International Journal for Uncertainty Quantification, 3 (4): 289–319 (2013)

AN OVERVIEW OF INVERSE MATERIAL IDENTIFICATION WITHIN THE FRAMEWORKS OF DETERMINISTIC AND STOCHASTIC PARAMETER ESTIMATION

Miguel A. Aguiló, Laura Swiler, & Angel Urbina*

Optimization and Uncertainty Quantification, Sandia National Laboratories, P.O. Box 5800, MS 1318, Albuquerque, New Mexico 87185-1320, USA

Original Manuscript Submitted: 04/25/2012; Final Draft Received: 09/15/2012

This work investigates the problem of parameter estimation within the frameworks of deterministic and stochastic parameter estimation methods. For the deterministic methods, we look at constrained and unconstrained optimization approaches. For the constrained optimization approaches we study three different formulations: L^2, error in constitutive equation method (ECE), and the modified error in constitutive equation (MECE) method. We investigate these formulations in the context of both Tikhonov and total variation (TV) regularization. The constrained optimization approaches are compared with an unconstrained nonlinear least-squares (NLLS) approach. In the least-squares framework we investigate three different formulations: standard, MECE, and ECE. With the stochastic methods, we first investigate Bayesian calibration, where we use Monte Carlo Markov chain (MCMC) methods to calculate the posterior parameter estimates. For the Bayesian methods, we investigate the use of a standard likelihood function, a likelihood function that incorporates MECE, and a likelihood function that incorporates ECE. Furthermore, we investigate the maximum a posteriori (MAP) approach. In the MAP approach, parameters' full posterior distribution are not generated via sampling; however, parameter point estimates are computed by searching for the values that maximize the parameters' posterior distribution. Finally, to achieve dimension reduction in both the MCMC and NLLS approaches, we approximate the parameter field with radial basis functions (RBF). This transforms the parameter estimation problem into one of determining the governing parameters for the RBF.

KEY WORDS: *inverse problems, Bayesian calibration, maximum a posteriori estimate, error in constitutive equation, nonlinear least squares, regularization*

1. INTRODUCTION

This paper investigates the problem of parameter estimation (also referred to as calibration or parameter identification) in the context of a particular problem: given observed displacement field(s) on a beam, determine the elastic modulus of a beam material. The approaches fall in two main classes: deterministic parameter estimation methods, which yield single-value point estimates of the parameters, and stochastic parameter estimation methods, which give some type of uncertainty about an optimal parameter value. For the deterministic methods, we investigate constrained and unconstrained optimization approaches. In the context of constrained optimization of a cost functional, we investigate three different first-order reduced-space formulations. Specifically, we look at the L^2 formulation [1, 2], the error in constitutive equation formulation [3–12], and the modified error in constitutive equation formulation [13–15]. Furthermore, these formulations are investigated in the context of both Tikhonov and total variation (TV) regularization. We compare the constrained optimization approaches, which minimize a specific cost functional plus a regularization operator, with a general unconstrained nonlinear least-squares approach (NLLS) [16]. The formulations considered

*Correspond to Miguel A. Aguiló, E-mail: maguilo@sandia.gov, URL: http://www.sandia.gov/

DOI: 10.1615/Int.J.UncertaintyQuantification.2012003668

for the solution of the unconstrained NLLS problem are the standard, modified error in constitutive equation (MECE), and error in constitutive equation (ECE) formulations. Radial basis functions are used to represent the elastic modulus field, transforming the parameter estimation problem into one of determining the governing parameters for the radial basis functions (RBF).

Reduced-space methods [17–19] have the advantage that only two linearized forward problems, state and adjoint problems, are solved at each optimization iteration. For a moderate number of parameters, the solution of the forward problems dominates the computational cost of an optimization iteration. Thus, when robust parallel solvers are available, this approach inherits the parallel efficiency and scalability of the partial differential equation (PDE) solver [20]. Moreover, if a first-order formulation is used, there is no need to derive second derivatives of the cost functional, at the expense of a superlinear rate of convergence [21]. However, the convergence of first-order reduced-space formulations often deteriorates as the number of parameters increases. As a result, such an approach often exhibits poor algorithmic scalability with respect to the number of parameters [22]. On the contrary, full-space methods [23, 24] are quadratically convergent close to the local solution and can be made to exploit the structure of the resulting Karush-Kuhn-Tucker (KKT) system in order to take advantage of existing parallel solvers. Furthermore, full-space methods are independent of the number of parameters. However, these methods require the calculation of the second derivatives of the cost functional. Moreover, the resulting KKT system can be ill-conditioned; thus, preconditioners are needed to improve the effectiveness of full-space methods. The question of how to precondition the KKT system is an important challenge and remains an active area of research.

