Multi-Model Polynomial Chaos Surrogate Dictionary for Bayesian Inference in Elasticity Problems

Andres A. Contreras[a,*], Olivier P. Le Maître[a,b], Wilkins Aquino[a], Omar M. Knio[a,c]

[a]*Duke University, Durham, NC*
[b]*LIMSI-CNRS, Orsay, France*
[c]*King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

Abstract

A method is presented for inferring the presence of an inclusion inside a domain; the proposed approach is suitable to be used in a diagnostic device with low computational power. Specifically, we use the Bayesian framework for the inference of stiff inclusions embedded in a soft matrix, mimicking tumors in soft tissues. We rely on a Polynomial Chaos (PC) surrogate to accelerate the inference process. The PC surrogate predicts the dependence of the displacements field with the random elastic moduli of the materials, and are computed by means of the Stochastic Galerkin (SG) projection method. Moreover, the inclusion's geometry is assumed to be unknown, and this is addressed by using a dictionary consisting of several geometrical models with different configurations. A model selection approach based on the evidence provided by the data (Bayes factors) is used to discriminate among the different geometrical models and select the most suitable one. The idea of using a dictionary of pre-computed geometrical models helps to maintain the computational cost of the inference process very low, as most of the computational burden is carried out off-line for the resolution of the SG problems. Numerical tests are used to validate the methodology, assess its performance, and analyze the robustness to model errors.

Keywords: Bayesian Inference, Model Selection, Uncertainty Quantification, Polynomial Chaos, Elasticity Imaging, Soft Tissue

1. Introduction

The nondestructive characterization of the parameters describing a physical system is a task of great importance and interest in various disciplines within science and engineering. Examples of such tasks include seismic imaging [1, 2], health monitoring of infrastructure [3, 4, 5], and more recently elasticity imaging [6, 7]. Elasticity imaging is a very promising branch of medical diagnosis which applies inverse problems techniques to compute the elasticity modulus given a set of measurements of a displacement or velocity field that is the result of some excitation force [8]. The idea is inspired by the palpation technique used by doctors to determine the presence of abnormal tissue through the sense of touch [9, 10]. Palpation, however, is limited in detecting anomalies that lie deep in the body or which are too small [11]; moreover, it tends to be qualitative as opposed to quantitative. Elasticity imaging takes palpation to the next level by extending its range and effectiveness, all in a more quantitative manner. The general goal of this work is to use a collection of models within a Bayesian framework to estimate the contrast between the elastic properties of different regions in a given domain.

The elasticity imaging technique encompasses three basic steps: first, the body is deformed through an applied external load, then the deformation field is measured (e.g. using ultrasound techniques), and finally the elastic properties are estimated by solving an inverse problem. To approach this problem, Oberai *et al.* [8]

---

*Corresponding author
Email address: `andres.contreras@duke.edu` (Andres A. Contreras)

assume that the displacements are governed by the equations of equilibrium of an incompressible, linear-elastic solid undergoing small, quasi-static deformation, and cast the problem as a non-linear optimization problem; the objective is to find a shear modulus field that minimizes the discrepancy between the measured and predicted displacement fields. Another optimization approach is based on minimizing the Modified Error in Constitutive Equation Functional [12], which measures the discrepancy in the constitutive equations that connect kinematically admissible strains and dynamically admissible stresses in addition to measuring the discrepancy between the measured and predicted displacement fields. Other approaches include direct inversion methods [13, 14, 15], but these methods, although computationally less expensive, tend to be more sensitive to noise in measurement data. All these approaches are deterministic, and consequently result in a single estimate of the elastic modulus, which doesn't accommodate for the quantification of uncertainty.

Important insights emerge by approaching inverse problems using a probabilistic framework. Some of the methods introduced to deal with this problem include the extended maximum likelihood method [16], the spectral stochastic method [17, 18], the sparse grid collocation approach [19, 20], stochastic reduced order models [21], and the Bayesian inference approach [22, 23]. In the Bayesian formalism, one obtains additional insight by computing a probability distribution that summarizes all available information about the elastic moduli (e.g., we can estimate moments, marginal distributions, quantiles), as opposed to the single value obtained in the deterministic setting. Specifically, in the context of elasticity imaging, Koutsourelakis [24] use a Bayesian framework to obtain probabilistic estimates of the material properties that account for various possible sources of uncertainty; this works deals with simplified geometries and large contrast ratio of the elastic properties. Another interesting approach is proposed by Iglesias [25], where the Bayesian framework is applied under an infinite dimensional setting; this work, however, is limited to deterministic (known) elastic properties and it requires an appropriate prior model.

For complex forward models, extracting information from the posterior distribution can be very computationally expensive. Several techniques are applied to address this computational challenge, such as the use of a two-stage MCMC to increase the acceptance rate of the algorithm by using an inexpensive approximation of the posterior distribution [26, 27, 28]; the use of proper orthogonal decomposition (POD) to construct a reduced-order model for the direct simulations [29, 30]; the use of adaptive hierarchical sparse grid collocation (ASGC) to obtain an approximate stochastic solution to the forward problem using piecewise linear interpolation [31]; and the use of Polynomial Chaos (PC) to approximate the solution of the stochastic forward model either through collocation [32, 33] or through the stochastic Galerkin method [34]. A related application of PC representations in the context of inverse acoustic scattering problems is found in [35], where PC expansions are integrated with optimization methods for the probabilistic characterization of hidden obstacles and inclusions in acoustic media.

The objective of this work is to develop a method that can be used in a diagnostic device with a low computational power to quickly assess the presence of an inclusion in a given domain. To achieve this, the proposed approach breaks the process in two steps: (1) an offline or pre-processing step where surrogate models are constructed for different geometrical models, and (2) an online step where a model selection and inference are performed on the basis of observations to assess the presence of an inclusion. This is advantageous, since the main computational cost is carried by the construction of the surrogate models, which is something that can be done offline with a dedicated computer. Thus, once the surrogate models have been constructed, the computational cost of the model selection and inference problem is relatively low and can be effectively handled by the diagnostic device with limited computational power. In more details, we extend the Bayesian approach proposed by Marzouk *et al.* [34] to the case of multiple geometrical models as follows. First, a dictionary of inclusion geometries is considered and for each of these geometries a suitable Polynomial Chaos expansion of the displacement field is computed, in terms of the unknown parameters (in our case the Young's modulus and the Poisson's ratio in soft matrix and inclusion) by means of the Stochastic Galerkin (SG) method [36, 37]. The SG allows for a fine control approximation error. When observations are made available, the PC surrogates can be used to derived corresponding approximations of the posterior distribution for the elastic properties given a geometry. Then, these posteriors can be compared by computing the evidences or Bayes factors of the geometrical models, in order to rank them and select the best one (or few best ones). The posterior distribution(s) of the elastic properties for the best model (or best ones) can then be used to reach a decision confirming or refuting the presence of an inclusion, analyzing

for instance the ratios between the mean properties in the inclusion and soft matrix domains.

The outline of the paper is as follows. In section 2, we introduce the mechanical model of the elasticity problem and derive the Polynomial Chaos expansions of the displacement field for a given geometry. In section 3, we describe the use of the Bayesian framework to solve both inverse problem and the model selection problem. In section 4 we present some numerical results showing the behavior of the approach when the exact geometry of the model is known. Then, in section 5, we look at the case when the exact model geometry is unknown and construct a dictionary of surrogate models and rank them based on the evidence provided by the data; also we test the robustness of the approach with respect to errors in the mechanical model. Finally, in section 6 we provide concluding remarks.

2. Physical Model and Polynomial Chaos Expansion

2.1. Physical Model

2.1.1. Continuous problem

The strong form of the equilibrium equations of a linear-elastic solid undergoing static deformation due to boundary loads and displacements can be expressed as:

$$\nabla \cdot \boldsymbol{\sigma} = 0 \quad \text{in } \Omega \tag{1}$$

with boundary conditions:

$$\boldsymbol{\sigma} \cdot \boldsymbol{n} = \boldsymbol{\tau} \ \text{ on } \Gamma_\tau \ , \quad \boldsymbol{u} = \boldsymbol{u_0} \ \text{ on } \Gamma_u, \tag{2}$$

where $\boldsymbol{\sigma} = \mathbb{C} : \boldsymbol{\epsilon} \equiv \mathbb{C}_{ijkl} \ \boldsymbol{\epsilon}_{kl}$ is the stress tensor; $\boldsymbol{\epsilon}(\boldsymbol{u}) = (\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T)/2$ is the linearized strain tensor; $\boldsymbol{u}$ is the displacement field; $\boldsymbol{n}$ the unit normal to the boundary; $\boldsymbol{\tau}$ is the traction vector; Ω is the spatial domain; Γ_τ and Γ_u form a partition of the boundary Γ of Ω; $\boldsymbol{u_0}$ is the essential boundary condition; and $\mathbb{C}$ is the fourth-order constitutive tensor of linear elasticity. Under the assumption of an isotropic medium, the constitutive tensor has only two independent elastic constants and can be written as:

$$\mathbb{C}_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu \left(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} \right) \tag{3}$$

where δ_{ij} is the Kronecker delta and λ and μ are the Lamé constants [38]. This decomposition of $\mathbb{C}$ is very advantageous for the computation of the PC coefficients described in section 2.2.1.

The forward problem consists in finding the displacement field $\boldsymbol{u}$ that satisfies (1) for a given constitutive tensor $\mathbb{C}$ (i.e. known material properties). The weak formulation of the forward problem is obtained after defining the space of trial functions, $\mathcal{S} = \left\{ \boldsymbol{u} \mid u_i \in H^1(\Omega), \boldsymbol{u} = \boldsymbol{u_0} \text{ on } \Gamma_u \right\}$, and the space of test functions, $\mathcal{V} = \left\{ \boldsymbol{v} \mid v_i \in H^1(\Omega), \boldsymbol{v} = \boldsymbol{0} \text{ on } \Gamma_u \right\}$. Multiplying (1) by an arbitrary $\boldsymbol{v} \in \mathcal{V}$, integrating over the spatial domain, using the divergence theorem, and the symmetry of $\mathbb{C}$ we get:

$$a\left(\boldsymbol{u}, \boldsymbol{v}\right) = \left(\boldsymbol{\tau}, \boldsymbol{v}\right), \qquad \forall \, \boldsymbol{v} \in \mathcal{V}, \tag{4}$$

where

$$a\left(\boldsymbol{u}, \boldsymbol{v}\right) \equiv \int_\Omega \boldsymbol{\epsilon}(\boldsymbol{v}) : \mathbb{C} : \boldsymbol{\epsilon}(\boldsymbol{u}) d\Omega \ , \quad \left(\boldsymbol{\tau}, \boldsymbol{v}\right) \equiv \int_{\Gamma_\tau} \boldsymbol{\tau} \cdot \boldsymbol{v} d\Gamma_\tau. \tag{5}$$

The function $\boldsymbol{u} \in \mathcal{S}$ that satisfies (4) is the equivalent weak solution of (1).

3

2.1.2. Finite element formulation

Using standard Voigt notation [39], the displacement fields, test functions, and their derivatives are expressed as:

$$\boldsymbol{u}^h = [N]\{u\}, \quad \boldsymbol{v}^h = [N]\{v\}, \quad \boldsymbol{\epsilon}^h(\boldsymbol{u}) = [B]\{u\}, \quad \boldsymbol{\epsilon}^h(\boldsymbol{v}) = [B]\{v\}, \tag{6}$$

where $\{u\}$ and $\{v\}$ belong to $\mathbb{R}^m$, m depends on the resolution of the finite element basis, whereas $[N]$ and $[B]$ represent matrices of finite element shape functions and their derivatives with respect to spatial coordinates, respectively. Substituting the above approximations into the variational problem in (4) we get:

$$[A]\{u\} = \{F\}, \tag{7}$$

where

$$[A] \equiv \sum_{\text{elements}} \int_{\Omega^e} [B]^T [C] [B] \, d\Omega \, , \quad \{F\} \equiv \sum_{\text{elements}} \int_{\Gamma^e_\tau} [N]^T \boldsymbol{\tau} \, d\Gamma, \tag{8}$$

and $[C]$ is the matrix representation of the fourth-oder tensor $\mathbb{C}$. Thus, the solution of the discretized forward problem reduces to inverting the linear system given by (7).

