{\xi} \to \mathbb{R}^n$ such that

$$f(\boldsymbol{\xi}) = \arg \inf_{\boldsymbol{\xi} \in \mathbb{R}^n} \|y(f(\boldsymbol{\xi})) - \bar{y}\|_{L^2(\Theta, \mathcal{F}, \mathbb{P})} \tag{12}$$

Following Radon-Nikodym theorem [40] the map f defines a new measure (say ν) on the space $(\Theta, \mathcal{F})$ which is absolutely continuous with $\mathbb{P}$ and is defined as $f = \mathrm{d}\nu / \mathrm{d}\mathbb{P}$.

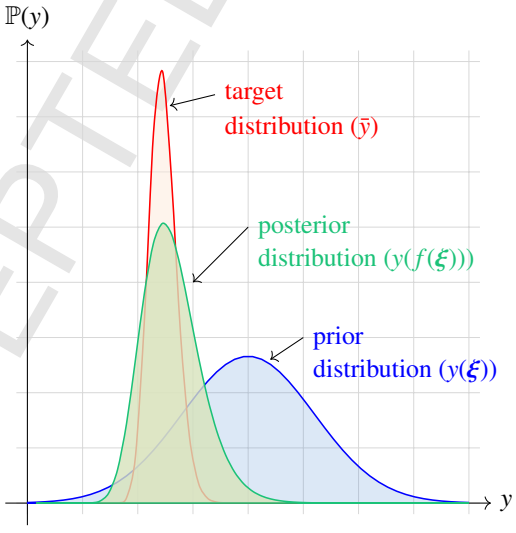


Figure (1): The schematic diagram describing the problem definition of prior and target distribution

This is illustrated in Fig. (1) where the *prior distribution* on y given the prior probability imposed on $\boldsymbol{\xi}$ is shown in blue while the target distribution for y (denoted as $\bar{y}$ in Eqn. (12)) is given in red. The objective is to find the optimal posterior distributions on $\boldsymbol{\xi}$ which will produce the best approximation

(shown in green) of the target distribution. The target distribution on $\bar{y}$ is a user-specified criterion on physical qoi and it can be in the form of

- samples from an arbitrary distribution as

$$\bar{\mathbf{y}} = \{\bar{y}_1, \ldots, \bar{y}_s\}^T \text{ where } \bar{\mathbf{y}} \sim \mathbb{P}(\bar{\mathbf{y}} \,|\, \alpha_{\bar{y}}) \quad \forall i = 1, \ldots, s \tag{13}$$

  where $\mathbb{P}(\bar{\mathbf{y}} \,|\, \alpha_{\bar{y}})$ is a probability distribution characterized by hyperparameter set $\alpha_{\bar{y}}$.

- constraints specified as upper/lower limits on $\bar{y}$ such that

$$\bar{\mathbf{y}} = \{\bar{y}_1, \ldots, \bar{y}_s\}^T \text{ where } \bar{y}_l \leq \bar{y}_i \leq \bar{y}_u \quad \forall i = 1, \ldots, s \tag{14}$$

  where $\bar{y}_u$ and $\bar{y}_l$ are designer prescribed upper and lower limits on observable outputs.

- extremizing objective functions $g(\bar{y})$ with sample values $\bar{\mathbf{y}} = \{\bar{y}_1, \ldots, \bar{y}_s\}^T$ derived from the criterion

$$\bar{\mathbf{y}} = \{\bar{y}_i : \|g(\bar{y}_i) - g_t\| < \epsilon \quad \forall i = 1, \ldots, s\} \tag{15}$$

  where g_t is a target value for the objective function and ϵ is a random variable with an associated probability distribution. ϵ can be considered as a loss term in regularization which defines the tightness of desired fit.

The samples $\bar{\mathbf{y}} = \{\bar{y}_1, \ldots, \bar{y}_s\}^T$ are derived from the constraints or optimization criterion imposed on the output qoi $\bar{y}$. This would form the dataset on which the prior probabilities imposed on the input parameter space $\boldsymbol{\xi}$ will be conditioned following the criterion in Eqn. (12).

2.4. Bayesian system identification

Following from the discussion presented in the previous section, we take $\mathcal{Y}$ as the observation operator which defines output data $y = \mathcal{Y}(q; u)$. The target distribution is specified by samples of $\bar{\mathbf{y}} \sim \mathbb{P}(\bar{y} \,|\, \alpha_{\bar{y}})$. The prior distribution on q is characterized by a set of hyperparameters α_q such that $q \sim \mathbb{P}(q \,|\, \alpha_q)$. Under classical Bayesian inference setting the problem of inferring optimal posterior distributions on parameter set q is given by

$$\mathbb{P}(q \,|\, \mathcal{Y}, \bar{\mathbf{y}}, \alpha_{\bar{y}}, \alpha_q) = \frac{\mathbb{P}(\bar{\mathbf{y}} \,|\, \mathcal{Y}, q, \alpha_{\bar{y}}) \mathbb{P}(q \,|\, \alpha_q)}{\mathbb{P}(\bar{\mathbf{y}} \,|\, \mathcal{Y}, \alpha_{\bar{y}})} \propto \mathcal{L}_{\bar{y}|\alpha_{\bar{y}}}(q) \mathbb{P}(q \,|\, \alpha_q) \tag{16}$$

where $\mathcal{L}_{\bar{y}|\alpha_{\bar{y}}}(q)$ is the likelihood function defined by $\mathcal{L}_{\bar{y}|\alpha_{\bar{y}}}(q) = \mathbb{P}(\bar{\mathbf{y}} \,|\, \mathcal{Y}, q, \alpha_{\bar{y}})$. Hence the posterior predictive distribution of observable output samples is given as

$$\mathbb{P}(y^* \,|\, \mathcal{Y}, \bar{\mathbf{y}}, \alpha_{\bar{y}}) = \int_q \int_{\alpha_q} \mathbb{P}(y^* \,|\, \mathcal{Y}, \bar{\mathbf{y}}, \alpha_{\bar{y}}) \mathbb{P}(q \,|\, \alpha_q) \mathbb{P}(\alpha_q) \; \mathrm{d}q \; \mathrm{d}\alpha_q \tag{17}$$

where the parameter q and the hyperparameter α_q has been marginalized out from the posterior predictive distribution.

The likelihood function is shaped by the observation data set $\bar{\mathbf{y}}$ since it is governed by the statistics of the observation dataset $\bar{\mathbf{y}}$. Assuming that the measurement are independent and identically distributed (iid), we can write from Eqn. (16)

$$\mathcal{L}_{\bar{y}|\alpha_{\bar{y}}}(q) = \prod_{i=1}^{s} \mathcal{L}_{\bar{y}_i|\alpha_{\bar{y}}}(q) \tag{18}$$

where $\mathcal{L}_{\bar{y}_i \,|\, \alpha_{\bar{y}}}(\theta)$ denotes the likelihood function associated with each data point i. For each datapoint i the likelihood function $\mathcal{L}_{\bar{y}_i \,|\, \alpha_{\bar{y}}}(q)$ is calculated as a function of the observation operator $\mathcal{Y}$. The probability measure associated with the response surface of observation operator $y = \mathcal{Y}(q; u)$ dictates the likelihood function. For every value of the parameter q, the distribution on the output y is given as $\mathbb{P}(y \,|\, q, \mathcal{Y})$. If the measurement point $\bar{y}_i$ is within the support of $\mathbb{P}(y \,|\, q, \mathcal{Y})$ then the likelihood for $\bar{y}_i$ is given as

$$\mathbb{P}(\bar{\mathbf{y}} \,|\, \mathcal{Y}, q, \alpha_{\bar{y}}) \equiv \prod_{i=1}^{s} \mathbb{P}(y = \bar{y}_i \,|\, \mathcal{Y}, q, \alpha_{\bar{y}}) \tag{19}$$

and the marginal likelihood (after marginalizing the hyperparameter $\alpha_{\bar{y}}$ associated with the dataset $\bar{\mathbf{y}}$) is non-zero and equals

$$\mathcal{L}_{\bar{y}}(q) = \int_{\alpha_{\bar{y}}} \mathcal{L}_{\bar{y} \,|\, \alpha_{\bar{y}}}(q)\, \mathbb{P}(\alpha_{\bar{y}})\, \mathrm{d}\alpha_{\bar{y}} = \int_{\alpha_{\bar{y}}} \mathbb{P}(\bar{\mathbf{y}} \,|\, \mathcal{Y}, q, \alpha_{\bar{y}}) \mathbb{P}(\alpha_{\bar{y}})\ \mathrm{d}\alpha_{\bar{y}} = \prod_{i=1}^{s} \int_{\alpha_{\bar{y}}} \mathbb{P}(y = \bar{y}_i \,|\, \mathcal{Y}, q, \alpha_{\bar{y}}) \mathbb{P}(\alpha_{\bar{y}})\ \mathrm{d}\alpha_{\bar{y}} \tag{20}$$

A discusssion of Eqns. (18)–(20) is given here to provide an insight into the calculation of likelihood function. The likelihood of a data point $\bar{y}_i \in \bar{\mathbf{y}}$ is calculated for each point q in the multidimensional parameter space. For each value of q, the distribution of the simulated output y is estimated in the first step, following which the quantity $\mathbb{P}(y = \bar{y}_i \,|\, \mathcal{Y}, q, \alpha_{\bar{y}})$ is evaluated based on its distance from the distribution. In the special case when the distribution of the multidimensional output vector for each q has a joint normal distribution with a mean vector $\mathbb{E}[y(q)]$ and a covariance matrix C_q, the negative log-likelihood is given as the squared Mahalanobis distance

$$-\ln[\mathbb{P}(y = \bar{y}_i \,|\, \mathcal{Y}, q, \alpha_{\bar{y}})] \ \propto\ [\bar{y}_i - \mathbb{E}[y(q)]]^T C_q^{-1} [\bar{y}_i - \mathbb{E}[y(q)]] \tag{21}$$

It is to be noted here that $\mathbb{E}[y(q)]$ is a function of q and the expectation is evaluated from the distribution of the simulated output y for each point in the input parameter space q. However, in the most general case the distribution of simulated y for specific values of the input parameter q would be an arbitrary probability distribution which would be fitted to a density function $f(y_q)$ using kernel density estimators. If a kernel $\mathcal{K}(\bullet)$ is defined with a smoothing parameter h with the samples of $y_j(q)$ from the true distribution of the observation vector $\pi(y(q))$, then from Eqn. (19) we can write

$$\mathbb{P}(y = \bar{y}_i \,|\, \mathcal{Y}, q, \alpha_{\bar{y}}) = \hat{\pi}(y(q) = \bar{y}_i)\ \ \forall i = 1, \ldots, s \quad \text{where} \quad \hat{\pi}(y(q)) = \frac{1}{sh} \sum_{j=1}^{s} \mathcal{K}\left(\frac{y_j - y}{h}\right) \tag{22}$$

where $\hat{\pi}(y(q))$ is an approximation of $\pi(y(q))$ with kernel density estimate.

With the likelihood function estimated using the above methods, the samples from the posterior distribution $\mathbb{P}(q \,|\, \mathcal{Y}, \bar{\mathbf{y}}, \alpha_{\bar{y}}, \alpha_q)$ in Eqn. (16) for various Bayesian model fitting algorithms can be used to sample from the target distributions. The Markov Chain Monte Carlo (MCMC) algorithm in its most generic form, as well as with various enhanced implementations for reduced computational overhead, would be applicable here. The MCMC convergence rate depends on the dimension of the parameter space and the nature of the posterior distributions (such as multi-modal distributions, which are challenging to handle). These require advanced implementations of the MCMC method a detailed review of which is beyond the scope of this work (please see [10, 41, 42]).

2.5. Identification on reduced parameter space

The uncertainty in the simulated output can result from the consideration of a reduced important parameter space, as shown in Fig. (2). We denote the reduced input parameter space $\boldsymbol{\xi}_r \in \mathbb{R}^r$ and its

complementary set $\boldsymbol{\xi}^c \in \mathbb{R}^{n-r}$ such that $\mathcal{S}(\xi_i) > \mathcal{S}(\xi_j) \quad \forall\ \xi_i \in \boldsymbol{\xi}_r,\ \xi_j \in \boldsymbol{\xi}^c$ where $\mathcal{S}(\xi_i)$ is the global sensitivity index (as discussed in Sec. 2.2). This implies the set $\boldsymbol{\xi}_r$ consists of all parameters with higher sensitivity indices than in $\boldsymbol{\xi}^c$. It must be mentioned that the applicability of the proposed method is not limited by the choice of the metric (such as the global sensitivity measure, in this case) used for constructing the reduced parameter space.