In the context of Bayesian calibration, which requires a prior distribution on the parameters and uses the experimental observations to update the prior and obtain "data-informed" posterior parameter distribution, Markov chain Monte Carlo (MCMC) methods [25, 26] were used to calculate the posterior parameter estimates. MCMC methods are widely used Bayesian methods due to their simplicity of coding and global convergence capability. However, MCMC methods tend to suffer from the curse of dimensionality. In this study, we investigate three different likelihood functions. Specifically, we look at the standard likelihood function, a likelihood function that incorporates ECE, and a likelihood function that incorporates MECE. The other stochastic method investigated was the maximum *a posteriori* (MAP) estimate [27–29]. In the MAP approach, we do not generate a full posterior distribution via sampling, but we determine the mode of the parameters' posterior distributions using the constrained deterministic optimization approaches study herein. These mode estimates can be computed by searching for the parameters that maximize the probability of the posterior distribution.

Overall, the set of methods described above allowed us to perform a comprehensive study and comparison of the different parameter estimation approaches. In addition to the analysis of the different methods, our studies represent specific contributions to the field of parameter estimation:

1. The MECE equations are derived for the elastostatic boundary value problem.

2. An iterative two-step approach for solving the ECE problem is presented (the alternating directions approach).

3. The ECE and MECE are incorporated directly into the Bayesian likelihood formulation. Our study shows that the ECE likelihood formulation provides a regularization behavior and improves the convergence of the Markov chain posterior.

4. The ECE and MECE are also incorporated into the MAP, where the ECE formulation provides improvements in the determination of the uncertainty in the parameter values.

This paper is organized as follows: Section 1 provides some background on the beam displacement problem and the problem formulation we are using. Section 2 presents the deterministic calibration approaches being used in this paper. The first three methods presented in Section 2 refer to approaches for constrained optimization problems, while the last method refers to an approach for unconstrained optimization problems. Section 3 presents the stochastic calibration approaches being used in this paper. Section 4 presents the results with discussion, and Section 5 provides conclusions.

1.1 Problem Overview

A prototypical inverse problem in linear elastostatics consists of predicting the elastic modulus field E from measurements u^m related to the solution u of the following boundary value problem:

$$\begin{aligned} \nabla \cdot \sigma &= f \quad \text{in} \quad D, \\ u &= 0 \quad \text{on} \quad D_u, \\ \sigma \cdot n &= \tau \quad \text{on} \quad D_\tau, \end{aligned} \tag{1}$$

where the stress tensor σ is given by $\sigma = E\nabla u$. In Eq. (1), $D \subseteq \mathbb{R}^d$ is the computational domain with boundary ∂D. The regions where displacements and surface tractions are applied are denoted by $D_u \subset \partial D$ and $D_\tau \subset \partial D$, respectively. Finally, u denotes displacement, f is the volume force, τ is the surface traction, n is the outward-pointing unit normal to D_τ, and superscript d denotes spatial dimension.

1.2 Material Representation

The physical characteristics of a material will determine how its properties are represented in the context of a parameter estimation problem. For instance, in homogeneous materials the elastic modulus is assumed constant. However, this assumption is not valid for heterogeneous materials. In this work, two distinct material representations were considered to characterize the elastic modulus field. First, the elastic modulus was assumed to vary pointwise (i.e., the elastic modulus was estimated at each point in space). Second, the elastic modulus was represented as an expansion of N_r radially symmetric Gaussian radial basis functions as follows:

$$E = E_c + \sum_{i=1}^{N_r} \widehat{E}_i \varphi(\|x - \widehat{x}_i\|),$$

where $\widehat{E}_i \in \mathbb{R}^{N_r}$ are real-valued coefficients, $E_c \in \mathbb{R}$ is the matrix elastic modulus, and $\widehat{x} \in \mathbb{R}^{N_r \times d}$ are the coordinates of the Gaussian basis functions.

The Gaussian basis function is defined as

$$\varphi(\|x - \widehat{x}_i\|) = \exp\left(-\frac{\|x - \widehat{x}_i\|^2}{c_i^2}\right),$$

where $c \in \mathbb{R}^{N_r}$ are locality parameters that control the radius of influence of the Gaussian basis function. For normalization purpose, the locality parameters were represented as $c_i = 10^{z_i}$, where each $z_i \in \mathbb{R}^{N_r}$ was estimated through the optimization scheme.