Notice that through (7) we can define a map

$$\mathcal{T} : [C] \to \{u\}, \tag{9}$$

that takes a constitutive matrix as its input, and outputs a discretized displacement field. This mapping, of course, is determined by the geometry of the problem and its boundary conditions. We will refer to this map as the Deterministic Forward Map.

2.2. Polynomial Chaos Surrogate Model

From now on we will study the situation where the Lamé constants, λ and μ, are stochastic quantities (i.e. $[C]$ is random) and will layout the framework to accommodate this situation. As an initial step we seek to construct a stochastic map $\hat{\mathcal{T}}$ that approximates (9) when $[C]$ is random and that can be efficiently sampled. This map will be constructed using polynomial chaos expansion [36, 37], and we will refer to it as the *surrogate model*.

In this work, we focus on the specific class of random processes that are in L^2, the space of second-order quantities. Following the notation in [37], let $(\Theta, \Sigma_\Theta, P_\Theta)$ be a probability space and θ a random event belonging to Θ. We denote $L^2(\Theta, P_\Theta)$ the space of second-order random variables defined on $(\Theta, \Sigma_\Theta, P_\Theta)$ equipped with the inner product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \|_\Theta$:

$$\langle X, Y \rangle = \int_\Theta X(\theta)Y(\theta)dP_\Theta(\theta) = \mathbb{E}[XY] \qquad \forall X, Y \in L^2(\Theta, P_\Theta) \tag{10}$$

$$X \in L^2(\Theta, P_\Theta) \iff \langle X, X \rangle = \|X\|_\Theta^2 < \infty, \tag{11}$$

where $\mathbb{E}[\cdot]$ is the expectation operator.

The discrete counterpart of (3) is

$$[C] = \lambda \left[C^\lambda\right] + \mu \left[C^\mu\right], \tag{12}$$

where $\left[C^\lambda\right]$ and $[C^\mu]$ are constant matrices independent of both λ and μ. The parameters λ and μ are bounded quantities and we will model them as a random processes $\lambda(\boldsymbol{x}, \theta), \mu(\boldsymbol{x}, \theta) \in L^\infty(\Omega) \times L^2(\Theta, P_\Theta)$, and represent them in terms of the truncated PC expansions:

$$\lambda(\boldsymbol{x}, \theta) \simeq \sum_{\gamma=0}^{P_\lambda} \lambda_\gamma(\boldsymbol{x}) \Psi_\gamma(\boldsymbol{\xi}(\theta)),$$

$$\mu(\boldsymbol{x}, \theta) \simeq \sum_{\gamma=0}^{P_\mu} \mu_\gamma(\boldsymbol{x}) \Psi_\gamma(\boldsymbol{\xi}(\theta)), \tag{13}$$

4

where $\boldsymbol{\xi}\left(\theta\right) = \{\xi_1\left(\theta\right), \xi_2\left(\theta\right), \cdots, \xi_N\left(\theta\right)\}$ is a N-dimensional random vector, with independent components having joint density $p_{\boldsymbol{\xi}}$ and $P_\lambda + 1$ and $P_\mu + 1$ are the number of terms in each corresponding expansion. The Ψ_γ form an orthonormal basis, and are indexed such that $\Psi_0 = 1$. For simplicity we will take the number of terms in both expansions to be the same, i.e. $P_\lambda = P_\mu = P$. Furthermore, the number of terms in each expansion is determined by the number of random variables, N, and the polynomial degree of the PC expansion, p (again, we take the same polynomial degree for both expansions); for a total order truncation, it is given by [37]:

$$P + 1 = \frac{(N + p)!}{N!\, p!}. \tag{14}$$

We also represent the stochastic discrete displacement field as a random field using a PC expansion of the form:

$$\{\hat{u}(\theta)\} \simeq \sum_{\alpha=0}^{P_u} \{u_\alpha\} \Psi_\alpha(\boldsymbol{\xi}(\theta)), \tag{15}$$

where $\boldsymbol{\xi}\left(\theta\right)$ is the same random vector introduced above, $\{u_\alpha\} \in \mathbb{R}^m$ are the PC coefficients, and $P + 1$ is the number of terms in the PC expansion of u. P_u is determined from N and p_u using (14). Without loss of generality we take $P_u = P$.

2.2.1. Computation of PC coefficients

In this section we will describe the method to compute the coefficients in (15). We start by substituting (13) into (12) to get the following PC expansion for $[C]$:

$$[C]\left(\boldsymbol{\xi}\right) \simeq \sum_{\gamma=0}^{P} [C_\gamma]\, \Psi_\gamma(\boldsymbol{\xi}) = \sum_{\gamma=0}^{P} \left(\lambda_\gamma\left[C^\lambda\right] + \mu_\gamma\left[C^\mu\right]\right) \Psi_\gamma(\boldsymbol{\xi}). \tag{16}$$

Notice that the decomposition of $[C]$ as given in (12) provides an explicit way to compute the $[C_\gamma]$'s from the coefficients in the PC expansions of λ and μ. Substituting (16) into (8) and using the Galerkin projection on (7) accounting for (15), we get the following coupled system of $P + 1$ equations for the PC coefficients:

$$[A_0]\{u_\beta\} + \sum_{\alpha=0}^{P} \sum_{\gamma=1}^{P} [A_\gamma]\{u_\alpha\} \langle \Psi_\gamma \Psi_\alpha, \Psi_\beta \rangle = \{F_\beta\}, \quad \beta = 0, \cdots, P, \tag{17}$$

where

$$[A_\gamma] = \sum_{\text{elements}} \int_{\Omega^e} [B]^T [C_\gamma] [B]\, d\Omega \qquad \gamma = 0, \cdots, P.$$

Equation (17) can be written more compactly as

$$[\mathcal{A}]\{\mathcal{U}\} = \{\mathcal{F}\}, \tag{18}$$

where $[\mathcal{A}] = [\bar{\mathcal{A}}] + [\tilde{\mathcal{A}}]$; $[\bar{\mathcal{A}}]_{\alpha,\beta} = [A_0]\,\delta_{\alpha,\beta}$; $[\tilde{\mathcal{A}}]_{\alpha,\beta} = \sum_{\gamma=1}^{P} [A_\gamma]\, \langle \Psi_\gamma \Psi_\alpha, \Psi_\beta \rangle$; $\{\mathcal{U}\}_\beta = \{u_\beta\}$; and $\{\mathcal{F}\}_\beta = \{F_\beta\}$; $0 \le \alpha, \beta \le P$.

We notice that the matrix $[A]$ in the deterministic case, (7), is $m \times m$, whereas the matrix $[\mathcal{A}]$ in (18) is $(m \cdot (P + 1)) \times (m \cdot (P + 1))$. Thus, if m is large to start with, which is typically the case, solving for $\{\mathcal{U}\}$ might incur a significant computational cost. However, effectively exploiting the structure of $[\mathcal{A}]$ could offer great computational savings. A simple approach [40, 41, 37] is to take advantage of the decomposition $[\mathcal{A}] = [\bar{\mathcal{A}}] + [\tilde{\mathcal{A}}]$, which allows us to solve for $\{\mathcal{U}\}$ iteratively according to:

$$\{\mathcal{U}\}_{k+1} = [\bar{\mathcal{A}}]^{-1}\left(\{\mathcal{F}\} - [\tilde{\mathcal{A}}]\{\mathcal{U}\}_k\right). \tag{19}$$

Since $[\bar{\mathcal{A}}]$ is block-diagonal, the iterations in (19) can be decoupled into $(P+1)$ independent updates for the $\{u_\beta\}$'s, requiring the unique factorization of $[A_0] \in \mathbb{R}^{m \times m}$. Thus, the computation can be parallelized, and the same deterministic solver can be used in the iterations.

So far we have expressed the constitutive matrix $[C]$ in term of the Lamé constants λ and μ because they facilitate the computation of the PC expansion of u. The proposed approach can be immediately adapted to work in terms of alternative pairs of elastic constants. For instance, it might be desirable to think in terms of the Young's modulus (E) and the Poisson's ratio (ν). When this is the case, we can first determine the corresponding PC expansions for E and ν and substitute them in the following relations

$$\lambda(\boldsymbol{x}, \theta) = \frac{E(\boldsymbol{x}, \theta)\, \nu(\boldsymbol{x}, \theta)}{(1 + \nu(\boldsymbol{x}, \theta))(1 - 2\nu(\boldsymbol{x}, \theta))}, \qquad \mu(\boldsymbol{x}, \theta) = \frac{E(\boldsymbol{x}, \theta)}{2(1 + \nu(\boldsymbol{x}, \theta))}, \tag{20}$$

using the Galerkin product and division [37]. The polynomial order in the expansions is selected such that the error introduced in this step is sufficiently small. The same can be done for other pairs of elastic constants such as the shear modulus and the bulk modulus. In the present work, we considered model based in (E, ν) variabilities, though we take the Poisson's ratio in a range $\nu \approx 0.5$ where the material is nearly incompressible and in agreement or consistent with other properties variability reported in the literature (see *e.g.* [42, 43, 44, 45]).

2.3. Error in the PC surrogate model

In order to assess the fidelity of the surrogate, we rely on the following error measure,

$$\hat{\vartheta}^2 \equiv \frac{\mathbb{E}\left[\|u(\boldsymbol{\xi}) - \hat{u}(\boldsymbol{\xi})\|^2_{L^2(\Omega)}\right]}{\mathbb{E}\left[\|u(\boldsymbol{\xi})\|^2_{L^2(\Omega)}\right]}, \tag{21}$$

where u is the displacement computed using the full forward model and $\hat{u} = \sum_{\alpha=0}^P u_\alpha \Psi_\alpha$ is the displacement computed using the surrogate model.

We approximate (21) by Monte Carlo sampling, computing the sample mean and integrating over the discretized finite-element domain, i.e. we estimate $\hat{\vartheta}$ using

$$\hat{\vartheta}^2 \approx \vartheta^2 \equiv \frac{\sum_{i=1}^{M_S} \|u(\boldsymbol{\xi}^{(i)}) - \hat{u}(\boldsymbol{\xi}^{(i)})\|^2_{L^2(\mathcal{S})}}{\sum_{i=1}^{M_S} \|u(\boldsymbol{\xi}^{(i)})\|^2_{L^2(\mathcal{S})}}, \tag{22}$$

where M_S is the number of Monte Carlo samples and $\boldsymbol{\xi}^{(i)}$ are i.i.d. random realizations of $\boldsymbol{\xi}(\theta)$. In practice, a low number of samples M_S can be used to get a sense of how well the surrogate approximates the full model over the entire range of $\boldsymbol{\xi}$. Further, we recall that this surrogate construction is performed offline, and not during the inference step.

One parameter that impacts ϑ is P (or p). Recall that this parameter represents the number of terms in the approximations of $[C]$ and $\hat{u}$. In section 4, we will analyze how the error ϑ changes as a function of P. This will allow us to characterize how the PC expansion of the displacement (i.e. the surrogate model) converges as we increase the number of terms in the approximation.

3. Bayesian Inference and Model Selection

This section focuses on applying the Bayesian framework to solve two tasks: 1) the inference problem of estimating $[C]$ for a given geometrical model and 2) for a given collection of models, how to rank them in order to select the ones that are more suitable.