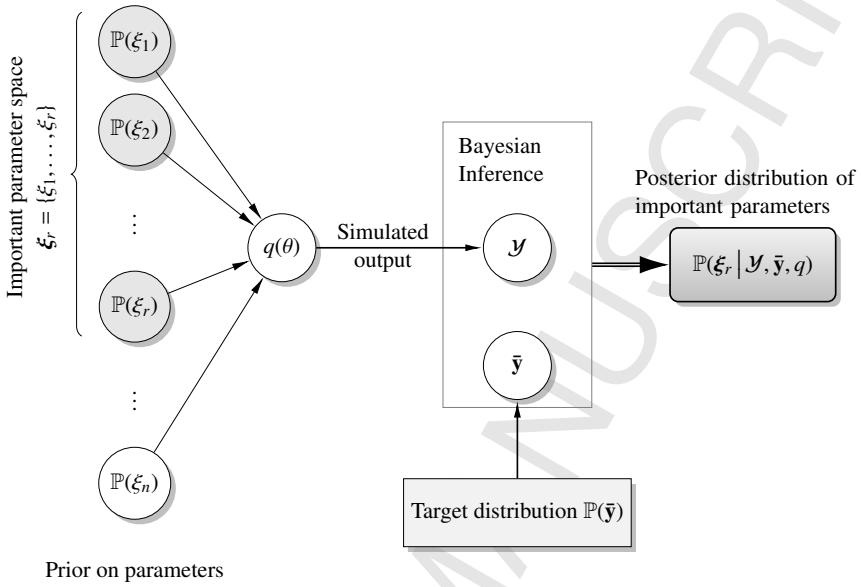


Figure (2): Schematic diagram of the Bayesian model fitting on a reduced important parameter space.

The probability distribution on the physical parameter q which depends on the basic random variables $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_n\}$ along with a set of hyperparameter α_q (which can describe the associated noise) can be expressed as $\mathbb{P}(q \,|\, \alpha_q)$ which depends on the set of underlying parameters $\boldsymbol{\xi}$. The probability distribution of q given a specified value for the set $\boldsymbol{\xi}_r$ is given as

$$q_{\boldsymbol{\xi}_r} \sim \mathbb{P}(q \,|\, \boldsymbol{\xi}_r, \alpha_q) = \mathbb{P}(q_{\boldsymbol{\xi}_r} \,|\, \alpha_q) \quad \text{where} \quad q_{\boldsymbol{\xi}_r} : \boldsymbol{\xi}_r \to q(\boldsymbol{\xi}_r) \quad \text{and} \quad q_{\boldsymbol{\xi}_r} \in (\Theta, F, \mathbb{P}), F \subset \mathcal{F}$$
$$\text{such that} \quad \mathbb{E}\left[q_{\boldsymbol{\xi}_r}\right] = \int_{\boldsymbol{\xi}^c} \mathbb{P}(q_{\boldsymbol{\xi}_r} \,|\, \alpha_q)\mathbb{P}(\boldsymbol{\xi}^c)\, \mathrm{d}\boldsymbol{\xi}^c \tag{23}$$

The random variable $q_{\boldsymbol{\xi}_r}$ follows the distribution $q_{\boldsymbol{\xi}_r} \sim \mathbb{P}(q_{\boldsymbol{\xi}_r} \,|\, \alpha_q)$ for each point in $\boldsymbol{\xi}_r$ and is a function of $\boldsymbol{\xi}_r$. This is discussed rigourously in light of Kolmogorov's conditional expectation in section 3. It is to be noted that the variability in $q_{\boldsymbol{\xi}_r}$ at a particular $\boldsymbol{\xi}_r$ is due to the dependence of q on the unimportant random variables $\boldsymbol{\xi}^c$. The probability space associated with $q_{\boldsymbol{\xi}_r}$ has the Borel σ-algebra F which is a subset of $\mathcal{F}$ defined in Sec. 2.1. Following from Eqn. (16), the posterior probability on $q_{\boldsymbol{\xi}_r}$ is

$$\mathbb{P}(q_{\boldsymbol{\xi}_r} \,|\, \mathcal{Y}, \bar{\mathbf{y}}, \alpha_{\bar{\mathbf{y}}}, \alpha_q) = \frac{\mathbb{P}(\bar{\mathbf{y}} \,|\, \mathcal{Y}, q_{\boldsymbol{\xi}_r})\mathbb{P}(\mathcal{Y}(q_{\boldsymbol{\xi}_r}) \,|\, \alpha_{\bar{\mathbf{y}}})\mathbb{P}(q_{\boldsymbol{\xi}_r} \,|\, \alpha_q)}{\mathbb{P}(\bar{\mathbf{y}} \,|\, \mathcal{Y}, \alpha_{\bar{\mathbf{y}}})} \propto \mathcal{L}_{\bar{\mathbf{y}}|\alpha_{\bar{\mathbf{y}}}} \mathbb{P}\left(q_{\boldsymbol{\xi}_r} \,|\, \alpha_q\right) \tag{24}$$

The likelihood function, defined here as $\mathcal{L}_{\bar{\mathbf{y}}|\alpha_{\bar{\mathbf{y}}}} = \mathbb{P}(\bar{\mathbf{y}} \,|\, \mathcal{Y}, q_{\boldsymbol{\xi}_r})\mathbb{P}(\mathcal{Y}(q_{\boldsymbol{\xi}_r}) \,|\, \alpha_{\bar{\mathbf{y}}})$, and the prior probability in the above equation does not have a straightforward relationship with the likelihood and prior probabilities described in Eqn. (16). It must be emphasized that the posterior distribution $\mathbb{P}(q_{\boldsymbol{\xi}_r} \,|\, \mathcal{Y}, \bar{\mathbf{y}}, \alpha_{\bar{\mathbf{y}}}, \alpha_q)$ in Eqn. (24) is a function of $\boldsymbol{\xi}_r$. Hence,

$$\mathbb{P}\left(\boldsymbol{\xi}_r \,|\, q_{\boldsymbol{\xi}_r}, \bar{\mathbf{y}}, \mathcal{Y}\right) \propto \mathbb{P}\left(\bar{\mathbf{y}}, \mathcal{Y} \,|\, q_{\boldsymbol{\xi}_r}\right) \mathbb{P}\left(q_{\boldsymbol{\xi}_r} \,|\, \boldsymbol{\xi}_r, \alpha_q\right) \mathbb{P}\left(\boldsymbol{\xi}_r\right) \propto \mathcal{L}_{\bar{\mathbf{y}}|\alpha_{\bar{\mathbf{y}}}} \mathbb{P}\left(q_{\boldsymbol{\xi}_r} \,|\, \boldsymbol{\xi}_r, \alpha_q\right) \mathbb{P}\left(\boldsymbol{\xi}_r\right) \tag{25}$$

Combining Eqns. (24) and (25), we can write

$$\mathbb{P}\left(\boldsymbol{\xi}_r \,\middle|\, \bar{\mathbf{y}}\right) = \int_{q_{\boldsymbol{\xi}_r}} \int_{\alpha_{\bar{y}}} \mathbb{P}(\boldsymbol{\xi}_r \,|\, q_{\boldsymbol{\xi}_r}, \bar{\mathbf{y}}, \mathcal{Y})\, \mathbb{P}(q_{\boldsymbol{\xi}_r} \,|\, \mathcal{Y}, \bar{\mathbf{y}})\, \mathbb{P}(\mathcal{Y}(q_{\boldsymbol{\xi}_r}) \,|\, \alpha_{\bar{y}})\ \mathrm{d}q_{\boldsymbol{\xi}_r}\ \mathrm{d}\alpha_{\bar{y}} \tag{26}$$

where the posterior distribution on the important parameter space $\boldsymbol{\xi}_r$ conditional on the measurement data $\bar{\mathbf{y}}$ is obtained by marginalizing out $q_{\boldsymbol{\xi}_r}$ and the observation operator parameter $\alpha_{\bar{y}}$.

3. Function space view of conditional expectation on reduced parameter space

In this section, we take a step back and discuss the theoretical aspects of the Bayesian inference based robust optimization from the functional analysis point of view using the notion of conditional expectation. The modern concept of conditional probability is explained in light of conditional expectation following Kolmogorov [40]. Given the probability space triplet $(\Theta, \mathcal{F}, \mathbb{P})$ associated with the input stochastic space and a parameter set $q \in \mathcal{Q} \otimes L^2(\Theta, \mathcal{F}, \mathbb{P}) := \mathcal{Q}$ (as introduced in Sec. 2.1), we assume there exists a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$.

The parameter q when described as a random field over the domain $q : \mathcal{D} \times \Theta \to \mathcal{Q}$, is assumed to be fairly smooth with bounded covariance, where $\mathcal{Q}$ is the image of q. Here we assume $\mathcal{Q}$ to be a Hilbert space with an inner product structure defined as

$$\langle\langle q_1, q_2 \rangle\rangle_{\mathcal{Q}} = \mathbb{E}\left[\langle q_1(\theta), q_2(\theta) \rangle_{\mathcal{Q}}\right] \quad \forall\ q_1, q_2 \in \mathcal{Q} \tag{27}$$

Remark 1. Given that $(\Theta, \mathcal{F}, \mathbb{P})$ is a measure space with $\mathcal{G} \subset \mathcal{F}$ and $\mathcal{G}$ contains all $S \in \mathcal{F}$ with $\mathbb{P}(S) = 0$, then $L^p(\Theta, \mathcal{G}, \mathbb{P}) \subset L^p(\Theta, \mathcal{F}, \mathbb{P}),\ \forall p \geq 1$ [40]. Thus, for the subsequent discussions of this article we can specifically write $L^2(\Theta, \mathcal{G}, \mathbb{P}) \subset L^2(\Theta, \mathcal{F}, \mathbb{P})$

Then the conditional expectation $\mathbb{E}\left[q \,\middle|\, \mathcal{G}\right]$ is obtained as the orthogonal projection of q on the subspace $\mathcal{Q} \otimes L^2(\Theta, \mathcal{G}, \mathbb{P})$. Mathematically,

$$\mathbb{E}\left[q \,\middle|\, \mathcal{G}\right] = q_{\mathcal{G}} = \arg\inf_{\tilde{q} \in \mathcal{Q}_{\mathcal{G}}} \|q - \tilde{q}\|_{\mathcal{Q}_{\mathcal{G}}} \quad \text{where} \quad \mathcal{Q}_{\mathcal{G}} := \mathcal{Q} \otimes L^2(\Theta, \mathcal{G}, \mathbb{P}) \tag{28}$$

where $\|\cdot\|_{\mathcal{Q}_{\mathcal{G}}}$ is an appropriately defined norm in $\mathcal{Q}_{\mathcal{G}}$ with an inner product structure defined in Eqn. (27). If we have a function $f : \Theta \to \mathbb{R}$ such that the sub-sigma algebra generated by f is equivalent to $\mathcal{G}$ i.e. $\mathcal{G} \equiv \sigma(f)$, then we can write following the Doob-Dynkin lemma [40] that all elements of $\tilde{q} \in \mathcal{Q}_{\mathcal{G}}$ can be expressed as $\tilde{q} = \phi \circ f$ where ϕ are Lebesgue measurable maps from f to $\mathcal{Q}_{\mathcal{G}}$, i.e.

$$\tilde{q} = \phi \circ f(q) \qquad \text{where} \quad \phi : f \to \mathcal{Q}_{\mathcal{G}},\ \forall \tilde{q} \in \mathcal{Q}_{\mathcal{G}} \tag{29}$$

Moreover, since $q_{\mathcal{G}} = \mathbb{E}\left[q \,\middle|\, \sigma(f)\right]$ is an orthogonal projection of q on $\mathcal{Q}_{\mathcal{G}}$, we can write

$$\langle\langle q - q_{\mathcal{G}}, \tilde{q} \rangle\rangle_{\mathcal{Q}} = 0 \quad \forall \tilde{q} \in \mathcal{Q}_{\mathcal{G}} \tag{30}$$

Assuming an observation operator $\mathcal{Y}$ as defined in Eqn. (11) which defines a σ-algebra denoted by $\sigma(\mathcal{Y})$, it follows from the previous discussion in Eqns. (28)–(30) that the conditional expectation $\mathbb{E}\left[q \,\middle|\, \sigma(\mathcal{Y})\right]$ is given by

$$\mathbb{E}\left[q \,\middle|\, \sigma(\mathcal{Y})\right] = \arg\inf_{\hat{q} \in \mathcal{Q}_{\mathcal{Y}}} \|q - \hat{q}\|_{\mathcal{Q}_{\mathcal{Y}}} \quad \text{where} \quad \mathcal{Q}_{\mathcal{Y}} := \mathcal{Q} \otimes L^2(\Theta, \sigma(\mathcal{Y}), \mathbb{P}) \tag{31}$$

$$\text{and} \quad \mathbb{E}\left[q \,\middle|\, \sigma(\mathcal{Y})\right] = \phi_{\mathcal{Y}} \circ \mathcal{Y}(q; u) \quad \text{such that} \quad \phi_{\mathcal{Y}} : \mathscr{Y} \to \mathcal{Q}_{\mathcal{Y}} \tag{32}$$

Thus $\phi_{\mathcal{Y}}$ are measurable maps from the vector space $\mathscr{Y}$ (in which the observation y exist) to $\mathcal{Q}_{\mathcal{Y}}$.