Given the elastic modulus at the center of the RBF, the coefficients $\widehat{E}_i$ were computed using the following interpolation scheme:

$$\Upsilon^{-1}\Theta = \widehat{E},$$

where

$$\Upsilon = \sum_{i=1}^{N_r}\sum_{j=1}^{N_r} \varphi\big(\|\widehat{x}_i - \widehat{x}_j\|^2\big),$$

$$\Theta = \sum_{j=1}^{R} E(\widehat{x}_j) - E_c.$$

The reader is referred to [30, 31] for more details on the interpolation scheme implemented herein.

2. DETERMINISTIC PARAMETER ESTIMATION METHODS

For the constrained optimization problem, a first-order reduced-space approach was used. The advantage of this approach is that only first derivatives of the cost functional and constraints are required. Furthermore, regardless of the number of unknown parameters, only two linearized forward problems are solved at each optimization iteration to compute the gradient of the cost functional. The main disadvantage of deterministic optimization methods is that these methods ensure just local extrema.

The discretization necessary for the solution of the constrained optimization problem can occur in two different manners. In the first, called "optimize-then-discretize," one can show that if the states and controls are a local solution to the nonlinear programming problem, then there exist Lagrange multipliers such that the KKT or first-order necessary optimality conditions are satisfied at the local solution [16]. Thus, under sufficient smoothness assumptions, one obtains a KKT or optimality system. Finally, one generates a mesh on the computational domain and proceeds to discretize the KKT system.

In the second approach, called "discretize-then-optimize," one generates a mesh on the computational domain and proceeds to discretize the nonlinear programming problem. If the states and controls are a local solution to the discrete nonlinear programming problem, there exist Lagrange multipliers such that the discrete KKT conditions are satisfied at the local solution, provided that some regularity conditions are met. Note that the discrete KKT systems obtained with both approaches are not necessarily the same [24, 32].

In this work we use an optimize-then-discretize approach. One has to be careful when such approach is used because the resulting optimality system can be nonsymmetric. This can lead to inexact gradients, which affects the convergence rate of gradient-based optimization algorithms. There is no general consensus on which approach should be preferred: it depends on the application and computational resources available to tackle the nonlinear programming problem [33, 34].

2.1 L^2 Formulation

We consider the nonlinear programming problem

$$\begin{aligned} &\arg \underset{\{u,E\}\in\, \mathcal{U}\times\mathcal{E}}{\text{minimize}} \quad \frac{\beta}{2}\langle u-u^m, u-u^m\rangle + R(E) \\ &\text{subject to} \\ &\qquad \nabla\cdot(E\nabla u) = f \quad \text{in} \quad D, \\ &\qquad (E\nabla u)\cdot n = \tau \quad \text{on} \quad D_\tau, \end{aligned} \tag{2}$$

where $\mathcal{U} = \{u\colon u \in H^1(D),\ u = 0 \text{ on } D_u\}$, $\mathcal{E} = \{E\colon E \in L^2(D),\ E > 0\}$, $R(\cdot)$ is the regularization operator, and $\beta > 0$ is a given parameter. For simplicity of implementation, we restrict the search space of the elastic modulus to $H^1(D)$.

The Lagrangian functional associated with the minimization problem defined in Eq. (2) is given by

$$\mathcal{L}(u,E,\lambda) = \frac{\beta}{2}\langle u-u^m, u-u^m\rangle + R(E) + \langle \lambda, \nabla\cdot(E\nabla u) - f\rangle + \langle \Pi(\lambda), \tau - (E\nabla u)\cdot n\rangle, \tag{3}$$

where $\Pi(\lambda)\colon H^1(D) \to L^2(D_\tau)$. If $\{u^*, E^*\}$ is a local solution of Eq. (2), there exists a Lagrange multiplier $\lambda \in \mathcal{U}$ such that the KKT conditions hold at the stationary point $\{u^*, E^*, \lambda\}$, i.e.,

$$D\mathcal{L}(u,E,\lambda)(\delta u, \delta E, \delta\lambda) = D_u\mathcal{L}\cdot\delta u + D_E\mathcal{L}\cdot\delta E + D_\lambda\mathcal{L}\cdot\delta\lambda = 0,$$

where

$$D_u\mathcal{L}\cdot\delta u = \beta\langle u-u^m, \delta u\rangle + \langle\nabla\cdot(E\nabla\lambda), \delta u\rangle = 0, \tag{4}$$

$$D_E\mathcal{L}\cdot\delta E = \langle\nabla R(E), \delta E\rangle - \langle\nabla\lambda\cdot\nabla u, \delta E\rangle = 0, \tag{5}$$

$$D_\lambda\mathcal{L}\cdot\delta\lambda = \langle\nabla\cdot(E\nabla u) - f, \delta\lambda\rangle + \langle\Pi(\delta\lambda), \tau - (E\nabla u)\cdot n\rangle_{D_\tau} = 0. \tag{6}$$