Figure 1 shows the schematic of the physical model and also two different approximations of the physical model that can be used for the construction of a surrogate model. In the first part of this section, we discuss the case in which the physical geometry is completely known and we describe how to estimate the material

properties of the domain using Bayesian inference. However, in practice, the actual shape and/or location of the inclusion is unlikely to be completely known, and in the second part of this section we present an approach that consists in creating a dictionary of surrogate models each built with a different geometrical configuration for the size and/or location of the inclusion. We apply a model selection analysis based on the evidence provided by the data to discriminate among the different models.
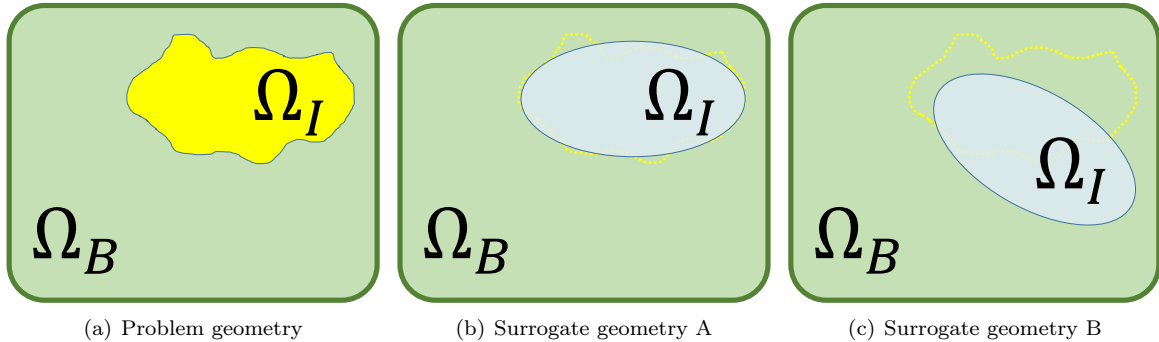

(a) Problem geometry   (b) Surrogate geometry A   (c) Surrogate geometry B

Figure 1: Schematic of the elasticity problem in which a given domain Ω has an inclusion Ω_I with different material properties. The part of the domain which excludes the inclusion is denoted by Ω_B. Figure (a) represents the geometry corresponding to the physical model. Figures (b) and (c) correspond to two different approximations of physical geometry used to build surrogate models.

3.1. Bayesian Inference

We now focus on estimating $[C]$ using Bayesian inference. We assume knowledge of a geometry $\mathcal{G}$ and the availability of m_o noisy observations of the displacement u at points $\boldsymbol{x}_i \in \Omega$. Since $[C]$ is parametrized by $\boldsymbol{\xi}$, the Bayesian inference is recast in terms of the parameter. We rely on a Metropolis-Hastings algorithm to determine the posterior. The covariance matrix of the proposal distribution is determined using information from the Hessian as outlined in Appendix A.

Using Bayes' formula, the posterior distribution of $\boldsymbol{\xi}$ is given by

$$p(\boldsymbol{\xi}, \epsilon | \{Y_i\}, \mathcal{G}) \propto p_Y(\{Y_i\} | \boldsymbol{\xi}, \epsilon, \mathcal{G}) p_\xi(\boldsymbol{\xi}) p_\epsilon(\epsilon), \tag{23}$$

where $p_Y(\{Y_i\} | \boldsymbol{\xi}, \epsilon, \mathcal{G})$ is the likelihood of the observations, $p_\xi(\boldsymbol{\xi})$ is the prior, $p_\epsilon(\epsilon)$ is the prior distribution of the noise hyper-parameter ϵ. The prior could be made dependent on the geometry $\mathcal{G}$, but for the sake of simplicity we will take the same prior for all the geometries.

We assume that the noise in the observations follows a Gaussian distribution, *i.e.* $Y_i = u(\boldsymbol{x}_i) + \eta_i$, $i = 1, \cdots, m_o$, where the η_i's are i.i.d centered Gaussian RVs, with variance ϵ^2. Since the η_i's are independent and Gaussian, it follows that

$$p_Y(\{Y_i\} | \boldsymbol{\xi}, \epsilon, \mathcal{G}) = \prod_{i=1}^{m_o} \frac{1}{\sqrt{2\pi\epsilon^2}} \exp\left(-\frac{(\hat{u}(\boldsymbol{x}_i, \boldsymbol{\xi}; \mathcal{G}) - Y_i)^2}{2\epsilon^2} \right). \tag{24}$$

Notice that in (24) we have replaced u by $\hat{u}$ (the surrogate model approximation). This substitution allows us to sample the posterior distribution at low computational cost [34, 46]. Of course, success of this approximation depends on the availability of a faithful surrogate, as further analyzed below.

We rely on an uninformative Jeffreys prior [47, 48] for the hyper-parameter $\epsilon > 0$, *i.e.* we set:

$$p_\epsilon(\epsilon) \sim \frac{1}{\epsilon}. \tag{25}$$

As for the prior, we assume that material properties of the background and the inclusion are homogeneous, i.e. they are constant along the spatial dimension in each respective domain. Furthermore, we work in terms

of Poisson's ratio and Young's modulus, and assume both of them to be uncertain. The background and the inclusion are assumed to have the same, but uncertain, Poisson's ratio ($\nu(\boldsymbol{x}, \theta) \equiv \nu(\theta)$) and a potentially different Young's modulus ($E(\boldsymbol{x}, \theta)$). Consequently, we express the Young's modulus as:

$$E(\boldsymbol{x}, \theta) = E_B(\theta)\left(1 - \mathbb{I}(\boldsymbol{x})\right) + E_I(\theta)\mathbb{I}(\boldsymbol{x}) \tag{26}$$

where E_B and E_I are the Young's moduli of the background and the inclusion, respectively; $\mathbb{I}(\boldsymbol{x})$ is an indicator function that takes a value of one inside the inclusion and zero outside. Thus, the prior distribution is parameterized using a three-dimensional germ, $\boldsymbol{\xi}(\theta) = \{\xi_1(\theta), \xi_2(\theta), \xi_3(\theta)\}$; one dimension is used to characterize the uncertainty in Poisson's ratio and the other two to characterize the uncertainty in the Young's modulus of the background and the inclusion.

The Poisson ratio (ν) and the background Young's modulus (E_B) are modeled as uniform random variables on the intervals (ν^{min}, ν^{max}) and (E_B^{min}, E_B^{max}), respectively. The inclusion Young's modulus (E_I) is modeled as $E_I = \kappa E_B$, where κ is uniformly distributed on ($1, \kappa^{max}$). Parameterizing in terms of the ξ_i's we have:

$$\nu(\theta) = \nu(\xi_1(\theta)) = \nu^{min} + (\nu^{max} - \nu^{min})\xi_1(\theta)), \qquad \xi_1 \sim \mathcal{U}(0,1) \tag{27}$$

$$E_B(\theta) = E_B(\xi_1(\theta)) = E_B^{min} + (E_B^{max} - E_B^{min})\xi_2(\theta)), \qquad \xi_2 \sim \mathcal{U}(0,1) \tag{28}$$

$$E_I(\theta) = E_I(\xi_1(\theta)) = [1 + (\kappa^{max} - 1)\xi_3(\theta)]E_B(\theta), \qquad \xi_3 \sim \mathcal{U}(0,1) \tag{29}$$

where the parameters $\nu^{min}, \nu^{max}, E_B^{min}, E_B^{max}$, and κ^{max} are variables that encode the prior knowledge. In particular, μ^{max} characterizes the amount by which E_I surpasses E_B. For this parametrization we have almost surely $E_I > E_B$. Note that following the definitions above, the ξ_i's are uniformly distributed over [0,1]. Therefore, the PC basis functions $\Psi_\alpha(\xi_1, \xi_2, \xi_3)$ consist in products of univariate shifted-Legendre polynomials.

Table 1 shows the values for the prior parameters that are used, unless otherwise specified, throughout this paper. The values are based on the ranges used in [43, 44, 45, 49, 50]. Note that the Poisson's ratio has been treated as random to show the generality of the method, although in imaging application it is usually considered known and fixed, assuming incompressible tissues. The prior distribution is however selected so that the soft tissue is modeled as nearly incompressible for almost any event.

Parameter	Value
E_B^{min}	1×10^4Pa
E_B^{max}	9×10^4Pa
ν^{min}	0.45
ν^{max}	0.49
κ^{max}	6

Table 1: Default parameters corresponding to the Young's modulus prior [49, 50] and the Poisson's ratio prior [43, 44, 45].

3.2. Evidence-Based Model Selection

As stated before, complete knowledge of the geometry might not be available. However, in order to build the surrogate model that speeds up the MCMC algorithm, we need the specifics of the geometry. To address this we propose to build not one, but many surrogates using different geometrical models. So, even if we don't know the true geometrical model, we can find a good approximation if our dictionary is extensive enough. For instance, for the physical model shown in figure 1(a), a surrogate built with the geometry shown in figure 1(b) would allow us to properly infer the presence of an inclusion in the domain; this will be further illustrated in section 5. A remaining question, however, is how to discriminate among the different models in our dictionary to select a suitable one. To address this, we consider the probability of observing the data given the geometrical model $\mathcal{G}_k$ (commonly referred as the evidence)

$$I_k \doteq p_Y(\{Y_i\}|\mathcal{G}_k) = \int p_Y(\{Y_i\}|\boldsymbol{\xi}, \epsilon, \mathcal{G}_k)p_\xi(\boldsymbol{\xi})p_\epsilon(\epsilon)d\boldsymbol{\xi}d\epsilon. \tag{30}$$

Evaluating the integrals $I_k = \int p_Y(\{Y_i\}|\boldsymbol{\xi}, \epsilon, \mathcal{G}_k) p_\xi(\boldsymbol{\xi}) p_\epsilon(\epsilon) d\boldsymbol{\xi} d\epsilon$ can be computationally difficult and expensive. However, when the posterior density is highly peaked about its maximum a posteriori (MAP) $(\tilde{\boldsymbol{\xi}}, \tilde{\epsilon})$, we can use Laplace's method to approximate the integral I as follows [51]:

$$I_k \approx \hat{I}_k = (2\pi)^{(N+1)/2} \left|-[H_{LP}]\right|^{-1/2} p_Y(\{Y_i\}|\tilde{\boldsymbol{\xi}}, \tilde{\epsilon}, \mathcal{G}_k) p_\xi(\tilde{\boldsymbol{\xi}}) p_\epsilon(\tilde{\epsilon}), \tag{31}$$

where N is the dimension of $\boldsymbol{\xi}$ and $[H_{LP}]$ is the Hessian matrix of the log-posterior distribution of model $\mathcal{G}_k$. The computation of the Hessian is made relatively straight forward thanks to the fact that we are using PC representations. Additional details regarding the computation of the Hessian are provided in Appendix A. Furthermore, the MAP point, $(\tilde{\boldsymbol{\xi}}, \tilde{\epsilon})$, can be obtained either by using an optimization approach or by using the MCMC algorithm to sample the posterior. In both cases, having a surrogate model provides significant computational savings. Moreover, if an optimization approach were to be used, one could take advantage of the availability of the exact PC Hessian to speed up the process. In this paper we use the sampled posterior to compute the MAP. Since several models are to be compared, we have to sample several posteriors which are expected to differ from a model to another. The sampling of the different models is however inexpensive, thanks to the availability of the surrogates, and it can be performed in parallel. More advanced sampling techniques could be considered, including the exploration of the models space in addition of the parameters space. This has not been considered in the present work as it was found easier to fit proposal parameters adapted to each model, than determining more general proposal parameters.

We have stressed the importance of evaluating the evidence at a low computational cost because this step would be performed by a diagnosis device with a relatively low computational power. Conversely, building an extensive dictionary of surrogates for different geometries can be costly, but this process is part of a pre-processing stage and can be done "offline" with a dedicated, more capable computer. Whence the evidences I_k have been approximated a Bayes factor analysis can be applied to rank the geometrical models, with the first rank corresponding to the model with the highest evidence.