Lemma 1. *Given an input set of parameters q with an associated probability space $(\Theta, \mathcal{F}, \mathbb{P})$, and a mapping $\kappa(u(q)) : u(q) \to \kappa,\ \forall q \in \mathcal{Q}$ ($\mathcal{Q}$-valued random variable) with $\sigma(\kappa) \subset \mathcal{F}$ being the sub-sigma algebra generated by κ, find $q_a \in \bar{\mathcal{Q}}$ where $\bar{\mathcal{Q}} := L^2(\Theta, \sigma(\kappa), \mathbb{P}) \otimes \mathcal{Q}$ such that*

$$q_a = \arg \inf_{\tilde{q} \in \bar{\mathcal{Q}}} \|q - \tilde{q}\|_{\mathcal{Q} \otimes L^2(\Theta, \mathcal{F}, \mathbb{P})} \tag{33}$$

This results in q_a becoming the conditional expectation $q_a = \mathbb{E}\left[q \,\middle|\, \sigma(\kappa)\right]$ which is an orthogonal projection given by

$$\left\langle \left\langle q - \mathbb{E}\left[q \,\middle|\, \sigma(\kappa)\right], \tilde{q} \right\rangle \right\rangle = 0 \quad \forall \tilde{q} \in \bar{\mathcal{Q}} \tag{34}$$

and the assimilated parameters $q_a = \mathbb{E}\left[q \,\middle|\, \sigma(\kappa)\right]$ are $\sigma(\kappa)$-measurable functions (using the Doob-Dynkin lemma). Hence $q_a = \psi_{q_a} \circ \kappa$ where $\psi_{q_a} : \kappa \mapsto q_a$. The map $P_{q_a} : q \mapsto q_a$ is the L^2-projection defined above and is a Markov operator [40].

To reiterate, the objective of this study is to obtain the optimal values of the design parameters which would satisfy a user-specified design criterion. Prior probability density $\mathbb{P}(\theta)$ is assumed on the input parameter space θ which ensures support of $\mathbb{P}(\theta)$ covers the target design space. This is a typical requirement for Bayesian regularization such that the priors cover the full probability density support containing the posterior distribution. Lemma 1 shows that the posterior calculated using the Bayes' theorem is optimal in the L^2 sense.

3.1. Probabilistic Optimization using Bayesian inference

The objective of this section is to pose the problem of robust optimization of the engineering system in a reduced parameter space. Assuming the parameter space consists of n iid random variables $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_n\}$ in the space $L^2(\Theta, \mathcal{F}, \mathbb{P})$ with a vector of physical system parameters $q \in \mathcal{Q}$ defined such that $q : \theta \to \mathcal{Q}$. The dimension of the parameter space is reduced by ranking each parameter in their order of global sensitivity indices. We assume that based on the observable qoi for the system model, we have the ordering of the reduced parameter space $\boldsymbol{\xi}_r \subset \boldsymbol{\xi}$ as

$$\boldsymbol{\xi}_r = \left\{\xi_{j_1}, \ldots, \xi_{j_r} : \ \mathcal{S}_{j_1} \geq \mathcal{S}_{j_2} \geq \ldots \geq \mathcal{S}_{j_r} \quad \text{where} \quad j_1, \ldots, j_r \in [1, \ldots, n]\right\} \tag{35}$$

where $\mathcal{S}_{j_i}$ is the first or total order Sobol' index of parameter ξ_{j_i} as discussed in Sec. 2.2. Thus $\boldsymbol{\xi}_r$ identifies a reduced number r of input random parameters from the original parameter space $\boldsymbol{\xi}(\theta)$ based on their sensitivity indices $\mathcal{S}_{j_i}$ for the system response $y(\boldsymbol{\xi})$. This gives a reduced parameter space of dimension $r < n$ (where n is original total number of independent identically distributed parameters of the system) which consists of parameters in ascending order of their global sensitivity indices. We denote the remaining parameters in the set $\boldsymbol{\xi}$ after the important parameters $\boldsymbol{\xi}_r$ as $\bar{\boldsymbol{\xi}}$.

Lemma 2. *A probability space $(\Theta, \mathcal{F}, \mathbb{P})$ associated with a set of parameters $\boldsymbol{\xi} \in \mathbb{R}^n$ can be partitioned into disjoint sets $\{A_1, A_2, \ldots\}$ based on the separation of important $\boldsymbol{\xi}_r$ and non-important $\bar{\boldsymbol{\xi}}$ parameters as $\left\{A_i : \boldsymbol{\xi}_r(\theta_i) \times \sigma_{\min}(\bar{\boldsymbol{\xi}})\right\}$, where $\sigma_{\min}(\bar{\boldsymbol{\xi}})$ is the smallest possible sub-sigma algebra of the non-important parameters $\bar{\boldsymbol{\xi}}$.*

In other words, we are interested in a partition of the sample space Θ, as $\{A_1, A_2, \ldots\}$ such that each A_i is associated with a specific value of the important variable $\boldsymbol{\xi}_r$ and all admissible values of the non-important parameters $\bar{\boldsymbol{\xi}}$. The set of all admissible values of $\bar{\boldsymbol{\xi}}$ is contained in the smallest sigma algebra $\sigma_{\min}(\bar{\boldsymbol{\xi}})$ which consists of $\left\{\Theta(\bar{\boldsymbol{\xi}}), \emptyset\right\}$ where $\Theta(\bar{\boldsymbol{\xi}})$ is the set of all admissible values of the non-important parameters $\bar{\boldsymbol{\xi}}$.

Lemma 3. *The partitioned disjoint sets $\{A_1, A_2, \ldots\}$ (as defined in Lemma 2) of the full-sample space Θ associated with the probability space $(\Theta, \mathcal{F}, \mathbb{P})$ defines a sub-σ algebra $\mathcal{G}_{\boldsymbol{\xi}_r} \subset \mathcal{F}$.*

It is easily seen that $\mathcal{G}(\boldsymbol{\xi}_r)$ satisfies all the properties of a sub-sigma algebra, since it is *closed under complementation* and *closed under countable unions*. This property will be frequently referred to in the subsequent discussions in this paper. It forms the basis of identifying the optimal posterior predictive distributions of the important parameter set which satisfies the target distribution of the qoi under the condition that the non-important (or uncontrollable parameters) follow their prior distributions.

We assume a set of system parameters $q : \Theta \to \mathcal{Q}$ which depend on the fundamental set of parameters $\boldsymbol{\xi}$, which defines a sub-σ-algebra $\sigma(q)$ defined as

$$\sigma(q) = \left\{\theta_i : q^{-1}(S), \quad \forall S \in \mathcal{Q} \quad \text{and} \quad \forall \theta_i \in \Theta\right\} \tag{36}$$

We define another functional mapping as the conditional expectation $q_{\boldsymbol{\xi}_r} = \mathbb{E}\left[q \,\middle|\, \mathcal{G}_{\boldsymbol{\xi}_r}\right]$ where $q_{\boldsymbol{\xi}_r} \in \mathcal{Q} \otimes L^2\left(\Theta, \mathcal{G}_{\boldsymbol{\xi}_r}, \mathbb{P}\right)$ and $q_{\boldsymbol{\xi}_r}$ is a $\mathcal{G}_{\boldsymbol{\xi}_r}$-measurable function. This implies

$$\int_{G_i} \mathbb{E}\left[q \,\middle|\, \mathcal{G}_{\boldsymbol{\xi}_r}\right] d\mathbb{P} = \int_{G_i} q \, d\mathbb{P} \qquad \forall\, G_i \in \mathcal{G}_{\boldsymbol{\xi}_r} \tag{37}$$

Considering Eqns. (36) and (37), we take $\mathcal{G}_{\boldsymbol{\xi}_r} \subset \sigma(q)$ since the set $\sigma(q)$ is dense in $\mathcal{F}$.

Next, we take an observation operator $\mathcal{Y}$, which defines a sub-σ-algebra $\sigma(\mathcal{Y})$ as discussed in Sec. 2.3. Our objective then is to obtain a functional mapping between the elements of the sets contained in the sub-σ-algebras $\sigma(\mathcal{Y})$ and $\mathcal{G}_{\boldsymbol{\xi}_r}$. This scenario is represented schematically in Fig. (3) where the dotted lines signify the functional relationship to be derived based on conditional expectation.

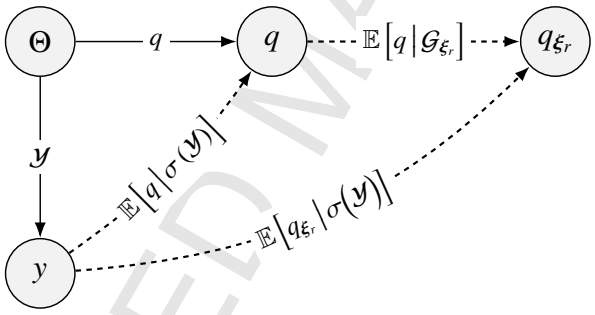


Figure (3): Schematic diagram of the Bayesian model fitting on a reduced important parameter space.

Theorem 1. *There is a basic probability space defined by the triplet $(\Theta, \mathcal{F}, \mathbb{P})$ consisting of a set of n independent identically distributed random variables $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_n\}$ which models a set of system parameters $q \in \mathcal{Q} \otimes L^2(\Theta, \sigma(q), \mathbb{P})$. The sub-$\sigma$-algebras obtained by the observation operator $\mathcal{Y}$ and the important parameter based partitioning of the sample space are defined as $\sigma(\mathcal{Y})$ and $\mathcal{G}_{\boldsymbol{\xi}_r}$ respectively. The conditional expectation $\mathbb{E}\left[\mathbb{E}\left[q \,\middle|\, \mathcal{G}_{\boldsymbol{\xi}_r}\right] \,\middle|\, \sigma(\mathcal{Y})\right]$ exists and when restricted to $L^2(\Theta, \mathcal{F}, \mathbb{P})$ is a projection onto a subspace $L^2(\Theta, \sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}}, \mathbb{P})$ where $\sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}} = \sigma(\mathcal{Y}) \cap \mathcal{G}_{\boldsymbol{\xi}_r}$*

Proof. It has already been discussed in Lemma 3 that the important parameter space generates a sub-σ-algebra denoted by $\mathcal{G}_{\boldsymbol{\xi}_r}$. Doob-Dynkin's lemma ensures that $\mathbb{E}\left[q \,\middle|\, \mathcal{G}_{\boldsymbol{\xi}_r}\right] = f(\boldsymbol{\xi}_r)$ is a $\mathcal{G}_{\boldsymbol{\xi}_r}$-measurable function of the important parameters $\boldsymbol{\xi}_r$. Similarly $\mathbb{E}\left[q \,\middle|\, \sigma(\mathcal{Y})\right]$ is $\sigma(\mathcal{Y})$-measurable and is a function of $\mathcal{Y}$ i.e. $\mathbb{E}\left[q \,\middle|\, \sigma(\mathcal{Y})\right] = \phi_(\mathcal{Y}) \circ \mathcal{Y}$. We also assume that $\sigma(\mathcal{Y}) \cap \mathcal{G}_{\boldsymbol{\xi}_r} \neq \emptyset$. Following Eqn. (30), $\mathbb{E}\left[q \,\middle|\, \mathcal{G}_{\boldsymbol{\xi}_r}\right]$ and $\mathbb{E}\left[q \,\middle|\, \sigma(\mathcal{Y})\right]$ are orthogonal projections of q on the measurable subspaces generated by $\mathcal{G}_{\boldsymbol{\xi}_r}$ and $\sigma(\mathcal{Y})$ and we denote these projections as $\mathcal{P}_{\mathcal{G}_\xi}$ and $\mathcal{P}_{\mathcal{Y}}$ respectively. Thus, the projection $\mathcal{P}_{\mathcal{Y}\mathcal{G}_\xi}$ onto $\sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}} = \mathcal{G}_{\boldsymbol{\xi}_r} \cap \sigma(\mathcal{Y})$ is obtained as $\mathcal{P}_{\mathcal{Y}\mathcal{G}_\xi} = \mathcal{P}_{\mathcal{Y}}\mathcal{P}_{\mathcal{G}_\xi} = \mathcal{P}_{\mathcal{G}_\xi}\mathcal{P}_{\mathcal{Y}}$, as per [40], and this implies

$$\mathbb{E}\left[q \,\middle|\, \mathcal{G}_{\boldsymbol{\xi}_r} \cap \sigma(\mathcal{Y})\right] = \mathbb{E}\left[\mathbb{E}\left[q \,\middle|\, \mathcal{G}_{\boldsymbol{\xi}_r}\right] \,\middle|\, \sigma(\mathcal{Y})\right] = \mathbb{E}\left[\mathbb{E}\left[q \,\middle|\, \sigma(\mathcal{Y})\right] \,\middle|\, \mathcal{G}_{\boldsymbol{\xi}_r}\right] \tag{38}$$

Thus the conditional expectation exists and is a projection on $L^2(\Theta, \sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}}, \mathbb{P})$. Again, we can write the double conditional expectation as