This allows us to reformulate the nonlinear programming problem in Eq. (2) into

$$\arg \underset{E\in\mathcal{E}}{\text{minimize}} \quad \frac{\beta}{2}\langle u(E)-u^m, u(E)-u^m\rangle + R(E),$$

where the gradient operator is given by

$$D_E\mathcal{L} = \nabla R(E) - \nabla\lambda(E)\cdot\nabla u(E).$$

Here, $u(E)$ is the solution to the boundary value problem

$$\begin{aligned} \nabla\cdot(E\nabla u) &= f \quad \text{in} \quad D,\\ (E\nabla u)\cdot n &= \tau \quad \text{on} \quad D_\tau, \end{aligned}$$

and $\lambda(E)$ is the solution to the adjoint problem

$$\nabla\cdot(E\nabla\lambda) = -\beta\,(u-u^m) \quad \text{in} \quad D.$$

2.1.1 Discretization

We define finite-dimensional subspaces $\mathcal{U}_h \in \mathcal{U}$ and $\mathcal{E}_h \in \mathcal{E}$ with basis $\{\phi_1,\ldots,\phi_P\}$ and $\{\psi_1,\ldots,\psi_Q\}$, respectively. This leads to the Galerkin approximation of the optimality system: Find $\{u_h, E_h, \lambda_h\} \in \mathcal{U}_h \times \mathcal{E}_h \times \mathcal{U}_h$ such that $F(u_h, E_h, \lambda_h, w) = 0\ \forall\ w \in U_h$. Here, $u_h = \sum_{i=1}^P u_i\phi_i$, $E_h = \sum_{i=1}^Q E_i\psi_i$, and $\lambda_h = \sum_{i=1}^P \lambda_i\phi_i$. Since $w = \sum_{i=1}^P w_i\phi_i$ for $w_i \in \mathbb{R}$, the Galerkin approximation of the optimality system is equivalent to

$$F(u_h, E_h, \lambda_h, w) = \left\{ \begin{array}{c} \langle w, \beta(u_h-u^m)\rangle - \langle\nabla w, E_h\nabla\lambda_h\rangle \\ \langle w, \nabla R(E_h)\rangle - \langle w, \nabla\lambda_h\cdot\nabla u_h\rangle \\ \langle w,\tau\rangle_{D_\tau} - \langle w, f\rangle - \langle\nabla w, E_h\nabla u_h\rangle \end{array} \right\} = 0. \tag{7}$$

Thus, the discrete optimality system of Eq. (7) is given by

$$[\mathbf{K}]\{\mathbf{u}\} = \{\mathbf{f}\}, \tag{8}$$

where

$$\mathbf{K} = \mathop{\mathbf{A}}_{e=1}^{n_e}\left(\int_{D^e}\sum_{i=1}^P\sum_{j=1}^P \nabla\phi_j\, E_h\nabla\phi_i\, dD\right), \tag{9}$$

$$\mathbf{f} = \mathop{\mathbf{A}}_{e=1}^{n_e}\left(\int_{D^e_\tau}\sum_{j=1}^P \phi_j\tau\, dD_\tau - \int_{D^e}\sum_{j=1}^P \phi_j f\, dD\right). \tag{10}$$

The discrete system of equations in Eq. (8) is used to solve the adjoint problem. However, the discrete force vector is given by

$$\widehat{\mathbf{f}} = \mathop{\mathbf{A}}_{e=1}^{n_e}\left(\beta\int_{D^e}\sum_{j=1}^P \phi_j\,(u_h-u^m)\, dD\right). \tag{11}$$

Finally, the discrete gradient is given by

$$D_E\mathcal{L} = \mathop{\mathbf{A}}_{e=1}^{n_e}\left(\int_{D^e}\sum_{j=1}^P \left[\phi_j\,\nabla R(E_h) - \phi_j\,(\nabla\lambda_h\cdot\nabla u_h)\right]\right), \tag{12}$$

where $\mathop{\mathbf{A}}_{e=1}^{n_e}$ is an assemble operator and n_e is the total number of elements.