4. Inference of Elastic Properties

In this section, we apply the methodology above to a simple example consisting of a rectangular solid with an inclusion. A uniform static pressure (magnitude $F = 1\text{Pa}$) is applied at the top of the solid while the bottom part is restrained from motion. A schematic representation of the setup is shown to scale in figure 2. We assume that the background and the inclusion are homogeneous materials, i.e. the Poisson ratio and Young's modulus are constant in each respective domain. Furthermore, we assume that the inclusion has a potentially higher Young's modulus than the background, whereas the Poisson ratio is the same in both regions. In addition, we set $L/H = 1$, $a_I/L = 1/2$, $b_I/H = 1/4$, and assume the inclusion is centered at $(x_c, y_c) = (0.65L, 0.65H)$. The problem is solved on a finite-element mesh with $T6$ elements, generated using the Triangle mesh generator [52] with the parameter $area = 0.001$ (which yields approximately 1,600 elements).

4.1. PC Surrogate

As stated earlier, it is essential to assess the fidelity of the PC surrogate model. To this end, we analyze the behavior of the error ϑ (defined in section 2.3) as function of p, the degree of the PC expansions of u, E, and ν. Specifically, we use (22) to analyze the convergence of the approximation as the polynomial orders p is increased. The error is estimated using a Monte Carlo approach with $M_s = 10,000$; a large number of samples was used to assure a high accuracy in the computation of the error.

Figure 3 shows a plots of ϑ versus p, for $\nu^{min} = 0.45$, $\nu^{max} = 0.49$, $E_B^{min} = 10$ kPa, $E_B^{max} = 90$ kPa, and three different values of κ^{max}. As expected, the relative error decreases as p increases. We notice that increasing κ^{max} results in a higher error for $p < 6$; this is because higher values of κ^{max} imply higher levels of uncertainty. Similarly, reducing (resp. increasing) the range for the Poisson's ratio and/or the Young's modulus would reduce (resp. increase) the error. For the present situation where only 3 parameters are involved in the PC expansion, an isotropic order is not an issue. So, unless otherwise specified, we shall use in the following $p = 6$, a value that ensures a relative error of approximately 1%.

4.2. Numerical Results

To assess the performance of the inference method, we generated a synthetic set of m_o observations of the local displacement field, $\{Y_i\} \equiv u(\boldsymbol{x}_i) + \eta_i, i = 1, \cdots, m_o$, where the displacement $u(x)$ is computed by solving (7) with $\nu = 0.48$, $E_B = 60$ kPa, $E_I = 240$ kPa, and the η_i's are independent, identically distributed, centered, Gaussian random variables with standard deviation $\epsilon = 0.03 \times u_{ref}$, where $u_{ref} = 2 \times F \times H/(E_B + E_I)$ is the deformation of a bar in uniaxial compression that behaves according to Hooke's law with modulus $(E_B + E_I)/2$. We scale the noise variance by the reference displacement so that the error is relative to the displacement's magnitude. Meanwhile, we constructed a surrogate model for $u(x, \boldsymbol{\xi})$ using $E_B^{min} = 10$ kPa, $E_B^{max} = 90$ kPa, $\nu^{min} = 0.45$, $\nu^{max} = 0.49$, $\kappa^{max} = 6$, and $p = 6$.

Figure 4 shows the marginal posterior distributions of ξ_1, ξ_2, ξ_3 and the hyper-parameter, ϵ. These posteriors were sampled using a Markov Chain Monte Carlo (MCMC) algorithm. Posterior distributions were then obtained from the chain samples using kernel density estimation [53]. We used a total of $m_o = 545$ observations (distributed according to the regular pattern described in section 4.2.3). We notice that the marginal posterior distributions are highly peaked and allow for very good estimates of the true values of the Poisson ration ν, Young moduli E_B, E_I and noise level ϵ to be obtained. Specifically, the calculated Maximum a Posteriori (MAP) estimates $\nu^{MAP} = 0.482$, $E_B^{MAP} = 60.1$ kPa, $E_I^{MAP} = 239.4$ kPa, and $\epsilon^{MAP} = 0.01$, are in close agreement with the true values.

From a diagnostic perspective, our main interest is to decide whether or not the value of the Young's modulus at the suspected inclusion is different from the value of the Young's modulus of the background, *i.e.* whether an anomaly is present. To this end, it is more effective to analyze, directly, the posterior distribution of $\rho = E_I/E_B$. We expect for this ratio to be one if the Young's modulus of the inclusion coincides with that of the background, and greater than one if they are different (because we assume that $E_I \geq E_B$). Consequently, the results below will focus on the posterior distribution of the contrast ratio $\rho = E_I/E_B$.

Note that the formulation of the proposed method accommodates naturally other priors, depending on the inference problem. For instance, one could relax the constrain $E_I \geq E_B$, using $E_I(\xi_1(\theta)) = [1/\kappa^{max} + (\kappa^{max} - 1/\kappa^{max})\xi_3(\theta)]E_B(\theta)$ with $\xi_3 \sim \mathcal{U}(0,1)$, and define $\rho = \max\{E_I/E_B, E_B/E_I\}$ as the diagnostic quantity. However, since the present work focuses on inferring inclusions stiffer than the background, we
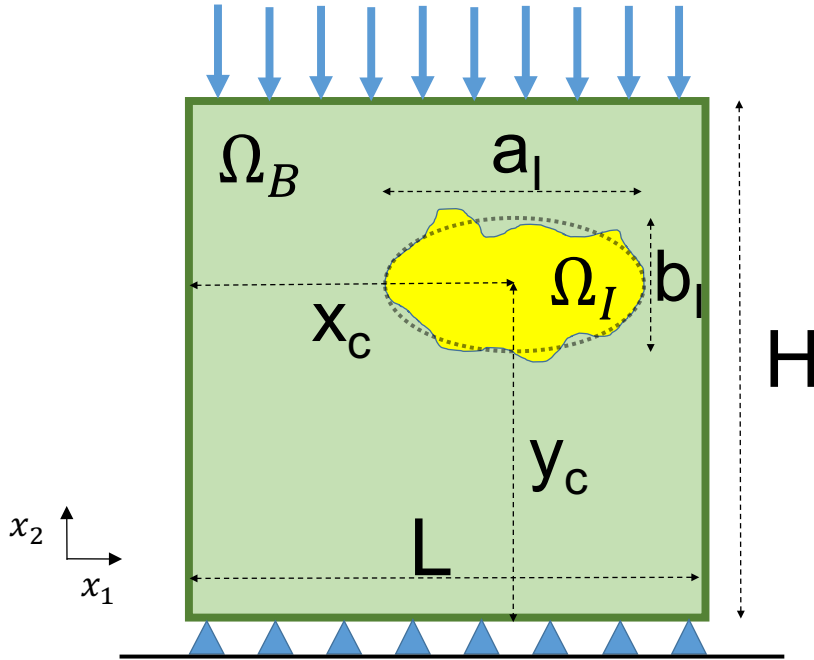

Figure 2: Schematic illustration of the computational domain, showing coordinate axes and geometry of inclusion.
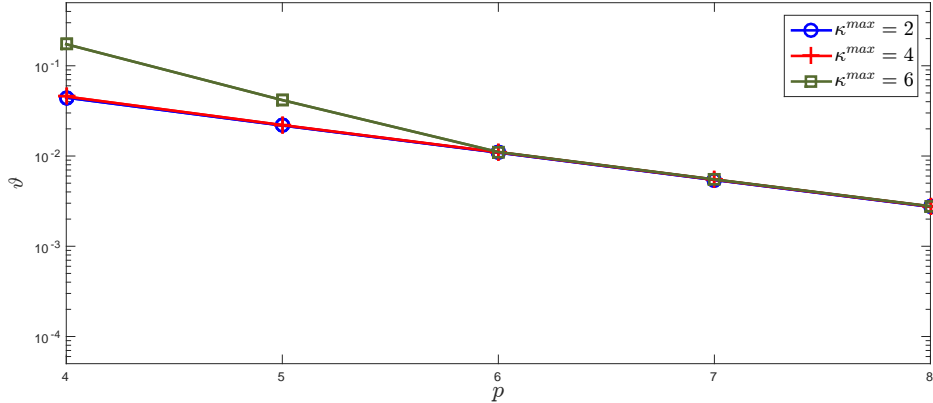
Figure 3: Error versus polynomial order for $(E_B^{min}, E_B^{max}) = (1e4, 9e4)$, $(\nu^{min}, \nu^{max}) = (0.45, 0.49)$, with curves generated for different values of κ^{max}, as indicated. Estimates are obtained using (22) with $M_s = 10,000$.

found it preferable to incorporate this information in the prior definition.

4.2.1. Effect of contrast

In the analysis above, we used a PC surrogate obtained for $\kappa^{max} = 6$, *i.e.* the prior anticipates that E_I falls in the range $E_B \leq E_I \leq 6E_B$. Using this PC surrogate, a synthetic inference exercise was conducted for $E_B = 60$ kPa and $E_I = 240$ kPa, showing satisfactory predictions. However, we expect the two extreme cases $E_I = E_B$ and $E_I = \kappa^{max} E_B$ to provide more challenging tests. The former is of great interest because it represents the case where there is no anomaly, and we want to be able to correctly infer this to avoid false-positives. The case where E_I/E_B is significantly larger than 1 is of lesser concern, because the presence of the inclusion will be easily ascertained; it is included nonetheless.

Figure 5 shows the inferred marginal distributions of $\rho = E_I/E_B$, for three different contrast values, generated with fixed $\nu = 0.48$ and $E_B = 60$ kPa, and $E_I = 60$, 240, and 360 kPa. Synthetic data are generated using the same procedure described above, and the PC surrogate constructed earlier is used for the purpose of sampling the posterior. In all cases, the MAP estimate of the contrast ratio is in good agreement with the true value used to generate the observations. Specifically, for the present tests, the situation involving absence of the anomaly is correctly detected with an essentially zero posterior probability for ratios $E_I/E_B \gtrsim 1.1$. Situations involving the presence of the anomaly are also correctly diagnosed, even for extreme values of the contrast ratio.

4.2.2. Effect of noise amplitude

One expects that the higher the noise level in the observations, the more challenging it would be to obtain accurate inference of the contrast ratio. To analyze the impact of the noise amplitude, we solved the inverse problem with $E_I/E_B = 4$ and different noise levels. Results are presented in Table 2, which provides estimates of the mean and standard deviation of the marginal posterior for the contrast ratio. We notice that, as expected, the standard deviation of the marginal distributions increases appreciably, roughly linearly, with the noise amplitude. On the other hand, the mean values (and the MAP estimates) are only weakly affected by the noise amplitude. Thus, for the present case and the noise range considered, the noise amplitude has insignificant impact on the diagnostic result.