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[q\,\middle|\,\mathcal{G}_{\boldsymbol{\xi}_r}\right]\,\middle|\,\sigma(\mathcal{Y})\right] &= \mathbb{E}\left[q\,\middle|\,\sigma\mathcal{Y}\right] - \mathbb{E}\left[\left(q - \mathbb{E}\left[q\,\middle|\,\mathcal{G}_{\boldsymbol{\xi}_r}\right]\right)\,\middle|\,\sigma(\mathcal{Y})\right] \\ &= \phi_{\mathcal{Y}} \circ \mathcal{Y} - \mathbb{E}\left[\left(q - \mathbb{E}\left[q\,\middle|\,\mathcal{G}_{\boldsymbol{\xi}_r}\right]\right)\,\middle|\,\left(\sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}} + \sigma^{\perp}_{\mathcal{Y}\mathcal{G}_{\xi_r}}\right)\right] \end{aligned} \tag{39}$$

where $\mathbb{E}\left[q\,\middle|\,\sigma\mathcal{Y}\right] = \phi_{\mathcal{Y}} \circ (\mathcal{Y})$ from Eqn. (30) where $\phi_{\mathcal{Y}} \in L_0(\mathcal{Y}, \mathcal{Q})$ and we have decomposed the sub-σ algebra $\sigma(\mathcal{Y})$ into two components as $\sigma(\mathcal{Y}) = \sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}} + \sigma^{\perp}_{\mathcal{Y}\mathcal{G}_{\xi_r}}$, where $\sigma^{\perp}_{\mathcal{Y}\mathcal{G}_{\xi_r}}$ is orthogonal to $\sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}}$. We also denote $q^{\perp}_{\mathcal{G}_{\xi}} = q - \mathbb{E}\left[q\,\middle|\,\mathcal{G}_{\boldsymbol{\xi}_r}\right]$ to be orthogonal to all vectors in $\mathcal{Q} \otimes L^2(\Theta, \mathcal{G}_{\boldsymbol{\xi}_r}, \mathbb{P})$. Following from the previous equation, we can write

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[q\,\middle|\,\mathcal{G}_{\boldsymbol{\xi}}\right]\,\middle|\,\sigma(\mathcal{Y})\right] &= \phi_{\mathcal{Y}} \circ \mathcal{Y} - \mathbb{E}\left[q^{\perp}_{\mathcal{G}_{\xi}}\,\middle|\,\left(\sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}} + \sigma^{\perp}_{\mathcal{Y}\mathcal{G}_{\xi_r}}\right)\right] \\ \text{or,}\quad \mathbb{E}\left[\mathbb{E}\left[q\,\middle|\,\mathcal{G}_{\boldsymbol{\xi}}\right]\,\middle|\,\sigma(\mathcal{Y})\right] &= \phi_{\mathcal{Y}} \circ \mathcal{Y} - \mathbb{E}\left[q^{\perp}_{\mathcal{G}_{\xi}}\,\middle|\,\sigma^{\perp}_{\mathcal{Y}\mathcal{G}_{\xi_r}}\right] \end{aligned} \tag{40}$$

Here we note that $q^{\perp}_{\mathcal{G}_{\xi}}$ being orthogonal to all elements in the sub-σ algebra $\mathcal{G}_{\boldsymbol{\xi}_r}$, would also be orthogonal to a subset of it $\sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}} \subset \mathcal{G}_{\boldsymbol{\xi}_r}$, hence $\mathbb{E}\left[q^{\perp}_{\mathcal{G}_{\xi}}\,\middle|\,\sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}}\right] = 0$. Equation (40) signifies that the LHS is equal to the projection of q on the sub-σ algebra generated by the observation operator $\mathcal{Y}$, i.e $\phi_{\mathcal{Y}} \circ \mathcal{Y}$, minus the projection of the non-important parameter value $q^{\perp}_{\mathcal{G}_{\xi_r}}$ on portion of the observation space orthogonal to the intersection between $\sigma(\mathcal{Y})$ and $\mathcal{G}_{\boldsymbol{\xi}_r}$. □

In the special case where $\sigma(\mathcal{Y}) \cap \mathcal{G}_{\boldsymbol{\xi}_r} = \sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}} = \emptyset$, we expect $\mathbb{E}\left[\mathbb{E}\left[q\,\middle|\,\mathcal{G}_{\boldsymbol{\xi}}\right]\,\middle|\,\sigma(\mathcal{Y})\right] = 0$. This can be proved as follows. Firstly, if $\sigma_{\mathcal{Y}\mathcal{G}_{\xi_r}} = \emptyset$ then $\sigma^{\perp}_{\mathcal{Y}\mathcal{G}_{\xi_r}} \equiv \sigma(\mathcal{Y})$. Thus, when $\sigma(\mathcal{Y}) \cap \mathcal{G}_{\boldsymbol{\xi}_r} = \emptyset$

$$\mathbb{E}\left[\mathbb{E}\left[q\,\middle|\,\mathcal{G}_{\boldsymbol{\xi}_r}\right]\,\middle|\,\sigma(\mathcal{Y})\right] = \phi_{\mathcal{Y}} \circ \mathcal{Y} - \mathbb{E}\left[q^{\perp}_{\mathcal{G}_{\xi_r}}\,\middle|\,\sigma(\mathcal{Y})\right] \tag{41}$$

We also note that since $\mathbb{E}\left[q^{\perp}_{\mathcal{G}_{\xi_r}}\,\middle|\,\sigma(\mathcal{Y})\right]$ is a projection of $q^{\perp}_{\mathcal{G}_{\xi_r}}$ on $\mathcal{Q} \otimes L^2(\Theta, \sigma(\mathcal{Y}), \mathbb{P})$, it is expressed as $\mathbb{E}\left[q^{\perp}_{\mathcal{G}_{\xi_r}}\,\middle|\,\sigma(\mathcal{Y})\right] = \psi_{\mathcal{Y}} \circ \mathcal{Y}$ where $\psi_{\mathcal{Y}}$ are measurable maps in $L_0(\mathcal{Y}, \mathcal{Q})$. Additionally, under the condition of $\sigma(\mathcal{Y}) \cap \mathcal{G}_{\boldsymbol{\xi}_r} = \emptyset$, the projection of $q^{\perp}_{\mathcal{G}_{\xi_r}}$ on $\mathcal{Q} \otimes L^2(\Theta, \sigma(\mathcal{Y}), \mathbb{P})$ is equivalent to the projection of q on the same. This results in $\psi_{\mathcal{Y}} \equiv \phi_{\mathcal{Y}}$. Thus, from Eqn. (41) we have

$$\mathbb{E}\left[\mathbb{E}\left[q\,\middle|\,\mathcal{G}_{\boldsymbol{\xi}_r}\right]\,\middle|\,\sigma(\mathcal{Y})\right] = 0 \quad \text{when} \quad \sigma(\mathcal{Y}) \cap \mathcal{G}_{\boldsymbol{\xi}_r} = \emptyset \tag{42}$$

4. Application to engineering problems

The robust optimization algorithm described in the previous sections are summarized here which would be helpful in making them amenable for implementations applied to practical engineering problems. The details of the MCMC sampling strategy has been removed to preserve clarity of the steps involved in calculating the optimum posterior on the important parameter space. Blackbox MCMC implementations can be utilized in conjunction with this algorithm and the primary quantities required for the calculation are listed in Algorithm 1.

We apply the proposed robust Bayesian optimization method to two example problems to demonstrate its performance and accuracy for real engineering problems. The first problem is a related to a structural dynamic system while the second deals with minimizing a non-linear six dimensional Hartmann function.

Algorithm 1 Robust optimization of important design variables under uncertainty

Input: Prior on n-dimensional parameter space $\boldsymbol{\xi} \in \mathbb{R}^n$, such that $\boldsymbol{\xi} \sim \mathbb{P}(\boldsymbol{\xi})$.
Input: Response surface of the output qoi, $\mathbf{y}(q(\boldsymbol{\xi})) \in \mathcal{Y}$.
Input: Target distribution on output qoi, $\bar{\mathbf{y}} \sim \mathbb{P}(\bar{\mathbf{y}})$.
Output: Posterior distribution on $\boldsymbol{\xi}_r \sim \mathbb{P}(\boldsymbol{\xi}_r \,|\, \bar{\mathbf{y}})$, where $\boldsymbol{\xi}_r \subset \boldsymbol{\xi}$, $r < n$.

1: Calculate importance measure $\mathcal{S}_i, i = 1, \ldots, n$.
2: Identify $\boldsymbol{\xi}_r = \left\{\xi_{j_1}, \ldots, \xi_{j_r} : \ \mathcal{S}_{j_1} \geq \mathcal{S}_{j_2} \geq \ldots \geq \mathcal{S}_{j_r}\right\}$, $j_1, \ldots, j_r \in [1, \ldots, n]$
3: Initialize $\boldsymbol{\xi}_r^0$.
4: % {MCMC sampling commences}
5: **for** i=1 to n_s **do**
6: $\quad \boldsymbol{\xi}_r^{(i)}$ from a proposal $\Pi(\boldsymbol{\xi}_r^i | \boldsymbol{\xi}_r^{(i-1)})$.
7: $\quad$ Likelihood $\mathcal{L}_{\bar{\mathbf{y}}|\alpha_{\bar{\mathbf{y}}}} = \mathbb{P}(\bar{\mathbf{y}} \,|\, \mathcal{Y}, q_{\boldsymbol{\xi}_r}) \mathbb{P}(\mathcal{Y}(q_{\boldsymbol{\xi}_r}) \,|\, \alpha_{\bar{\mathbf{y}}})$
8: $\quad$ Posterior probability on $\boldsymbol{\xi}_r$ as $\mathbb{P}\left(\boldsymbol{\xi}_r \,|\, q_{\boldsymbol{\xi}_r}, \bar{\mathbf{y}}, \mathcal{Y}\right) = \mathcal{L}_{\bar{\mathbf{y}}|\alpha_{\bar{\mathbf{y}}}} \mathbb{P}\left(q_{\boldsymbol{\xi}_r} \,|\, \boldsymbol{\xi}_r, \alpha_q\right) \mathbb{P}\left(\boldsymbol{\xi}_r\right)$.
9: $\quad$ Accept or reject $\boldsymbol{\xi}_r^{(i)}$ % {based on MCMC acceptance probability}
10: **end for**
11: Posterior predictive distribution $\mathbb{P}(\mathbf{y} \,|\, \bar{\mathbf{y}}, q, \mathcal{Y}) = \int_{\boldsymbol{\xi}_r} \mathbb{P}(\mathbf{y} \,|\, \bar{\mathbf{y}}, q, \mathcal{Y}) \mathbb{P}\left(\boldsymbol{\xi}_r \,|\, q_{\boldsymbol{\xi}_r}, \bar{\mathbf{y}}, \mathcal{Y}\right) \, \mathrm{d}\boldsymbol{\xi}_r$

4.1. Randomly parametrized structural dynamic system

We consider a bounded domain $\mathscr{D} \in \mathbb{R}^d$ for a structural dynamic system with piecewise Lipschitz boundary $\partial\mathscr{D}$, $d \leq 3$ is the spatial dimension and $t \in \mathbb{R}^+$ is the time variable. The linear stochastic partial differential equation (pde) governing the system with uncertainty in parameter values is given as

$$\rho(\mathbf{r}; \theta) \frac{\partial^2 u(\mathbf{r}, t; \theta)}{\partial t^2} + \mathfrak{Q}_{\mathbf{c}} \frac{\partial u(\mathbf{r}, t; \theta)}{\partial t} + div\left(\sigma_\alpha(u(\mathbf{r}, t; \theta))\right) = p(\mathbf{r}, t); \quad \mathbf{r} \in \mathscr{D}, \ t \in [0, T], \ \theta \in \Theta \tag{43}$$

with the associated Dirichlet condition $u(\mathbf{r}, t, \theta) = 0; \ \mathbf{r}$ on $\partial\mathscr{D}$. Here $\sigma_\alpha(u(\mathbf{r}, t; \theta))$ is the stress tensor with stiffness coefficient $\alpha(\mathbf{r}, \theta)$ modeled as a stationary, square integrable random field such that $\alpha : \mathbb{R}^d \times \Theta \to \mathbb{R}$. The random field exists in the probability space defined by the triplet $(\Theta, \mathscr{F}, \mathbb{P})$ as before. The operator $div(\sigma_a)$ is the self-adjoint stochastically parametrized stiffness operator and $\mathfrak{Q}_{\mathbf{c}}$ is the damping operator containing the stochastic coefficient vector $\mathbf{c}(\mathbf{r}, \theta)$. The system solution $u(\mathbf{r}, t, \theta)$ is sought under the condition of external excitation field $p(\mathbf{r}, t)$. To perform harmonic analysis, Eqn. (43) is transformed to the frequency domain as

$$-\omega^2 \rho(\mathbf{r}; \theta) \widetilde{u}(\mathbf{r}, \omega; \theta) + i\omega \mathfrak{Q}_{\mathbf{c}} \widetilde{u}(\mathbf{r}, \omega; \theta) + div\left(\sigma_\alpha(\widetilde{u}(\mathbf{r}, \omega; \theta))\right) = \widetilde{p}(\mathbf{r}, \omega); \quad \theta \in \Theta \tag{44}$$

where ω denotes the harmonic frequency, with $\widetilde{p}$ and $\widetilde{u}$ representing the corresponding complex harmonic amplitudes. $\mathbf{E}(\alpha)$ in the stress-strain relationship $\sigma_\alpha = \mathbf{E}(\alpha) : \varepsilon$ is the symmetric positive definite elasticity tensor which depends on the scalar random parameter α with ε being the strain tensor expressed as $\varepsilon = \mathbf{D}\widetilde{\mathbf{u}}$. Well established techniques of variational formulation of the displacement-based deterministic finite-element methods [43, 44] gives the following bilinear form for the elastodynamic system (dropping the ω for the sake of simplicity)

$$\mathscr{B}(\widetilde{\mathbf{v}}, \widetilde{\mathbf{u}}; \theta) = -\omega^2 \int_{\mathscr{D}} \widetilde{\mathbf{v}} \rho(\mathbf{r}; \theta) \widetilde{\mathbf{u}} \, d\mathscr{D} + i\omega \int_{\mathscr{D}} \widetilde{\mathbf{v}} \mathfrak{Q}_{\mathbf{c}} \widetilde{\mathbf{u}} \, d\mathscr{D} + \int_{\mathscr{D}} \{\mathbf{D}\widetilde{\mathbf{v}}\}^T \mathbf{E}(\alpha) \{\mathbf{D}\widetilde{\mathbf{u}}\} \, d\mathscr{D} \tag{45}$$

$$\mathscr{L}(\widetilde{\mathbf{v}}; \theta) = \int_{\mathscr{D}} \widetilde{\mathbf{v}} \widetilde{p} \, d\mathscr{D}$$

$$\text{so that,} \quad \mathscr{B}(\widetilde{\mathbf{v}}, \widetilde{\mathbf{u}}; \theta) = \mathscr{L}(\widetilde{\mathbf{v}}; \theta) \qquad \forall \ \widetilde{\mathbf{v}} \in \mathscr{X} \times \mathscr{S} \tag{46}$$

where $\mathscr{X} \times \mathscr{S}$ the tensor product space of admissible functions as introduced Eqn. (2). Equation (46) gives a set of discretized linear algebraic equations in terms of the mass, damping and stiffness matrices. These can be expressed in a compact form as

$$\mathbf{A}(\omega,\theta)\widetilde{\mathbf{u}}(\omega,\theta) = \widetilde{\mathbf{p}}(\omega,\theta); \quad \forall\ \theta \in \Theta;\ \mathbf{A} \in \mathbb{C}^{n\times n};\ \widetilde{\mathbf{u}}, \widetilde{\mathbf{p}} \in \mathbb{C}^n \tag{47}$$

where $\mathbf{A}(\omega,\theta)$ is the complex frequency dependent coefficient matrix which inherits the uncertainty of the random parameters involved in the governing pde.

The uncertainty elicited using random field description of a parameter α is generally associated with a covariance function $C_\alpha : \mathscr{D} \times \mathscr{D} \to \mathbb{R}$ defined on the open, bounded polygonal domain in $\mathscr{D}$. For second order random fields, there is a compact self-adjoint operator

$$T_\alpha v(\cdot) = \int_{\mathscr{D}} C_\alpha(\mathbf{r},\cdot)v(\mathbf{r})d\mathbf{r} \quad \forall v \in L^2(\mathscr{D}) \quad \text{s.t.} \quad T_\alpha\varphi_i = \lambda_i\varphi_i, \quad \left\langle \varphi_i, \varphi_j \right\rangle_{L^2(\mathscr{D})} = \delta_{ij} \tag{48}$$

where $\langle\cdot,\cdot\rangle_{L^2(\mathscr{D})}$ denotes the inner product in $L^2(\mathscr{D})$. The $\{(\lambda_i,\varphi_i)\}_{i=1}^{\infty}$ are a sequence of non-negative eigenpairs of the operator $T_\alpha v(\cdot)$. The truncated KL expansion of the stochastic process $\alpha(\mathbf{r},\theta)$ is thus expressed using these eigen-functions as

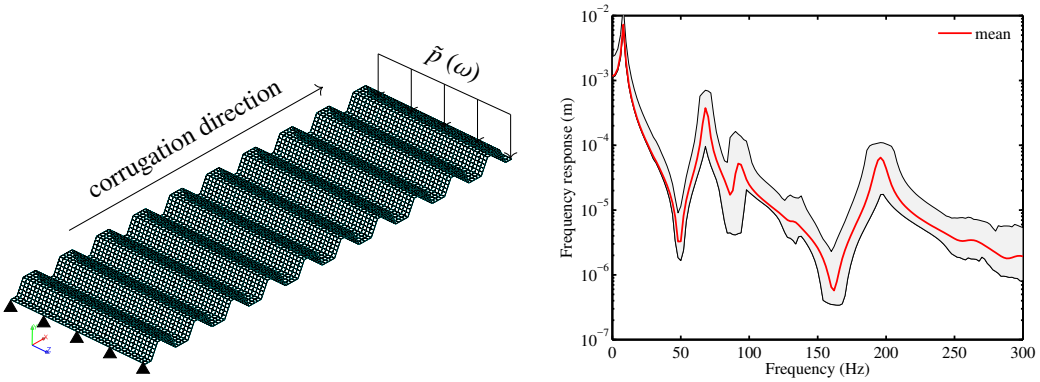
$$\alpha(\theta,\mathbf{r}) = \underbrace{\mathbb{E}\left[\alpha\right](\mathbf{r})}_{\text{mean part}} + \underbrace{\sum_{i=1}^{m} \sqrt{\lambda_i}\varphi_i(\mathbf{r})\xi_i(\theta)}_{\text{deviatoric part}} \quad \forall m \in \mathbb{N}_+ \tag{49}$$

where $\mathbb{E}\left[\alpha\right](\mathbf{r})$ is the mean function and $\{\xi_i(\theta)\}_{i=1}^{m}$ is a set of pairwise uncorrelated random variables. In the special case where α is a Gaussian stochastic process, ξ_i-s are modeled with independent standard (zero mean, unit variance) Gaussian random variables. The eigenfunctions $\varphi_i(\mathbf{r})$ can be assumed to have sufficient smoothness for smooth covariance functions, and if the eigenpairs are decaying according to at least $\sqrt{\lambda_k}\,\|\varphi_k\|_{L^\infty(\mathscr{D})} = O(\frac{1}{1+k^s})$ for some decay exponent $s > 1$, then $\|\alpha - \hat{\alpha}\|_{L^\infty(\mathscr{D})} \to 0$, as $m \to \infty$ increases [45]. For practical engineering problems, the parametric randomness is modeled with a finite set of random variables $\boldsymbol{\xi} = (\xi_1,\xi_2,\ldots,\xi_m) : \Theta \to \mathbb{R}^m$, using first few largest eigenpairs in the reduced probability space [46]. This is facilitated by the fact that the non-negative eigenvalues satisfy the relation $\sum_{i=1}^{\infty_i} \lambda_i = \int_{\mathscr{D}} Var[\alpha](\mathbf{r})d\mathbf{r}$ and decay in accordance with the aforementioned relation. The choice of m is informed by the decaying eigenvalues and it is ensured $\ln(\lambda_m/\lambda_1) << 0.01$ which implies that at least 99% of the spectrum has been accounted for in the truncated KL expansion in Eqn. (49). For arbitrary random field models, the expansion in Eqn. (49) can be written in terms of Weiner-Askey chaos using finite order polynomial basis functions following [47]. Additionally, to ensure strict positivity of stochastic processes (such as when modelling elastic modulus, material thickness), it is standard to use the lognormal transformation [2, 48].

The solution $u(\boldsymbol{\xi})$ is obtained using a finite order chaos expansion as discussed in Sec. 2.1 using Galerkin's method. The order of polynomial expansion has been chosen to ensure that the mean and standard deviation of the solution has converged to the working level of accuracy for the chosen application. Following the stochastic response surface constructed in the parameter space, the sensitivity indices are analytically obtained from the coefficients of the expansion as discussed in Sec. 2.2. The important parameter space $\boldsymbol{\xi}_r$ is identified based on the sensitivity indices on which the posterior probability distributions are estimated conditioned on the optimal distributions prescribed on the target qoi.

The structural system being studied here is a corrugated panel (as shown in Fig. 4(a)) which is fixed at the left edge and a line force acts on the opposite edge. This panel is of particular interest in aerospace applications due to its highly anisotropic behavior of high compliance in the corrugation direction and high stiffness in the transverse direction. The uncertainties in the material and geometric properties of the corrugated skin have been described probabilistically with random fields and variables.

The idealized geometry of the corrugated panel is described with the geometrical properties of each repetitive corrugation unit (such as width, height and the angle of corrugation). But it is impossible (under the cost constraints of industrial production) to have the exact replication of the prescribed geometry during manufacturing. As a result uncertainty always exists in the geemetrical shape of the corrugated panel.



(a) Corruagted panel

(b) Mean response and the $\pm 2\sigma$ envelope

Figure (4): The corrugated panel used as an example of the structural dynamic system being studied here and the frequency response function (mean and standard deviation) of the system under the action of a point load.

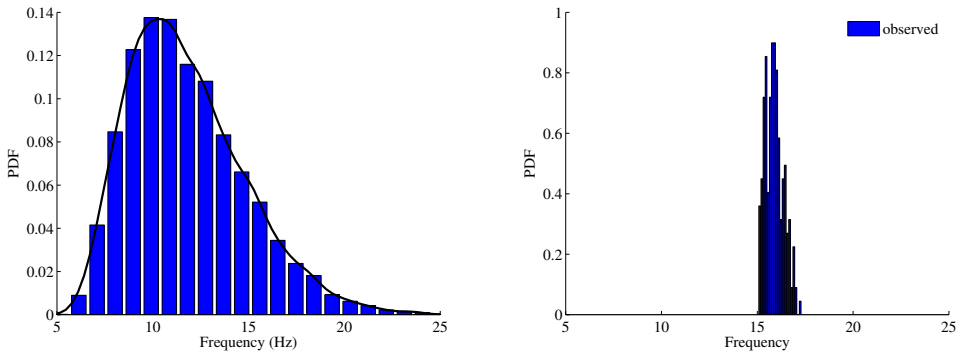
The uncertainty in the geometric shape or random surface roughness is quantified here with a random field description characterized by a correlation length. Additionally, the thickness and Young's modulus has also been considered uncertain. There are a total of 12 random variables which are used to model the stochastic properties as given below.

- The topology of the corrugated skin is assumed to be a random field with an exponential covariance function $\left(C_r(\mathbf{r}_1, \mathbf{r}_2) = \exp\left\{-\frac{\|\mathbf{r}_1 - \mathbf{r}_2\|_{L^2}}{L_c}\right\}\right)$. The random field has been approximated with 10 uniform random variables which models the perturbation of the (x, y) coordinates of the skin.
- The 11-th random variable models the thickness of the corrugated panel as a uniform random variable around the nominal thickness value.
- The 12-th random variable models the Young's modulus as a uniform random variable around the nominal value.

Each random parameter is modeled with a scaled version of a standard uniform random variable, such that a random parameter a is given as $a = a_0(1 + \delta\xi)$ where a_0 is the baseline parameter value, δ models the variability and ξ is a uniform random variable (between $[-1, 1]$). The value of δ has been chosen for each of these random parameters has been chosen to be equal for all cases, $\delta = 0.2$ which represents a 40% variability of the parameters about their baseline value. The correlation length L_c has been taken to be half of the overall length (along x-axis) of the corrugated panel. Hence the input parameter space is 12 dimensional which models the random topology, thickness and elastic constant of the corrugated panel.

The stochastic solution of this system, as shown in Fig. 4(b), has been obtained with 4-th order Legendre polynomials which are orthogonal with respect to the joint uniform distribution on the input random variables. This results in 1820 Legendre basis functions. This has been tested to approximate the response surface with sufficient accuracy in context of the present computations. The figure shows

the mean curve along with the $\pm 2\sigma$ envelope (σ being the standard deviation) which signifies that 95% of the possible FRF realizations lie within this envelope given the prior uniform distribution on the 12 input random variables.



(a) Prior distribution of 1st resonant frequency

(b) Target distribution, $\Pi(\mathcal{D})$

Figure (5): The distribution of the first resonant frequency given the prior distributions on the multidimensional stochastic parameter space. The figure on the right shows the target distribution $\Pi(\mathcal{D})$ from which the observation data $\mathcal{D}$ is sampled.

This leads to a distribution of the first resonant frequency of the corrugated panel as shown in Fig. 5(a). The objective of the problem is to optimize the first resonant frequency so that it follows a target distribution function shown in Fig. 5(b). The objective of the design optimization in this study is to isolate the structural modes of vibration from the aerodynamic modes. As a result, the objective is to have the structural modes pushed near (and beyond) the 15 Hz range. This target distribution of the first resonant frequency is specified by a beta distribution whose pdf is specified as $x^{a-1}(1-x)^{b-1}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ with $a = 2.0$, $b = 5.0$, Γ being the gamma function and the pdf is defined over the support of $[15, 18]$ Hz. We have synthetically generated $1,000$ samples from this target distribution conditioned on which we seek the joint posterior distribution in the parameter space of interest.