4.2.3. Effects of observation density

To analyze how the inference is affected by the number and location of observations, we investigated the posterior distributions of the contrast ratio for different values of m_o. In addition, we obtained results for observations made at locations distributed according a regular pattern that is illustrated in figure 6. The pattern starts with an observation mesh with five points and the observation density is then systematically
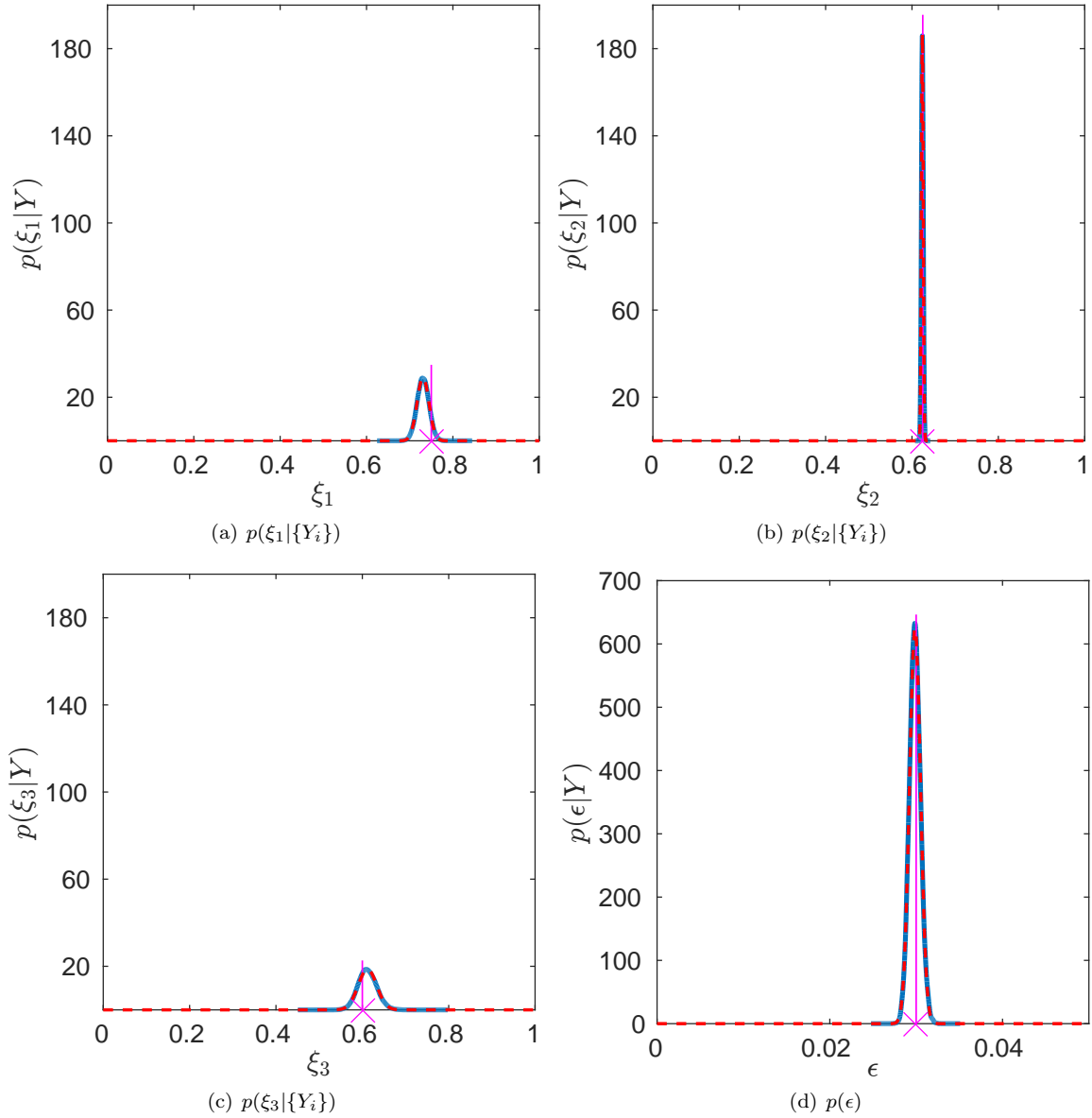
11

Figure 4: Posterior distributions. In the plots, the red dashed-line corresponds to a Gaussian fit of the distribution and the purple vertical line indicates the location of the true values of ξ_1, ξ_2, ξ_3, and ϵ, respectively.

increased by introducing new locations such that the minimum distance between observation points is maximized. The figure depicts the observation mesh corresponding to $m_o = 13$, 25 and 41.

Figure 4.2.3 shows the mean and standard deviation of the posterior density of E_I/E_B as a function of m_o. We notice that the mean fluctuates for small values of m_o, but that as we increase m_o the inferred mean starts to converge to a value close to 4 (the true value of E_I/E_B). Furthermore, we see that, as expected, the standard deviation rapidly decreases as m_o increases. It is very satisfactory that even with a very limited number of observations (*e.g.* 5), we can satisfactorily infer the presence of the inclusion.
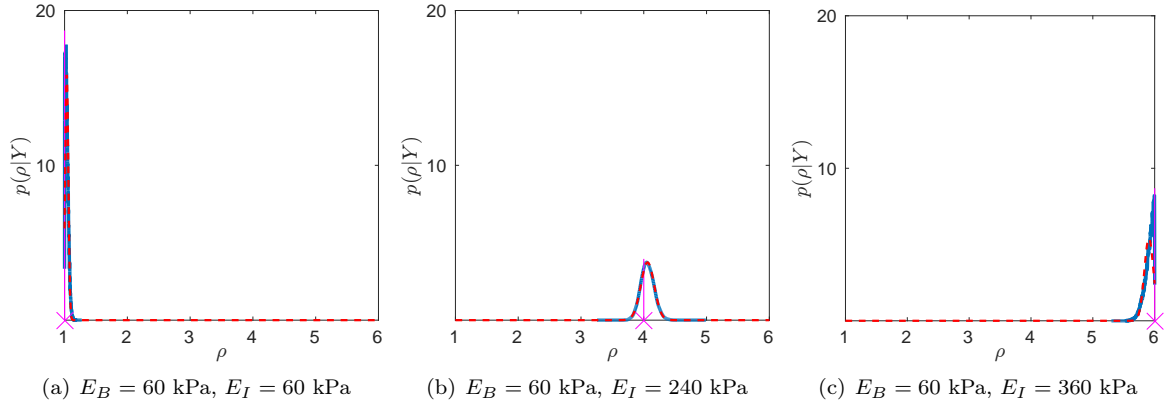
(a) $E_B = 60$ kPa, $E_I = 60$ kPa   (b) $E_B = 60$ kPa, $E_I = 240$ kPa   (c) $E_B = 60$ kPa, $E_I = 360$ kPa

Figure 5: Posterior distribution of $\rho = E_I/E_B$ (blue curves) for different values of E_I, with 1% noise. The dashed red curve corresponds to the Gaussian fit of the distributions, and the purple vertical line indicates the true value of ρ.

	$p(\rho\,|Y)$	
	μ	σ
$\epsilon = 0.5\%$	3.955	0.0169
$\epsilon = 1\%$	3.965	0.0333
$\epsilon = 3\%$	4.003	0.1021
$\epsilon = 5\%$	4.057	0.1773
$\epsilon = 10\%$	4.206	0.3837
$\epsilon = 20\%$	4.482	0.7048

Table 2: Comparison of inferred estimates of $\rho = E_I/E_B$ for different noise levels. The values μ and σ correspond to the mean and standard deviation of the posterior density of ρ.

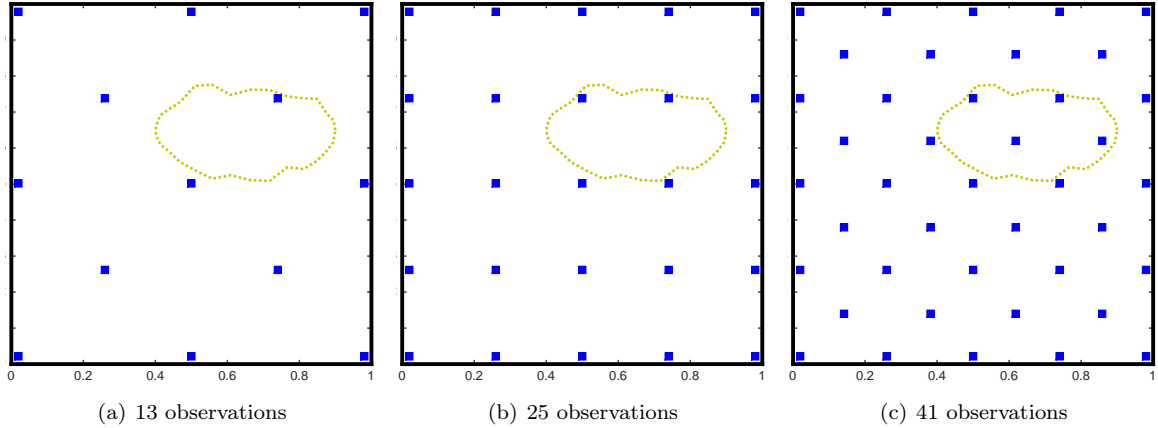(a) 13 observations   (b) 25 observations   (c) 41 observations

Figure 6: Schematic of the location of the observations. New points are selected following a regular pattern in which the distant to existing locations is maximized.

5. Model Selection

In this section we analyze the case when the actual geometry of the physical problem is not completely known (which is expected in most practical applications). To address this issue, we propose to compute
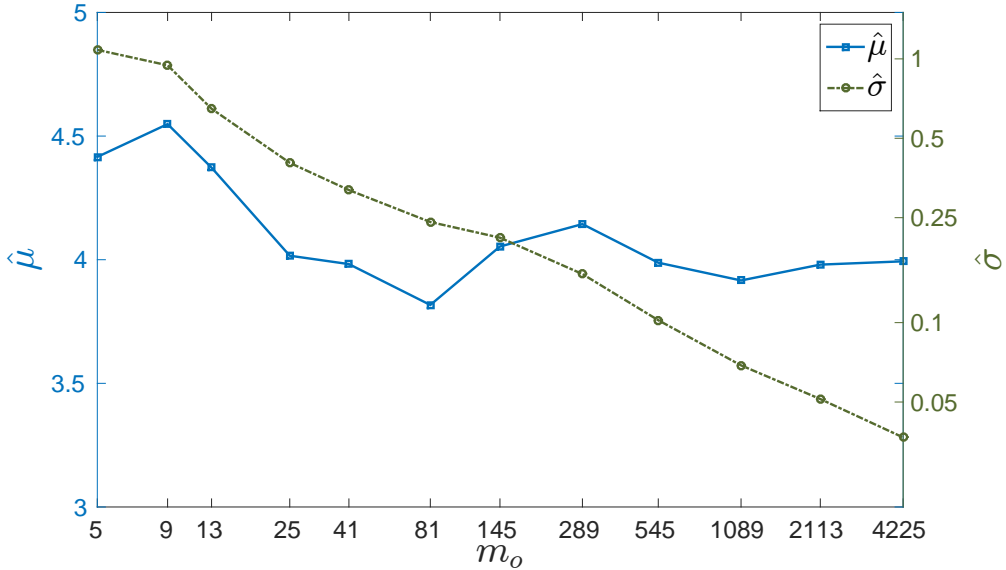
Figure 7: Mean and standard deviation of the posterior density of $\rho = E_I/E_B$ as a function of the number of observations, m_o.

not only one, but several surrogates using different geometrical models. As described in section 3.2, we use $p_Y(\{Y_i\}|\mathcal{G}_k)$, the evidence provided by the data, to rank the different models.

In the present work, a dictionary composed of 12 geometrical models is used. Schematics of the models are shown in figure 8. The models are ordered from left-to-right and top-to-bottom based on the value of the evidence; with the first model corresponding to the one with the highest evidence. In the figures, the area enclosed by the blue solid-line corresponds to the geometry of the inclusion assumed by the model, and the area enclosed by the yellow doted-line corresponds to the actual geometry of the physical model. Visually, we notice that, as expected, the highest ranked models are the ones in which the surrogate geometries agrees the most with the true geometry.

Figure 9 shows the normalized log-evidence (left plot) and also the MAP estimate of ρ (right plot) for each model. The evidence is normalized as follows:

$$\mathcal{E}_k = \frac{\log\left(p_Y(\{Y_i\}|\mathcal{G}_k)\right) - \min_j \log\left(p_Y(\{Y_i\}|\mathcal{G}_j)\right)}{\log\left(p_Y(\{Y_i\}|\mathcal{G}_0)\right) - \min_j \log\left(p_Y(\{Y_i\}|\mathcal{G}_j)\right)}, \tag{32}$$

where $p_Y(\{Y_i\}|\mathcal{G}_0)$ is the evidence given the true geometry. Since the true geometry is expected to provide the highest value for the evidence, the normalized log-evidence effectively falls in the range $[0, 1]$. We appreciate that the Evidence heavily favors models 1 through 7 (which is what one would expect from figure 8). From the MAP estimates of ρ shown in figure 9(b), we notice that for model $\mathcal{G}_1$ the estimated value is very close to the true value; this is not surprising, since $\mathcal{G}_1$ approximates well $\mathcal{G}_0$. For other models, we see some deviation in the estimate from the true value. The estimates provided by models $\mathcal{G}_3$ and $\mathcal{G}_5$ are higher than the true value because these geometries have a good overlap with the true inclusion, but they enclose a smaller area in the domain. Because of the smaller area the surrogate believes that the Young's modulus of the inclusion in higher than what it actually is. The converse is true for models $\mathcal{G}_2$, $\mathcal{G}_4$, $\mathcal{G}_6$, and $\mathcal{G}_7$. Using nested geometries might provide some means to get a closer estimate of ρ in this kind of setting.