The sensitivity of the first resonant frequency to the individual random variables have been estimated based on the discussion presented in Sec. 2.2. Here table 1 gives the first order sensitivity values of the

Table 1: Sobol's sensitivity indices for the first two eigenvalues of the randomly parametrized (12 parameters) corrugated skin. The first order Sobol' indices of the four most sensitive parameters to the first and second resonant frequencies are shown along with the respective 95% confidence intervals (CI).

R.V.	1st eigenvalue (λ_1)		2nd eigenvalue (λ_2)	
	1st order	95% CI	1st order	95% CI
ξ_1	0.552858	0.054858	0.572822	0.061312
ξ_2	0.069691	0.009317	0.081345	0.009861
ξ_{11}	0.051539	0.008933	0.051970	0.009138
ξ_{12}	0.203692	0.035671	0.206073	0.036156

first and second resonant frequencies of the stochastic corrugated panel to the input random variables $\{\xi_1, \xi_2, \xi_{11}, \xi_{12}\}$. These four parameters have the highest values of first order Sobol' indices with an acceptable confidence interval (less than 10% of the sensitivity value). Hence the first robust optimization problem has considered ξ_1 and ξ_{12} as the important parameters whose posterior distributions are inferred conditional on the samples from the target distribution (as shown in Fig. 5(b)).

The MCMC method has been used to sample from the joint posterior distribution of ξ_1 and ξ_{12} using the Metropolis-Hastings algorithm. The prior probability support for all input parameters has

been standardized over $[-1, 1]$. The likelihood function depends on the choice of the reduced important parameter space as has been discussed in Sec. 2.5. In this case, the distribution of the first resonant frequency at each point in the important parameters space due to the non-important variables (i.e. all variables except for ξ_1 and ξ_{12}), governs the likelihood function. The optimal joint posterior distribution along with the marginal distributions are shown in Fig. (6). The best fit regression line through the cluster of points through the samples of the joint posterior distribution is shown in red. The joint posterior of

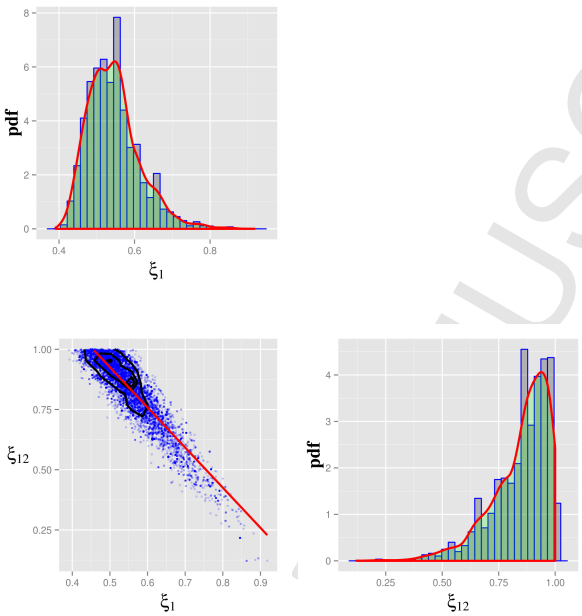


Figure (6): Posterior joint probability density functions of the two random variables ξ_1 and ξ_{12} obtained from the samples generated using the MCMC algorithm.

these two parameters shows a strong correlation, the correlation coefficient has been calculated to be 0.2502. This is expected since the variability of the posterior predictive distribution of the first resonant frequency is primarily due to variation of the unimportant parameters (the remaining 10 parameters out of the total of 12 input except for ξ_1 and ξ_{12}, which follow the original prior uniform distribution). The derived posterior distribution on ξ_1 and ξ_{12}, in order to fit the target distribution closely, would have small variability and a strong correlation, as observed in Fig. (6).

Figure (7) shows the prior, target (observation) and *posterior* (or optimized) distribution of the first resonant frequency of the corrugated panel. The posterior distribution has been obtained from the joint posterior distribution on the important parameters ξ_1 and ξ_{12} while the remaining parameters follow their uniform prior distribution. The figure indicates that the mean of the posterior distribution of the first resonant frequency has moved closer to that of the target distribution. The mean of this distribution is obtained approximately at 16 Hz which is in the desired range of values with a standard deviation of around 1.7 Hz. This shows that the posterior Bayesian model calibration framework has produced the posterior samples distributions based on the observations such that it successfully drove the natural frequency of the stochastic system in the desired range of values.

Next the important parameter space has been expanded to include the four parameters $\{\xi_1, \xi_2, \xi_{11}, \xi_{11}\}$ which are the parameters with the highest Sobol' indices (as indicated in table 1). The joint posterior distribution on the parameters (obtained using the MCMC method) and the distribution of the first resonant frequencies are shown in Figs. (8) and (9) respectively. It can be seen from Fig. (9) that the four parameter optimization produces a better approximation of the target pdf compared to the two parameter case,

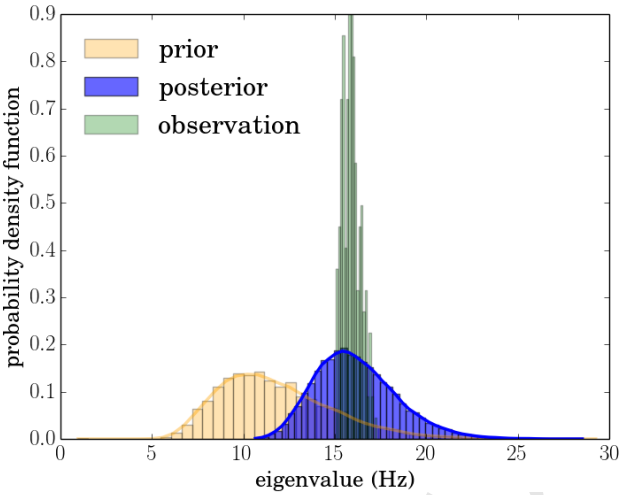


Figure (7): The prior, observed and posterior distributions of the first resonant frequency of the randomly parametrized corrugated skin.

especially in the form of reduced variance around the optimum mean. This is expected since in this case we are seeking the optimal solution in a higher dimensional (four) parameter space and the uncertainty in the optimized result is in large part due to the eight remaining input parameters following their prior distribution (compared to ten for the two parameter case). It is also to be noted that the joint posterior on the important parameters in Figs. (6) and (8) indicates the robustness of the solution by specifying the allowable range of variation (using the probability distributions) around the optimal point values of the input parameters.

Thus as the important parameter space is extended to include more parameters, the target pdf will be approximated more closely compared to the case when only a few parameters are included. This leads to the criterion where the input parameter space can be adaptively increased based on distance metrics between the acheived posterior distribution on the qoi vis-a-vis the target distribution on these qunatities (such as the distance between summary statistical indicators as mean or variance, or the Kullback-Leibler (KL) divergence estimates between the full distributions).

4.2. Test: 6-dimensional Hartmann function

The six dimensional Hartmann function is used here as another example to demonstrate the applicability of the proposed optimal probabilistic design framework. The 6D Hartmann function in variables $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_6\}$ is given as

$$f(\boldsymbol{\xi}) = -\sum_{i=1}^{4} \alpha_i \exp\left\{-\sum_{j=1}^{6} A_{ij}\left(\xi_j - P_{ij}\right)^2\right\} \quad \text{where} \quad \boldsymbol{\alpha} = [1.09, 1.2, 3.0, 3.2]\,; \quad \xi_i \in [0,1]\ \forall i \tag{50}$$

$$\mathbf{A} = \begin{bmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{bmatrix}; \quad \mathbf{P} = 10^{-4} \times \begin{bmatrix} 1312 & 1696 & 5569 & 124 & 8283 & 5886 \\ 2329 & 4135 & 8307 & 3736 & 1004 & 9991 \\ 2348 & 1451 & 3522 & 2883 & 3047 & 6650 \\ 4047 & 8828 & 8732 & 5743 & 1091 & 381 \end{bmatrix};$$

The Hartmann function is a highly non-linear function of the variables $\boldsymbol{\xi}$ and given that a prior uniform distribution in $[0, 1]$ is imposed on these variables the sensitivities of $f(\boldsymbol{\xi})$ on the individual input variables are shown in table 2. The total order Sobol' indices are shown here along with the associated 95%

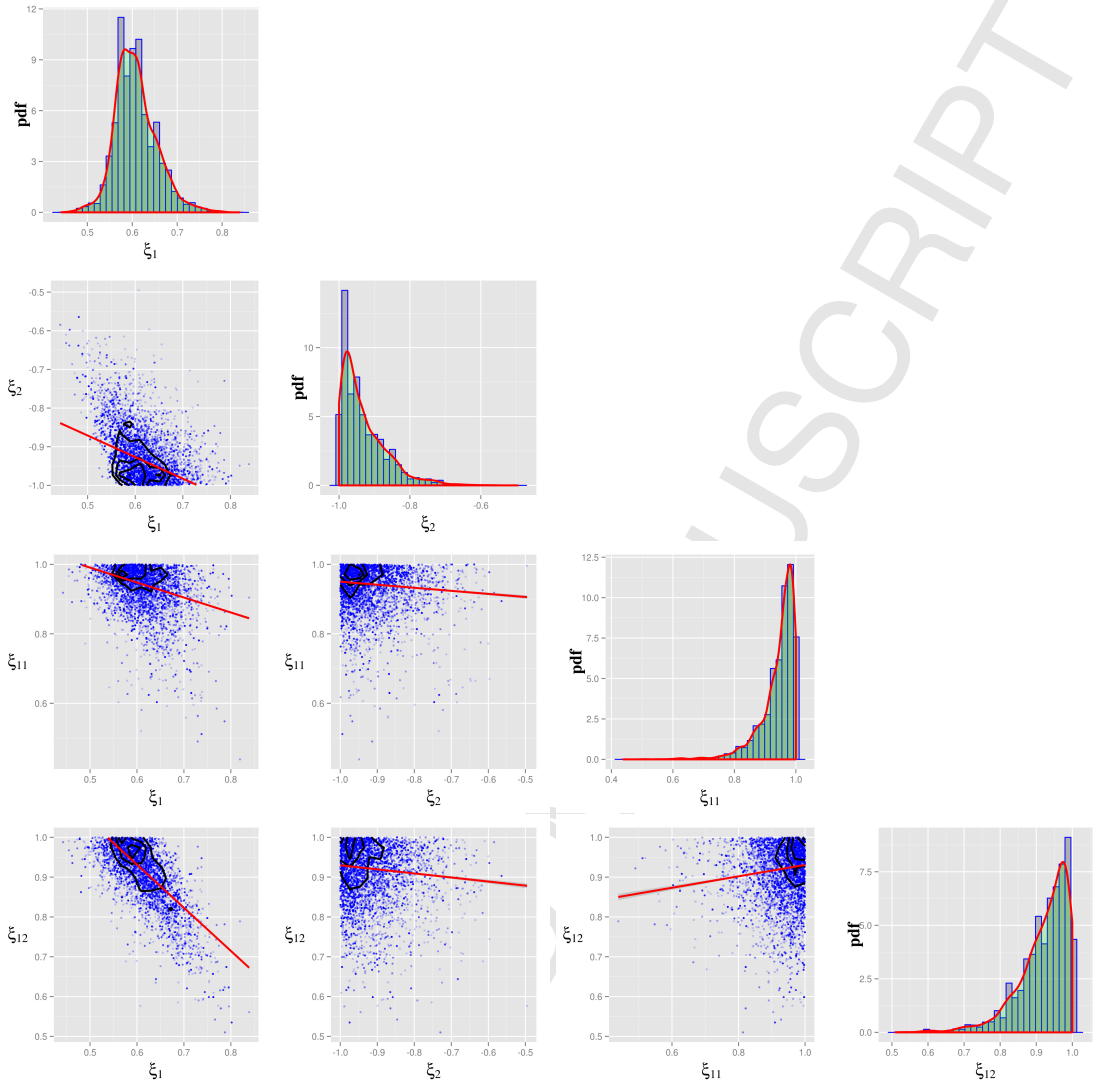


Figure (8): Posterior joint probability density functions of the four random variables $\{\xi_1, \xi_2, \xi_{11}, \xi_{11}\}$ obtained from the samples generated using the MCMC method.

Table 2: Sobol's total order sensitivity indices for the 6-dimensional Hartmann function over the domain $[0, 1]$ along with the associated confidence intervals.

Variable	Total order	
	Sobol' indices	95% C.I.
ξ_1	0.342296	0.012293
ξ_2	0.398686	0.014681
ξ_3	0.051675	0.002197
ξ_4	0.381338	0.011978
ξ_5	0.297879	0.010183
ξ_6	0.477963	0.015618

confidence interval to capture the effect of individual random variables on the function value. Due to the highly non-linear nature of the Hartmann function there is little contribution of the individual variables

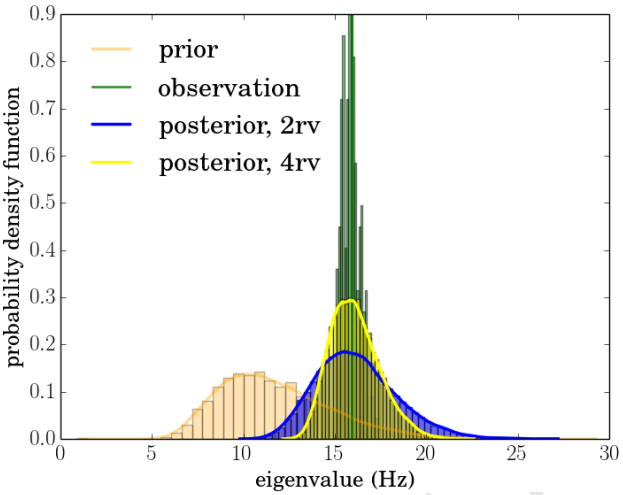


Figure (9): Comparison of the prior, posterior and observed values for MCMC model fit with 4 and 2 random variables (rv).

to the function variance, rather the interaction terms are much more significant. Hence the total order Sobol' indices have been considered to rank the random variables according to their order of importance. Table 2 shows that parameter ξ_3 and ξ_5 have the lowest sensitivity index (in ascending order) and hence the important parameter space is constructed by successively eliminating them.

The global minimum of the 6D Harmann function exists when the input variables are bound in $[0, 1]$ and following from Eqn. (50) is given as,

$$f_{\text{min}} = -3.322 \quad \text{with} \quad \boldsymbol{\xi}^* = \underset{\xi_i \in [0,1], \forall i}{\arg\ \inf} \left| f(\boldsymbol{\xi}) \right| \quad \text{where} \quad \boldsymbol{\xi}^* = \{0.202, 0.150, 0.477, 0.275, 0.312, 0.657\} \tag{51}$$

Our objective is to specify a target distribution close to the global minimum f_{min} as has been shown in Fig. (10) where the histogram in green shows the target distribution with its median at 3.0 with samples from a beta distribution. The prior distribution of the Hartmann function is obtained by sampling $f(\boldsymbol{\xi})$ for uniform priors on $\boldsymbol{\xi}$ within the $[0, 1]$ interval. Hence the objective here is to obtain the optimal distribution on the important parameter space which brings the prior (yellow) distribution on $f(\boldsymbol{\xi})$ closer to the target distribution (specified by the green histogram).

We use five parameters with highest total order Sobol' indices (from table 2) as our design variables while ξ_3 (with the lowest sensitivity) is treated as a stochastic variable which follows their prior distribution before and after optimization. The joint posterior distribution on the five parameters are shown in Fig. (11) which has been obtained by using the MCMC method to sample from the posterior distribution. The figure shows pairwise joint posterior distributions of the five parameters along with their individual marginalized distributions. It can be observed that the posterior probability support for each of these variables are quite narrow (compared to the $[0, 1]$ interval), which implies a small variance, and this support covers the input variable $\boldsymbol{\xi}$ values which give the global minimum of $f(\boldsymbol{\xi})$ as discussed in Eqn. (51).

The probability distribution of the Hartmann function given the optimal joint posterior distributions on the important parameters and the prior distribution on the non-important parameters is shown in Fig. (12). The calculation of the posterior predictive distribution has been performed with different dimensions (4 and 5) of the important parameter space (based on the highest sensitivity indices). It is seen that the closeness of the posterior predictive distribution to the target distribution in Fig. (12) is more for the 5 parameter case compared to the one with 4. Once again, this is similar to the observation

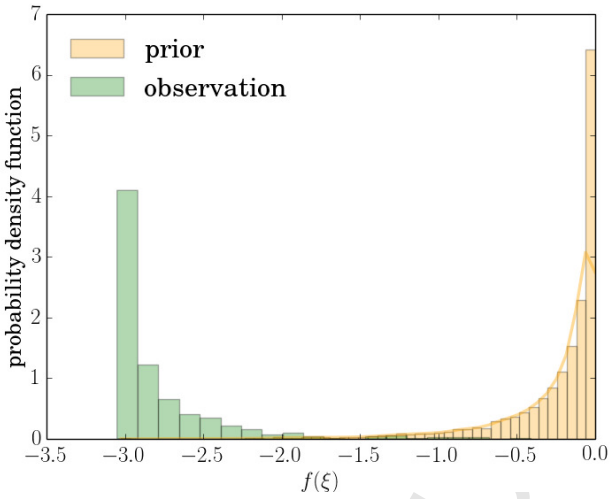


Figure (10): Prior distribution of the six-dimensional Hartmann function with uniform priors on the six variables $\{\xi_1, \ldots, \xi_6\}$ and samples (observation) from the target distribution around the global minimum of the Hartmann function.

in Fig. (9) where increasing the dimension of the design space results in improved capacity of meeting the target distribution. A closer look at Fig. (12) (and comparing with Fig. (9)) indicates that the target distribution has high skewness (asymmetry) and is far from the prior (with respect to distance measures between probability distributions, for e.g. KL divergence). The better fit of the posterior predictive distribution to the target distribution with 5 parameters highlights that the proposed method performs well even when the distance between the target probability distribution (in green in Fig. (12)) and the prior is large in a statistical sense (such as entropy based metrics). This in turn implies that the proposed method performs well even when the rate of acceptance of samples for the MCMC step is quite small.

5. Conclusions and future work

The study demonstrates the applicability of a novel robust optimization approach based on Bayesian system identification to achieve ad-hoc user-prescribed target distributions on output qoi of a stochastic system. This differs from the deterministic optimization schemes in the sense that Bayesian inference is used to obtain optimal joint posterior probability distributions on a chosen important parameters space in order to get the probability distribution of the qoi of the system close to a specified target distribution. Hence the optimization criterion is specified by incorporating the uncertainty description (either by specifying acceptable margins of deviation around the optimum or by identifying a robust region of operation of the physical system) on the target function. This alleviates the problem of having to deal with optimal point estimates (as is the case with deterministic optimization schemes) which fail to provide a qualitative/quantitative idea of the robustness of the solution in the neighbourhood of the optimal point. This offers distinct advantages in a practical design optimization problem by a) clearly identifying the best optimal design regions in the important parameter space, b) having a quantitative estimate of the confidence with which design targets would be met, c) ensuring robustness of operation from the joint posterior distribution on the parameters (key indicators would be correlation coefficient, lower order statistical moments, number of modes) and d) obtaining a one-to-one map between the input distribution of important design variables around the optimal design point and the distribution of the target qoi in presence of system level uncertainty. Thus the proposed method satisfies these key requirements which are necessary to ensure its suitability for practical industrial applications.

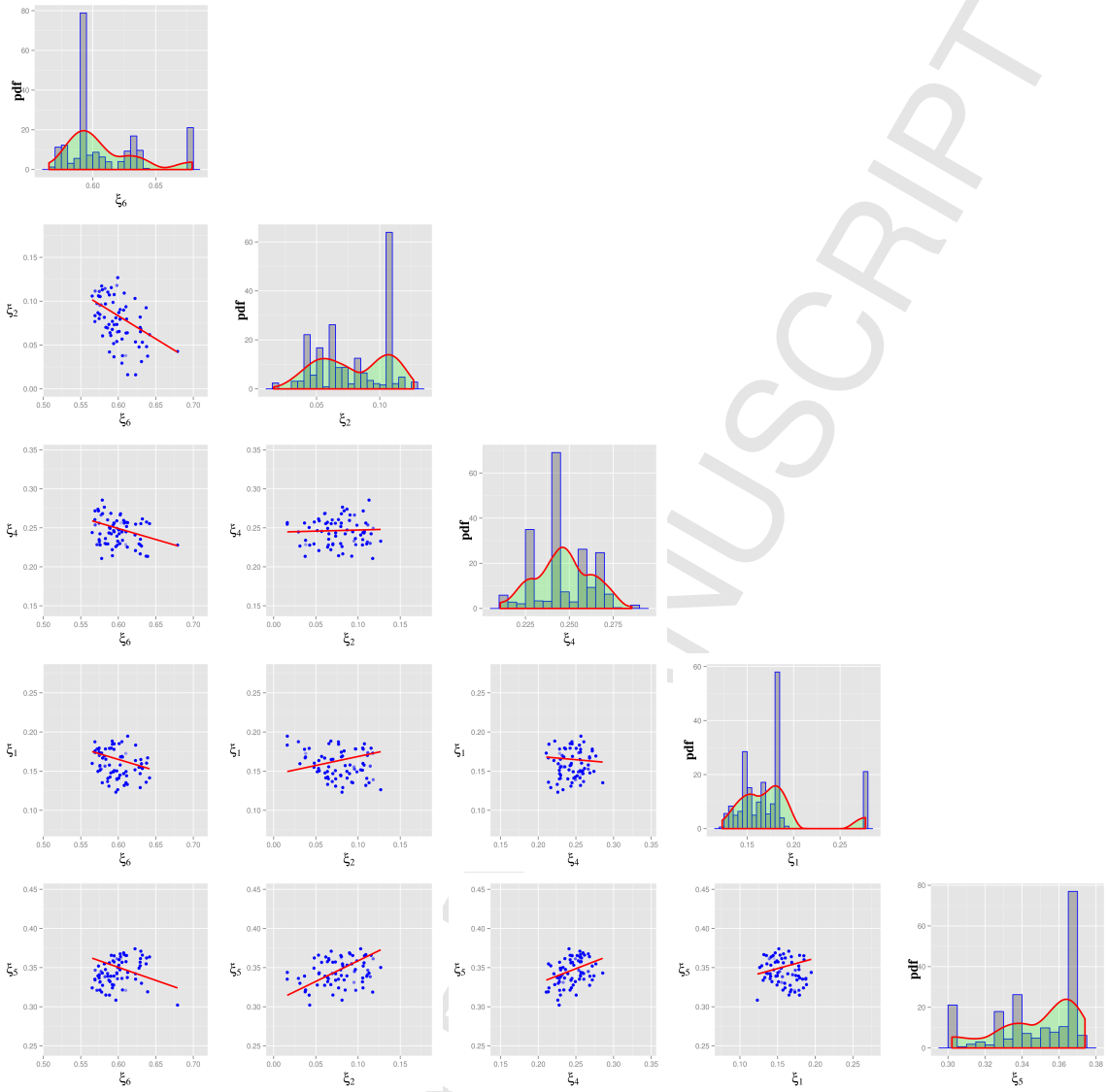


Figure (11): Posterior joint probability density functions of the two random variables ξ_1 and ξ_{12} obtained from the samples generated using the MCMC algorithm.

The proposed robust Bayesian optimization method evaluates the optimal solution which satisfies the target distribution subject to the assumed priors since it has been shown in Sec. 3.1 that the joint posterior distribution derived using the proposed method is an L^2 projection of the important parameters onto the sub-σ algebra generated by the intersection of the reduced parameter space and the observation samples from the target distribution. The likelihood function for the robust optimization scheme is informed by the variability of the response surface at each point of the important parameter space (the design space) due to the prior distributions of the non-important parameters. Thus efficient uncertainty propagation methods are required which can allow for rapid sampling from the multi-dimensional stochastic response surface. The results show that increasing the dimension of the design space gets the posterior distribution of the qoi closer to the prescribed target distribution.

The proposed method has been demonstrated with two examples in which the input parameters have been considered as random inputs with prescribed prior distributions while the target distributions are

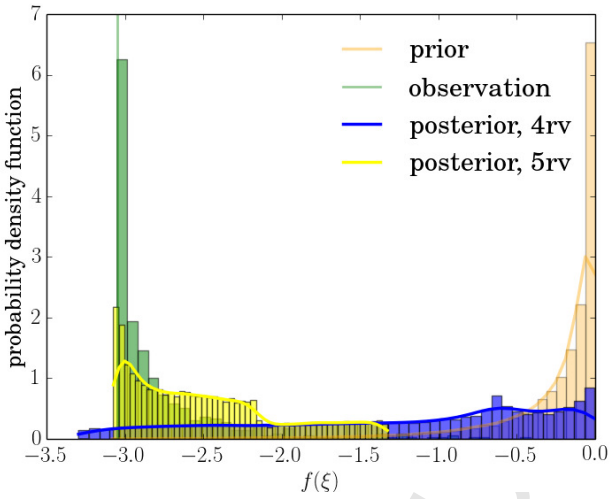


Figure (12): Probability distribution of the Hartmann function with the prior and joint optimal posterior distribution on the input parameters. The 'observation' histogram denotes the samples from the specified target distribution and the optimal posterior distribution have been performed with 4 and 5 parameters (random variables (rv)) with the highest global sensitivity indices.

based on ad-hoc prescribed distributions on the qoi. The structural dynamic system consists of a corrugated structural component where selected geometric and elastic parameters (important parameters) were used for robust optimization to match the target distribution for the first resonant frequency. The second example dealt with a six-dimensional Hartmann function where the problem was posed so as to find the optimal distribution on the important parameters which would bring the Hartmann function close to its global minima. The applicability of the method has been demonstrated even when the target distribution is very dissimilar to the output qoi with the assumed prior distribution on the input parameters.

It is imperative that the proposed robust optimization method based on Bayesian inference is a first attempt at providing a theoretical basis and a proof of concept for the proposed scheme. A number of important directions for the further investigation and consolidation of the proposed method has to be undetaken some of which can be listed as follows.

- Include distance measures between the prior and target distributions which would enable an alternative approach of estimating the likelihood function without resorting to explicitly sampling from the specified target distribution.