Even though the inferred MAP of the contrast ratio are quite sensitive to geometrical errors, the predictions are quite robust in the sense that in all cases one correctly concludes that the domain is likely to contain an inclusion. In a diagnostic framework, this equates to having a significant probability of encountering an anomaly inside the inspected sample. In addition, the sensitivity to geometrical errors suggests that the inference of the inclusion geometry is feasible from the observations.

Note that even though our dictionary in this specific example consists exclusively of ellipses, the method actually allows for different types of geometries to be considered simultaneously. This provides a great
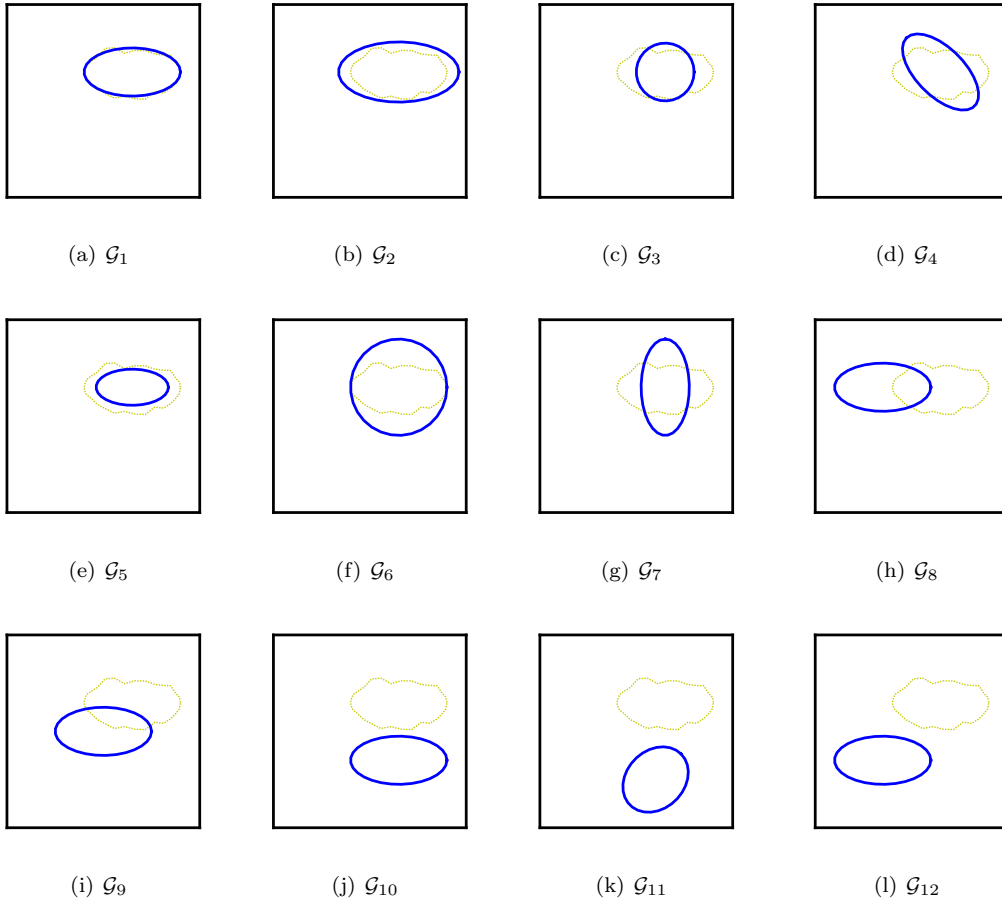
14

Figure 8: Schematics of 12 different geometrical configurations. In the figures the area enclosed by yellow doted-line correspond to the actual geometry of the inclusion, and the area enclosed by the blue solid-line corresponds to the location and shape of the inclusion assumed by the surrogate. The geometries are shown to scale. The models are ordered from left to right, top to bottom according to the value of the evidence.

deal of flexibility because one is not restricted to a single family of shapes, which is a limitation of some approaches where the geometry is parametrized by a small set of parameters. In fact, the geometries in the dictionary can be generated based on any prior information available regarding the shape and size of the inclusion (e.g. images of known existing tumors). Moreover, the method has shown to be robust to model errors, meaning that we can still detect the presence of an inclusion despite discrepancies between the true geometry and the geometries considered in the dictionary.

5.1. Non-homogeneous Young's modulus

Now, we analyze the impact of model errors on the robustness of the inference method. Specifically, we focus on the mechanical model errors arising due to a non-homogeneous distribution of the Young's modulus. So far, we have assumed that the background and the inclusion are both homogeneous. However, we now analyze how a surrogate model built assuming a homogeneous Young's moduli would perform when the observations correspond to a non-homogeneous distribution of the Young's modulus.

The aim of this analysis is to understand how tissue inhomogeneity would be interpreted by the model. In particular, in the case when there is not an inclusion, we want to know if inhomogeneities would be interpreted as the presence of an inclusion (i.e. a false-positive diagnostic). To this end, we consider observations
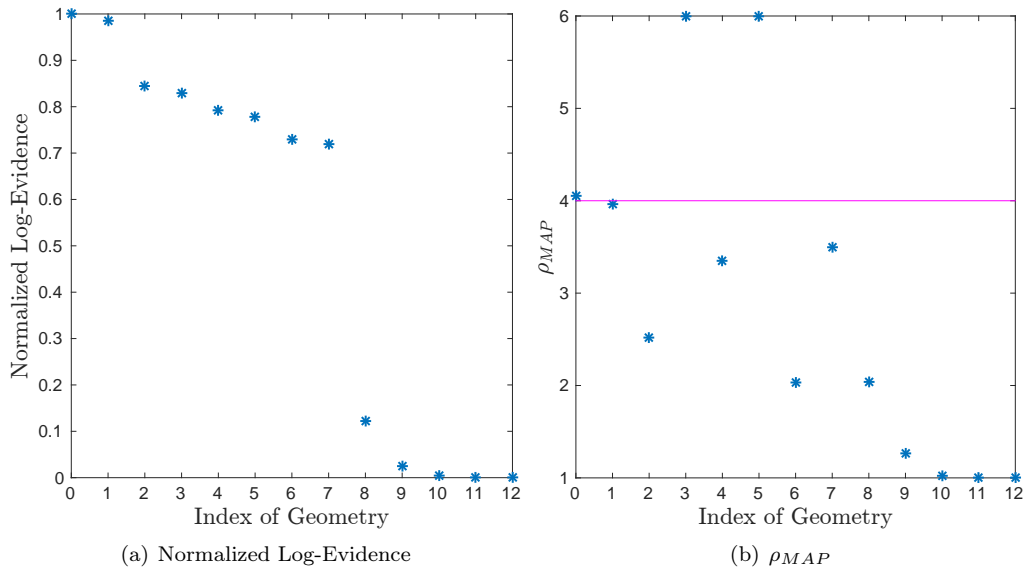
15

(a) Normalized Log-Evidence   (b) ρ_{MAP}

Figure 9: The left plot shows the normalized evidence of the models. The plot on the right show the MAP estimate of ρ for each model. The index of the model corresponds to the order of the geometries shown in figure 8 (index zero correspond to the true geometry). The purple line in the right plot corresponds to the true value of ρ.

generated with a non-homogeneous Young modulus field, $\tilde{E}(\boldsymbol{x}, \theta)$, that is constructed by perturbing the homogeneous field $E(\boldsymbol{x}) = \bar{E}$ with a realization of a Gaussian process, $K(\boldsymbol{x}, \theta)$, namely according to:

$$\tilde{E}(\boldsymbol{x}, \theta) = \bar{E} \exp\left[K(\boldsymbol{x}, \theta)\right] \tag{33}$$

where $K(\boldsymbol{x})$ is the centered Gaussian Process with correlation matrix

$$C(\boldsymbol{x}_1, \boldsymbol{x}_2) = \bar{\alpha}^2 \exp\left[-\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}{2l^2}\right].$$

The variance $\bar{\alpha}^2$ and the correlation length l are used to control the spatial variability of $\tilde{E}(\boldsymbol{x}, \theta)$. Moreover, we define $\bar{\rho} = (\mu_E + 3\sigma_E)/(\mu_E - 3\sigma_E) = (1 + 3\sqrt{e^{\bar{\alpha}^2} - 1})/(1 - 3\sqrt{e^{\bar{\alpha}^2} - 1})$, where μ_E and σ_E are the mean and variance of $\tilde{E}(\boldsymbol{x}, \theta)$ at a given fixed value of $\boldsymbol{x}$. This expression for $\bar{\rho}$ provides a sense of how the parameter $\bar{\alpha}$ impacts the variability of the Young's modulus field. Note that for $\bar{\alpha} = 0$ one recovers almost surely the homogeneous field $\tilde{E}(\boldsymbol{x}, \theta) = \bar{E}$.

For the analysis we generate perturbation fields for $l = 0.001, 0.1, 0.5, 1, 10$ (these values are relative to the dimension of the domain) and $\bar{\alpha} = 0.01, 0.05, 0.1$ (i.e. $\bar{\rho} \approx 1.062, 1.352, 1.860$). For each combination of these values we generated 30 realizations of the Young's modulus field.

For each realization of the field, we selected the best model according to the evidence and plotted the value of ρ_{MAP} in figure 10. First we notice that, as expected, the larger the value of $\bar{\alpha}$ the more our estimate for ρ deviates from the actual value. Moreover, we see that for large value of the correlation length the value of ρ_{MAP} remains close to one; this is because the field is highly correlated and we do not expect to see much variability. We also observe a similar behavior when l is very small; in this case the there is more fluctuation in the field but the fluctuations are small relative to the size of the suspected inclusion, and as a result they are "averaged-out". Finally, we see the largest deviations in the estimate when the correlation is about the size of the domain.

We can better understand the behavior observed in the previous figure by looking at some of the realizations of the field shown in figure 11. We see that for $l = 10$ the field is nearly constant, and as a result the inferred value for ρ is close to one as previously seen. For $l = 0.5$ we observe some high contrast regions

16

which are comparable in size to the inclusion's assumed geometries, which explains the higher values of ρ_{MAP}. Finally, for $l = 0.01$ we also see high contrast in the material, but the fluctuations are very small compared to the size of the assumed inclusions. Fluctuations like the ones depicted in figures 11(b) and 11(e) could in fact lead to a false-positive diagnostic, since they mimic the presence of an inclusion. However, if fluctuations like these are expected, the decision threshold for the value of ρ could be adjusted accordingly. Specific analysis about calculating the probabilities of miss-detection and false-positives, and decisions about the selection of an appropriate threshold for ρ are left for a future study and will be reported elsewhere.
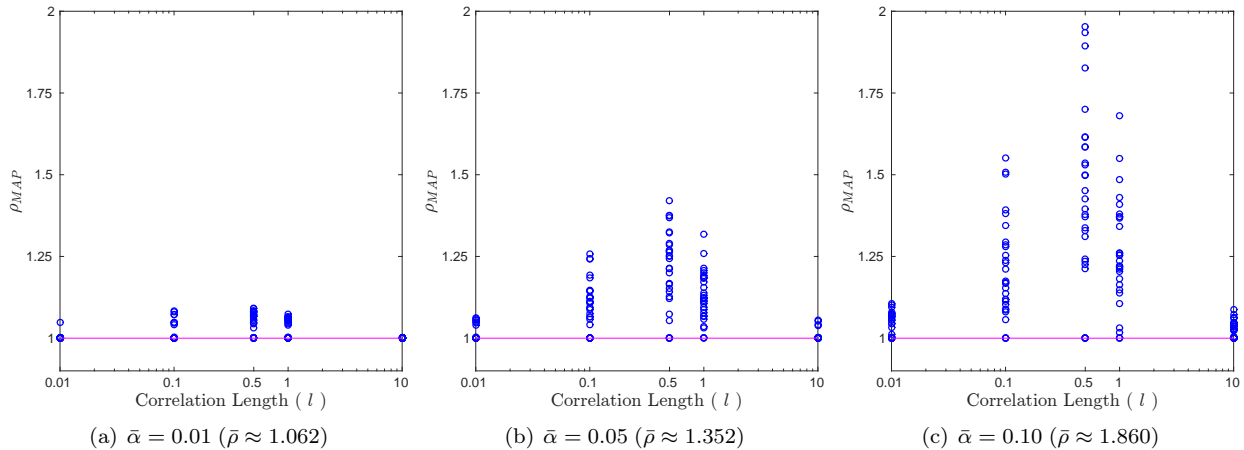

(a) $\bar{\alpha} = 0.01$ ($\bar{\rho} \approx 1.062$)   (b) $\bar{\alpha} = 0.05$ ($\bar{\rho} \approx 1.352$)   (c) $\bar{\alpha} = 0.10$ ($\bar{\rho} \approx 1.860$)

Figure 10: MAP estimate of the ratio ρ for the case of a non-homogeneous Young's modulus for different combinations of l and $\bar{\alpha}$. The MAP corresponds to the model with the highest Evidence. Each blue marker represent a different realization of the field. The purple line corresponds to a value of $\rho = 1$ that would be expected in the presence of a homogeneous field.