- It is necessary to provide the theoretical basis and demonstration of the applicability of the proposed method to multiobjective robust optimization problem. This would be particularly interesting since the Bayesian inference methodology is amenable to accommodating additional constraints and conditions on the study parameters using the general product rule [40] to extend simple problems.

- The investigation of efficient sampling algorithms to improving the speed of convergence of the MCMC algorithm [10, 41, 42] especially when the probability support for the joint optimal posterior distribution is small in a high dimensional parameter space. This would also present the opportunity to reformulate the problem in terms of a dual problem where sampling from the complement space would significantly increase the sample acceptance rate.

- A multi-stage robust optimization problem, where a joint identification of the important parameter space and the associated optimal joint posterior distribution is performed. This would imply that

the definition of the important parameter space can change following the identification of the robust optimum zone based on additional local sensitivity/importance metrics.

Additionally, the considerations of other future research directions would include studying the performance of the proposed method for the case of hierarchical uncertainty models, adaptively enriching the solution accuracy around the areas of high joint posterior densities on the important parameters and choice of the important parameter space on the fly (during the robust optimization process) with local sensitivity measure of the input parameters on the qoi.

Acknowledgement

AK acknowledges the fellowship award from Sêr Cymru National Research Network, project number NRNC25.

References

[1] I. Babuška, R. Tempone, G. E. Zouraris, Galerkin finite element approximations of stochastic elliptic partial differential equations, Siam Journal on Numerical Analysis 42 (2) (2004) 800–825.

[2] R. G. Ghanem, P. D. Spanos, Stochastic Finite Elements: A Spectral Approach, Dover Publications, New York, USA, 2003.

[3] A. Nouy, Recent developments in spectral stochastic methods for the numerical solution ofstochastic partial differential equations, Archives of Computational Methods in Engineering 16 (3) (2009) 251–285.

[4] D. C. Charmpis, G. I. Schuëller, M. F. Pellissetti, The need for linking micromechanics of materials with stochastic finite elements: A challenge for materials science, Computational Materials Science 41 (1) (2007) 27–37.

[5] A. Kundu, F. DiazDelaO, S. Adhikari, M. Friswell, A hybrid spectral and metamodeling approach for the stochastic finite element analysis of structural dynamic systems, Computer Methods in Applied Mechanics and Engineering 270 (0) (2014) 201 – 219.

[6] B. Ganapathysubramanian, N. Zabaras, Sparse grid collocation schemes for stochastic natural convection problems, Journal of Computational Physics 225 (1) (2007) 652–685.

[7] J. Foo, G. E. Karniadakis, Multi-element probabilistic collocation method in high dimensions, Journal of Computational Physics 229 (5) (2010) 1536 – 1557.

[8] H. J. Pradlwarter, G. I. Schuëller", On advanced monte carlo simulation procedures in stochastic structural dynamics, International Journal of Non-Linear Mechanics 32 (4) (1997) 735 – 744, third International Stochastic Structural Dynamics Conference.

[9] F. Yamazaki, M. Shinozuka, Digital generation of non-gaussian stochastic fields, Journal of Engineering Mechanics 114 (7) (1988) 1183–1197.

[10] S. K. Au, J. L. Beck, A new adaptive importance sampling scheme for reliability calculations, Structural Safety 21 (2) (1999) 135 – 158.

[11] R. Sivaraj, T. Ravichandran, A review of selection methods in genetic algorithm, International journal of engineering science and technology 1 (3) (2011) 3792–3797.
URL `http://journals.indexcopernicus.com/abstract.php?icid=946101`

[12] X.-S. Yang, Review of meta-heuristics and generalised evolutionary walk algorithm, International Journal of Bio-Inspired Computation 3 (2) (2011) 77–84.
URL http://www.inderscienceonline.com/doi/abs/10.1504/IJBIC.2011.039907

[13] J. Mokus, On Bayesian methods for seeking the extremum, in: Optimization Techniques IFIP Technical Conference, Springer, 1975, pp. 400–404.
URL http://link.springer.com/content/pdf/10.1007/978-3-662-38527-2_55.pdf

[14] J. Otto, M. Paraschivoiu, S. Yesilyurt, A. T. Patera, Bayesian-validated computer-simulation surrogates for optimization and design: Error estimates and applications, Mathematics and Computers in Simulation 44 (4) (1997) 347–367. doi:10.1016/S0378-4754(97)00061-X.
URL http://www.sciencedirect.com/science/article/pii/S037847549700061X

[15] J. Gu, G. Y. Li, Z. Dong, Hybrid and adaptive meta-model-based global optimization, Engineering Optimization 44 (1) (2012) 87–104. doi:10.1080/0305215X.2011.564768.
URL http://www.tandfonline.com/doi/abs/10.1080/0305215X.2011.564768

[16] L. Shi, R.-J. Yang, P. Zhu, An adaptive response surface method for crashworthiness optimization, Engineering Optimization 45 (11) (2013) 1365–1377. doi:10.1080/0305215X.2012.734815.
URL http://www.tandfonline.com/doi/abs/10.1080/0305215X.2012.734815

[17] G. G. Wang, Adaptive Response Surface Method Using Inherited Latin Hypercube Design Points, Journal of Mechanical Design 125 (2) (2003) 210. doi:10.1115/1.1561044.
URL http://MechanicalDesign.asmedigitalcollection.asme.org/article.aspx?articleid=1447457

[18] G. Gary Wang, Z. Dong, P. Aitchison, ADAPTIVE RESPONSE SURFACE METHOD - A GLOBAL OPTIMIZATION SCHEME FOR APPROXIMATION-BASED DESIGN PROBLEMS, Engineering Optimization 33 (6) (2001) 707–733. doi:10.1080/03052150108940940.
URL http://www.tandfonline.com/doi/abs/10.1080/03052150108940940

[19] S. Shan, G. G. Wang, Failure Surface Frontier for Reliability Assessment on Expensive Performance Function, Journal of Mechanical Design 128 (6) (2006) 1227. doi:10.1115/1.2337311.
URL http://MechanicalDesign.asmedigitalcollection.asme.org/article.aspx?articleid=1449152

[20] L. Wang, S. Shan, G. G. Wang, Mode-pursuing sampling method for global optimization on expensive black-box functions, Engineering Optimization 36 (4) (2004) 419–438. doi:10.1080/03052150410001686486.
URL http://www.tandfonline.com/doi/abs/10.1080/03052150410001686486

[21] G. G. Wang, S. Shan, Review of Metamodeling Techniques in Support of Engineering Design Optimization, Journal of Mechanical Design 129 (4) (2007) 370. doi:10.1115/1.2429697.
URL http://MechanicalDesign.asmedigitalcollection.asme.org/article.aspx?articleid=1449318

[22] W. L. Oberkampf, C. J. Roy, Verification and Validation in Scientific Computing, Cambridge University Press, Cambridge, 2010, dOI: 10.1017/CBO9780511760396.
URL http://ebooks.cambridge.org/ref/id/CBO9780511760396

[23] C. J. Roy, W. L. Oberkampf, A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing, Computer Methods in Applied Mechanics and Engineering 200 (25-28) (2011) 2131–2144. doi:10.1016/j.cma.2011.03.016.
URL http://linkinghub.elsevier.com/retrieve/pii/S0045782511001290

[24] H. G. Matthies, A. Keese, Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations, Computer Methods in Applied Mechanics and Engineering 194 (12-16) (2005) 1295–1331.

[25] G. E. B. Archer, A. Saltelli, I. M. Sobol, Sensitivity measures, ANOVA-like techniques and the use of bootstrap, Journal of Statistical Computation and Simulation 58 (2) (1997) 99–120.

[26] A. Saltelli, J. Marivoet, Non-parametric statistics in sensitivity analysis for model output: A comparison of selected techniques, Reliability Engineering & System Safety 28 (2) (1990) 229–253. doi:10.1016/0951-8320(90)90065-U.
URL http://linkinghub.elsevier.com/retrieve/pii/095183209090065U

[27] A. van Griensven, T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, R. Srinivasan, A global sensitivity analysis tool for the parameters of multi-variable catchment models, Journal of Hydrology 324 (1-4) (2006) 10–23. doi:10.1016/j.jhydrol.2005.09.008.
URL http://linkinghub.elsevier.com/retrieve/pii/S0022169405004488

[28] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index, Computer Physics Communications 181 (2) (2010) 259 – 270.

[29] A. K. Gupta, T. F. Mri, G. J. Szkely, How to transform correlated random variables into uncorrelated ones, Applied Mathematics Letters 13 (6) (2000) 31–33. doi:10.1016/S0893-9659(00)00050-1.
URL http://www.sciencedirect.com/science/article/pii/S0893965900000501

[30] S. Oladyshkin, W. Nowak, Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion, Reliability Engineering & System Safety 106 (2012) 179–190. doi:10.1016/j.ress.2012.05.002.
URL http://www.sciencedirect.com/science/article/pii/S0951832012000853

[31] I. Lee, K. K. Choi, Y. Noh, L. Zhao, D. Gorsich, Sampling-Based Stochastic Sensitivity Analysis Using Score Functions for RBDO Problems With Correlated Random Variables, Journal of Mechanical Design 133 (2) (2011) 021003. doi:10.1115/1.4003186.
URL http://MechanicalDesign.asmedigitalcollection.asme.org/article.aspx?articleid=1449968

[32] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering and System Safety 93 (7) (2008) 964–979.

[33] A. O'Hagan, J. E. Oakley, Probability is perfect, but we can't elicit it perfectly, Reliability Engineering & System Safety 85 (1-3) (2004) 239–248. doi:10.1016/j.ress.2004.03.014.
URL http://linkinghub.elsevier.com/retrieve/pii/S0951832004000638

[34] J. Oakley, Eliciting Gaussian process priors for complex computer codes, The Statistician 51 (1) (2002) 81–97.

[35] F. A. Moala, A. O'Hagan, Elicitation of multivariate prior distributions: A nonparametric Bayesian approach, Journal of Statistical Planning and Inference 140 (7) (2010) 1635–1655.

[36] W. P. Aspinall, R. M. Cooke, Quantifying scientific uncertainty from expert judgement elicitation, in: Risk and uncertainty assessment for natural hazards, Cambridge University Press Cambridge, UK, 2013, p. 64.

[37] J. E. Oakley, A. O'Hagan, Uncertainty in prior elicitations: a nonparametric approach, Biometrika 94 (2) (2007) 427–441. doi:10.1093/biomet/asm031.
URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/asm031

[38] A. O'Hagan, C. Buck, A. Daneshkhah, J. Eiser, P. Garthwaite, D. Jenkinson, J. Oakley, T. Rakow, Uncertain Judgements: Eliciting Experts' Probabilities, Statistics in Practice, Wiley, 2006.
URL https://books.google.co.uk/books?id=H9KswqPWIDQC

[39] S. Abe, C. Beck, E. G. D. Cohen, Superstatistics, thermodynamics, and fluctuations, Physical Review E 76 (3) (2007) 031102. doi:10.1103/PhysRevE.76.031102.
URL https://link.aps.org/doi/10.1103/PhysRevE.76.031102

[40] A. Bobrowski, Functional Analysis for Probability and Stochastic Processes, Cambridge University Press, 2005.

[41] D. Gamerman, H. F. Lopes, Markov chain Monte Carlo: stochastic simulation for Bayesian inference, CRC Press, 2006.

[42] H. Haario, M. Laine, A. Mira, E. Saksman, DRAM: Efficient adaptive MCMC, Statistics and Computing 16 (4) (2006) 339–354. doi:10.1007/s11222-006-9438-0.
URL http://link.springer.com/10.1007/s11222-006-9438-0

[43] M. Kleiber, T. D. Hien, The Stochastic Finite Element Method, John Wiley, Chichester, 1992.

[44] H. G. Matthies, C. E. Brenner, C. G. Bucher, C. G. Soares, Uncertainties in probabilistic numerical analysis of structures and solids - stochastic finite elements, Structural Safety 19 (3) (1997) 283–336.

[45] I. Babuška, R. Tempone, G. E. Zouraris, Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation, Computer Methods in Applied Mechanics and Engineering 194 (12-16) (2005) 1251 – 1294.

[46] A. Kundu, S. Adhikari, Transient response of structural dynamic systems with parametric uncertainty, Journal of Engineering Mechanics 140 (2) (2014) 315–331.

[47] X. L. Wan, G. E. Karniadakis, Beyond wiener-askey expansions: Handling arbitrary pdfs, Journal of Scientific Computing 27 ((-3) (2006) 455–464.

[48] A. Kundu, S. Adhikari, M. I. Friswell, Stochastic finite elements of discretely parameterized random systems on domains with boundary uncertainty, International Journal for Numerical Methods in Engineering 100 (3) 183–221. doi:10.1002/nme.4733.

Highlights

- A novel Bayesian-inference based probabilistic optimal design is proposed.
- Optimal distribution on a reduced set of important design parameters is derived.
- The likelihood explicitly considers the variability due to the reduced parameter set.
- Conditional expectation theory is used to prove the existence of the posterior.
- The method is demonstrated with non-linear functions and structural dynamics systems.