6. Conclusions

In this paper, we presented a method to quickly assess the presence of inclusion in soft tissue at a low computational cost. Specifically, our focus was to create a dictionary of surrogate models in order to determine the posterior probability distribution of the Young's modulus based on noisy observations of the displacement field. In order to reduce the cost of the inverse problem, we used a Polynomial Chaos expansion to construct a surrogate model that provides a faithful approximation of the forward problem. A Galerkin methodology was applied for this purpose, and the accuracy of the predictions was established using systematic refinement of the expansion order.

To handle the lack of complete information about the geometry of the physical model, a family of PC surrogates was built, each with a different assumed geometry. A model selection approach based on the evidence provided by the data was used to discriminate among the different models. The availability of the surrogate was exploited in the computation of the posterior's Hessian, which was used to estimate the value of the evidence. Furthermore, the Hessian was used compute the covariance of the proposal distribution in the the M-H algorithm, improving its convergence.

The proposed method was shown to be advantageous since it divides the computational cost into more costly preprocessing stage where the surrogate models are computed for several geometries, and a lower cost model selection and inference stage where the decision regarding the presence of the inclusion is made using observations. Thus, the approach makes feasible the use of a diagnostic device with a low computational cost to identify the presence of an inclusion in soft tissue

Implementation of the methodology was illustrated in a simplified setting, consisting of a two-dimensional rectangular solid containing an inclusion. Attention was focused on demonstrating the possibility of inferring the presence of the inclusion from observations of the displacement field. In the wide range of conditions
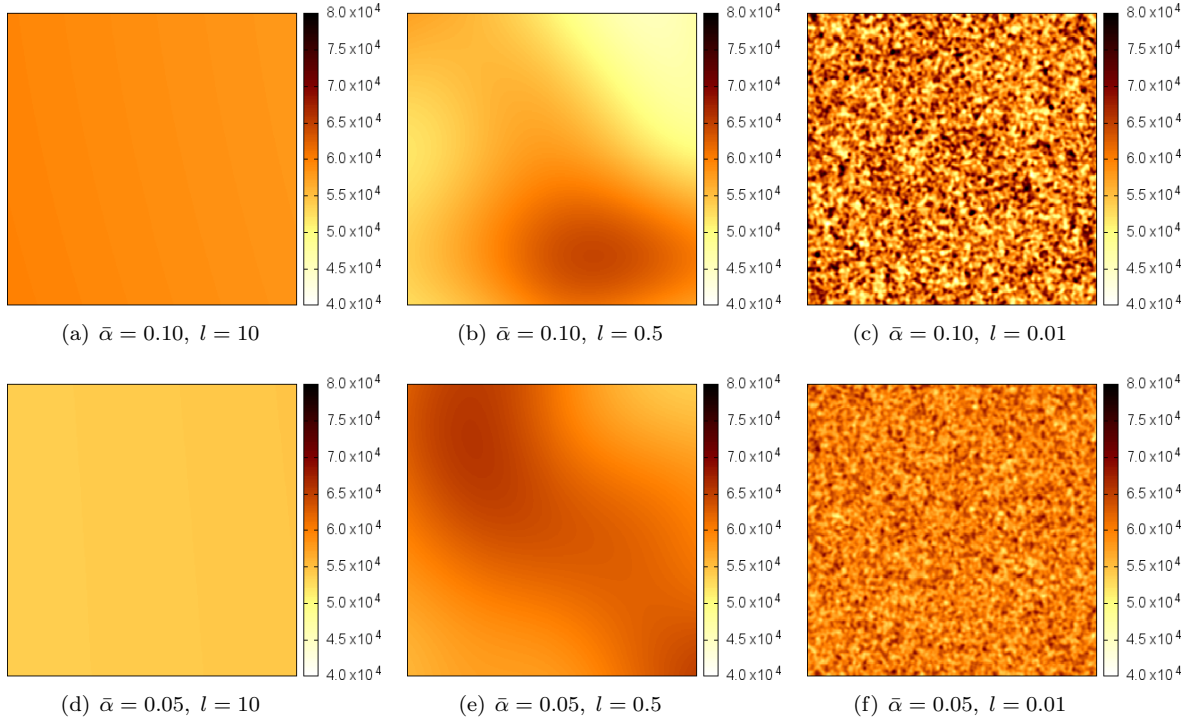
17

Figure 11: Selected realizations of the non-homogeneous Young's modulus field for different values of l and $\bar{\alpha}$ as depicted.

considered, the computations indicate that the approach is suitable for a wide range of contrast ratios, and that the predictions are robust to measurement noise, geometrical model errors, as well as the presence of spatial inhomogeneities.

The present developments motivate several avenues for further refinement. These include non-stationary problems, the possibility of simultaneously inferring both elastic problems and geometric features, as well as optimal selection of experimental observables. These topics will be the focus of future studies, and will be reported elsewhere.

Acknowledgments

Appendix A. MCMC Posterior Sampling

Given observations Y_i, the joint posterior distribution of $\boldsymbol{\xi}$ and ϵ can be sampled by Markov Chain Monte Carlo (MCMC) algorithms [54, 55] in the corresponding $(\boldsymbol{\xi}, \epsilon)$ domain. In this work, we use an adaptive Metropolis-Hastings (M-H) algorithm with a multivariate centered Gaussian proposal distribution. The approach consists in using a proposal distribution with a $(\boldsymbol{\xi}, \epsilon)$-dependent covariance matrix based on the decomposition of the Hessian of the log-posterior distribution.

From the expressions of the priors and likelihood function, the log-posterior distribution can be expressed as

$$LP(\boldsymbol{\xi}, \epsilon | \{Y_i\}) = -\sum_{i=1}^{m_o} \frac{(\hat{u}(\boldsymbol{x}_i, \boldsymbol{\xi}) - Y_i)^2}{2\epsilon^2} - (m_o + 1)\log(\epsilon) + \log(p_{\xi}(\boldsymbol{\xi})) + C. \qquad (A.1)$$

where C groups all the constant terms in the expression. Under mild conditions on p_{ξ} (recall that the component ξ_i of $\boldsymbol{\xi}$ are independent), the log-posterior can be differentiated twice, and we denote $[H_{LP}] \in \mathbb{R}^{(N+1)\times(N+1)}$ the Hessian matrix of the log-posterior distribution. Observing that $[H_{LP}]$ is symmetric, it can be decomposed as $[H_{LP}] = [V][\Lambda][V]^T$, where $[\Lambda]$ is a diagonal matrix whose elements are the real eigenvalues of $[H_{LP}]$, and $[V]$ is the matrix whose columns are the eigenvectors of $[H_{LP}]$ such that $[V][V]^T = [I]$. We then define the covariance matrix Σ^2 of the centered Gaussian proposal distribution according to:

$$\Sigma^2 = \sigma_{\Sigma}^2 \times [V]\mathcal{D}\left([\Lambda]\right)[V]^T, \qquad (A.2)$$

where the non-negative diagonal matrix $\mathcal{D}$ is any convenient transformation of $[\Lambda]$ and $\sigma_{\Sigma}^2 > 0$ is a scaling factor. A suitable choice for $\mathcal{D}$ is

$$\mathcal{D}_{ii}\left([\Lambda]\right) = \begin{cases} \frac{\min_k |\Lambda_{kk}|}{|\Lambda_{ii}|} & \Lambda_{ii} \neq 0 \\ 1 & \Lambda_{ii} = 0 \end{cases} \qquad (A.3)$$

This transformation allows us to take a larger step along the direction where the log-posterior is flat (the posterior varies slowly) and smaller steps along the directions of high curvature.

Note that in general the Hessian matrix $[H_{LP}]$ will be a function of the current state $(\boldsymbol{\xi}, \epsilon)$ of the chain. Thus, as the chain evolves from one state to another the Hessian matrix will change accordingly. However, in many cases the changes might not be significant, and it is possible to further reduce the computational cost by reducing the frequency at which the covariance matrix of the proposal distribution is updated, without significant performance degradation. Furthermore, in some instances we only need to compute the Hessian once to determine an adequate proposal distribution.

An advantage of using PC representations is that it is straightforward to compute derivatives [56, 57]. In particular one can derive expressions for the successive derivatives of $\hat{u}(\boldsymbol{x}, \boldsymbol{\xi})$ with respect to components ξ_i as needed for the computation of the Hessian. For instance, considering the partial derivative with respect to ξ_i of $\hat{u}(\boldsymbol{x}, \boldsymbol{\xi})$, we can write it as a PC expansion with the same basis used for $\hat{u}(\boldsymbol{x}, \boldsymbol{\xi})$:

$$\frac{\partial}{\partial \xi_i}\hat{u}(\boldsymbol{x}, \boldsymbol{\xi}) = \frac{\partial}{\partial \xi_i}\sum_{\alpha=0}^{P}\hat{u}_{\alpha}(\boldsymbol{x})\Psi_{\alpha}(\boldsymbol{\xi}) = \sum_{\alpha=0}^{P}(\partial_i \hat{u})_{\alpha}(\boldsymbol{x})\Psi_{\alpha}(\boldsymbol{\xi}), \qquad (A.4)$$

where the PC coefficients $(\partial_i \hat{u})_{\alpha}$ of the derivative are related to the PC coefficients of $\hat{u}$ through

$$\begin{Bmatrix} (\partial_i \hat{u})_0 \\ (\partial_i \hat{u})_1 \\ \vdots \\ (\partial_i \hat{u})_P \end{Bmatrix} = [L^i] \begin{Bmatrix} \hat{u}_0 \\ \hat{u}_1 \\ \vdots \\ \hat{u}_P \end{Bmatrix} = \begin{bmatrix} \left\langle \frac{\partial}{\partial \xi_i}\Psi_0, \Psi_0 \right\rangle & \left\langle \frac{\partial}{\partial \xi_i}\Psi_1, \Psi_0 \right\rangle & \cdots & \left\langle \frac{\partial}{\partial \xi_i}\Psi_P, \Psi_0 \right\rangle \\ \left\langle \frac{\partial}{\partial \xi_i}\Psi_0, \Psi_1 \right\rangle & \left\langle \frac{\partial}{\partial \xi_i}\Psi_1, \Psi_1 \right\rangle & \cdots & \left\langle \frac{\partial}{\partial \xi_i}\Psi_P, \Psi_1 \right\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \left\langle \frac{\partial}{\partial \xi_i}\Psi_0, \Psi_P \right\rangle & \left\langle \frac{\partial}{\partial \xi_i}\Psi_1, \Psi_P \right\rangle & \cdots & \left\langle \frac{\partial}{\partial \xi_i}\Psi_P, \Psi_P \right\rangle \end{bmatrix} \begin{Bmatrix} \hat{u}_0 \\ \hat{u}_1 \\ \vdots \\ \hat{u}_P \end{Bmatrix}. \qquad (A.5)$$

The derivative operators $[L^i]$, $i = 1, \cdots, N$ can be precomputed exactly and stored, so that when computing the Hessian in the MCMC algorithm, finding $\frac{\partial}{\partial \xi_i}\hat{u}$ reduces to a single matrix multiplication of $[L^i]$ with the coefficients vector $\hat{u}$. Higher-order derivatives amount to additional matrix multiplications. Consequently, the computational complexity of the adaptive proposal covariance essentially reduces to the spectral decomposition of the Hessian, which remains low provided that N is not too large. Finally, the scale factor σ_{Σ}^2 is classically tuned as to obtain a prescribed rejection rate for the chain.

References

[1] L. Kallivokas, A. Fathi, S. Kucukcoban, K. Stokoe II, J. Bielak, O. Ghattas, Site characterization using full waveform inversion, Soil Dynamics and Earthquake Engineering 47 (2013) 62–82.

[2] K. T. Tran, M. McVay, Site characterization using gauss–newton inversion of 2-d full seismic waveform in the time domain, Soil Dynamics and Earthquake Engineering 43 (2012) 16–24.

[3] F. Schoefs, H. Yáñez-Godoy, F. Lanata, Polynomial chaos representation for identification of mechanical characteristics of instrumented structures, Computer-Aided Civil and Infrastructure Engineering 26 (3) (2011) 173–189.

[4] A. P. Adewuyi, Z. Wu, Vibration-based damage localization in flexural structures using normalized modal macrostrain techniques from limited measurements, Computer-Aided Civil and Infrastructure Engineering 26 (3) (2011) 154–172.

[5] R. Jafarkhani, S. F. Masri, Finite element model updating using evolutionary strategy for damage detection, Computer-Aided Civil and Infrastructure Engineering 26 (3) (2011) 207–224.

[6] M. A. Aguilo, W. Aquino, J. C. Brigham, M. Fatemi, An inverse problem approach for elasticity imaging through vibroacoustics, IEEE Transactions on Medical Imaging 29 (4) (2010) 1012–1021.

[7] E. Crossen, M. S. Gockenbach, B. Jadamba, A. A. Khan, B. Winkler, An equation error approach for the elasticity imaging inverse problem for predicting tumor location, Computers & Mathematics with Applications 67 (1) (2014) 122–135.

[8] A. A. Oberai, N. H. Gokhale, G. R. Feijóo, Solution of inverse problems in elasticity imaging using the adjoint method, Inverse Problems 19 (2) (2003) 297.

[9] L. Sandrin, M. Tanter, J.-L. Gennisson, S. Catheline, M. Fink, Shear elasticity probe for soft tissues with 1-d transient elastography, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 49 (4) (2002) 436–446.

[10] R. Sinkus, J. Bercoff, M. Tanter, J.-L. Gennisson, C. El Khoury, V. Servois, A. Tardivon, M. Fink, Nonlinear viscoelastic properties of tissue assessed by ultrasound, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 53 (11) (2006) 2009–2018.

[11] M. Fatemi, J. F. Greenleaf, Vibro-acoustography: An imaging modality based on ultrasound-stimulated acoustic emission, Proceedings of the National Academy of Sciences 96 (12) (1999) 6603–6608.

[12] B. Banerjee, T. F. Walsh, W. Aquino, M. Bonnet, Large scale parameter estimation problems in frequency-domain elastodynamics using an error in constitutive equation functional, Computer Methods in Applied Mechanics and Engineering 253 (2013) 60–72.

[13] K. Raghavan, A. E. Yagle, Forward and inverse problems in elasticity imaging of soft tissues, IEEE Transactions on Nuclear Science 41 (4) (1994) 1639–1648.

[14] C. Sumi, A. Suzuki, K. Nakayama, Estimation of shear modulus distribution in soft tissue from strain distribution, IEEE Transactions on Biomedical Engineering 42 (2) (1995) 193–202.

[15] D. B. Plewes, J. Bishop, A. Samani, J. Sciarretta, Visualization and quantification of breast cancer biomechanical properties with magnetic resonance elastography, Physics in Medicine and Biology 45 (6) (2000) 1591.

[16] A. Emery, A. V. Nenarokomov, T. D. Fadale, Uncertainties in parameter estimation: the optimal experiment design, International Journal of Heat and Mass Transfer 43 (18) (2000) 3331–3339.

[17] V. A. B. Narayanan, N. Zabaras, Stochastic inverse heat conduction using a spectral approach, International Journal for Numerical Methods in Engineering 60 (9) (2004) 1569–1593.

[18] B. Jin, J. Zou, Inversion of Robin coefficient by a spectral stochastic finite element approach, Journal of computational Physics 227 (6) (2008) 3282–3306.

[19] N. Zabaras, B. Ganapathysubramanian, A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach, Journal of Computational Physics 227 (9) (2008) 4697–4735.

[20] I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM review 52 (2) (2010) 317–355.

[21] J. E. Warner, W. Aquino, M. D. Grigoriu, Stochastic reduced order models for inverse problems under uncertainty, Computer Methods in Applied Mechanics and Engineering 285 (2015) 488–514.

[22] L. Tenorio, Statistical regularization of inverse problems, SIAM review 43 (2) (2001) 347–366.

[23] A. Mohammad-Djafari, Bayesian inference for inverse problems, in: AIP Conference Proceedings: Bayesian inference and Maximum Entropy Methods in Science and Engineering, Vol. 617, 2002, pp. 477–496.

[24] P.-S. Koutsourelakis, A novel Bayesian strategy for the identification of spatially varying material properties and model validation: an application to static elastography, International Journal for Numerical Methods in Engineering 91 (3) (2012) 249–268.

[25] M. A. Iglesias, K. Lin, A. M. Stuart, Well-posed Bayesian geometric inverse problems arising in subsurface flow, Inverse Problems 30 (11) (2014) 114001.

[26] Y. Efendiev, T. Hou, W. Luo, Preconditioning Markov chain Monte Carlo simulations using coarse-scale models, SIAM Journal on Scientific Computing 28 (2) (2006) 776–803.

[27] Y. Efendiev, A. Datta-Gupta, V. Ginting, X. Ma, B. Mallick, An efficient two-stage Markov chain Monte Carlo method for dynamic data integration, Water Resources Research 41 (12) (2005) 1–6.

[28] J. A. Christen, C. Fox, Markov chain Monte Carlo using an approximation, Journal of Computational and Graphical Statistics 14 (4) (2005) 795–810.

[29] J. Wang, N. Zabaras, Using Bayesian statistics in the estimation of heat source in radiation, International Journal of Heat and Mass Transfer 48 (1) (2005) 15–29.

[30] B. Jin, Fast Bayesian approach for parameter estimation, International Journal for Numerical Methods in Engineering 76 (2) (2008) 230–252.

[31] X. Ma, N. Zabaras, An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method, Inverse Problems 25 (3) (2009) 035013.

[32] S. Balakrishnan, A. Roy, M. G. Ierapetritou, G. P. Flach, P. G. Georgopoulos, Uncertainty reduction and characterization for complex environmental fate and transport models: An empirical Bayesian framework incorporating the stochastic response surface method, Water Resources Research 39 (12) (2003) 1–13.

[33] Y. Marzouk, D. Xiu, A stochastic collocation approach to Bayesian inference in inverse problems, Journal of Computational Physics 6 (4) (2009) 826–847.

[34] Y. M. Marzouk, H. N. Najm, L. A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, Journal of Computational Physics 224 (2) (2007) 560–586.

[35] B. Faverjon, R. Ghanem, Stochastic inversion in acoustic scattering, The Journal of the Acoustical Society of America 119 (6) (2006) 3577–3588.

[36] R. G. Ghanem, P. D. Spanos, Stochastic finite elements: a spectral approach, Springer, 1991.

[37] O. P. Le Maître, O. M. Knio, Spectral methods for uncertainty quantification: with applications to computational fluid dynamics, Scientific Computation, Springer, 2010.

[38] W. S. Slaughter, The linearized theory of elasticity, Birkhauser, 2002.

[39] B. D. Reddy, A. T. McBride, Introduction to finite element analysis and recent developments, in: A. Patnaik, R. Anandjiwala (Eds.), Modeling and Simulation in Fibrous Materials: Techniques and Applications, Nova Science Publishers, New York, 2012, pp. 1–28.

[40] M. Pellissetti, R. Ghanem, Iterative solution of systems of linear equations arising in the context of stochastic finite elements, Adv. Eng. Softw. 31 (2000) 607–616.

[41] O. Le Maître, O. Knio, B. Debusschere, H. Najm, R. Ghanem, A multigrid solver for two-dimensional stochastic diffusion equations, Computer Methods in Applied Mechanics and Engineering 192 (2003) 4723–4744.

[42] T. Glozman, H. Azhari, A method for characterization of tissue elastic properties combining ultrasonic computed tomography with elastography, Journal of Ultrasound in Medicine 29 (3) (2010) 387–398.

[43] E. J. Chen, J. Novakofski, W. K. Jenkins, W. O'Brien, Young's modulus measurements of soft tissues with application to elasticity imaging, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 43 (1) (1996) 191–194.

[44] J. F. Greenleaf, M. Fatemi, M. Insana, Selected methods for imaging elastic properties of biological tissues, Annual Review of Biomedical Engineering 5 (1) (2003) 57–78.

[45] T. P. Harrigan, E. E. Konofagou, Estimation of material elastic moduli in elastography: a local method, and an investigation of Poisson's ratio sensitivity, Journal of Biomechanics 37 (8) (2004) 1215–1221.

[46] D. Higdon, H. Lee, C. Holloman, Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems, Bayesian Statistics 7 (2003) 181–197.

[47] A. Mohammad-Djafari, On the estimation of hyperparameters in Bayesian approach of solving inverse problems, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, 1993, pp. 495–498.

[48] A. Mohammad-Djafari, A full Bayesian approach for inverse problems, in: Maximum Entropy and Bayesian Methods, Springer, 1996, pp. 135–144.

[49] J. Bercoff, M. Pernot, M. Tanter, M. Fink, Monitoring thermally-induced lesions with supersonic shear imaging, Ultrasonic Imaging 26 (2) (2004) 71–84.

[50] K. Parker, S. Huang, R. Musulin, R. Lerner, Tissue response to mechanical vibrations for sonoelasticity imaging, Ultrasound in Medicine & Biology 16 (3) (1990) 241–246.

[51] R. E. Kass, A. E. Raftery, Bayes factors, Journal of the American Statistical Association 90 (430) (1995) 773–795.

[52] J. R. Shewchuk, Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator, in: M. C. Lin, D. Manocha (Eds.), Applied Computational Geometry: Towards Geometric Engineering, Vol. 1148 of Lecture Notes in Computer Science, Springer-Verlag, 1996, pp. 203–222, from the First ACM Workshop on Applied Computational Geometry.

[53] Z. Botev, J. Grotowski, D. Kroese, Kernel density estimation via diffusion, The Annals of Statistics 38 (5) (2010) 2916–2957.

[54] L. Tierney, Markov chains for exploring posterior distributions, The Annals of Statistics (1994) 1701–1728.

[55] J. Besag, P. Green, D. Higdon, K. Mengersen, Bayesian computation and stochastic systems, Statistical science (1995) 3–41.

[56] A. Alexanderian, J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W. Thacker, O. Knio, Global sensitivity analysis in ocean global circulation models: a sparse spectral projection approach, Comput. Geosci. 16 (2012) 757–778.

[57] I. Sraj, M. Iskandarani, C. Thacker, A. Srinivasan, O. Knio, Drag parameter estimation using gradients and Hessian from a polynomial chaos model surrogate, Monthly Weather Review 142 (2014) 933–941.