2.1.2 Identification Procedure

The following algorithm is used to compute a new estimate for the elastic modulus at every kth iteration of the minimization process:

Repeat until convergence:

Given an estimate for the elastic modulus E^k

1. Solve Eq. (8) with $\mathbf{f}$ given by Eq. (10) to compute $u(E^k) \equiv u_h$.
2. Solve Eq. (8) with $\mathbf{f}$ given by Eq. (11) to compute $\lambda(E^k) \equiv \lambda_h$.
3. Use Eq. (12) to compute the discrete gradient $D_E\mathcal{L}$.
4. Compute a new estimate for the elastic modulus E^{k+1}.

2.2 MECE Formulation

Mathematically, we can describe the MECE functional as the sum of two errors: the error in the constitutive equation and the error in the unreliable experimental information (e.g., observed measurements, boundary conditions, etc.). The main idea of the MECE formulation is to use as reference the reliable experimental information (e.g., location of sensors, location of the source, etc.) at hand as well as the reliable equations of the model (e.g., equilibrium equations). The reliable information is verified exactly, while the constitutive equation and the unreliable information are verified in an average sense.

Define the following error functional:

$$\mathcal{J}(\sigma, u, E) = \frac{\alpha}{2}\left\langle (\sigma - E\nabla u), E^{-1}(\sigma - E\nabla u)\right\rangle, \tag{13}$$

which denotes the error in the constitutive eqaution. Here, $\mathcal{J} \geq 0$ and $\mathcal{J} = 0 \Longleftrightarrow \sigma = E\nabla u$. Now, we consider the nonlinear programming problem

$$\begin{aligned} \arg \underset{\{u,\sigma,E\}\in\, \mathcal{U}\times\mathcal{S}\times\mathcal{E}}{\text{minimize}} \quad & \frac{\gamma}{2}\langle \sigma\cdot n - \tau^m, \sigma\cdot n - \tau^m\rangle_{D_\tau} + \frac{\beta}{2}\langle u - u^m, u - u^m\rangle \\ & + \frac{\alpha}{2}\langle (\sigma - E\nabla u), E^{-1}(\sigma - E\nabla u)\rangle + R(E) \end{aligned} \tag{14}$$

subject to

$$\nabla\cdot\sigma = f \quad \text{in} \quad D,$$

where $\mathcal{U} = \{u\colon u \in H^1(D),\ u = 0 \text{ on } D_u\}$, $\mathcal{S} = \{\sigma\colon \sigma \in L^2(D),\ \nabla\cdot\sigma = f \text{ in } D,\ \sigma\cdot n = \tau \text{ on } D_\tau\}$, and $\mathcal{E} = \{E\colon E \in L^2(D),\ E > 0\}$. Here, $\alpha > 0,\ \beta > 0$ and $\gamma > 0$ are given parameters, and $R(\cdot)$ denotes a regularization operator.

The Lagrangian functional associated with the minimization problem defined in Eq. (14) is given by

$$\begin{aligned} \mathcal{L}(u,\sigma,E,\lambda) = & \frac{\gamma}{2}\langle \sigma\cdot n - \tau^m, \sigma\cdot n - \tau^m\rangle_{D_\tau} + \frac{\beta}{2}\langle u - u^m, u - u^m\rangle \\ & + \frac{\alpha}{2}\langle (\sigma - E\nabla u), E^{-1}(\sigma - E\nabla u) + R(E) + \langle \lambda, \nabla\cdot\sigma - f\rangle. \end{aligned} \tag{15}$$

If $\{u^*, \sigma^*, E^*\}$ is a local solution of Eq. (14), there exist Lagrange multipliers $\lambda \in \mathcal{U}$ such that the KKT conditions hold at the stationary point $\{u^*, \sigma^*, E^*, \lambda\}$, i.e.,

$$D\mathcal{L}(u,\sigma,E,\lambda)(\delta u, \delta\sigma, \delta E, \delta\lambda) = D_u\mathcal{L}\cdot\delta u + D_\sigma\mathcal{L}\cdot\delta\sigma + D_E\mathcal{L}\cdot\delta E + D_\lambda\mathcal{L}\cdot\delta\lambda = 0,$$

where

$$D_u\mathcal{L}\cdot\delta u = \alpha\langle\nabla\cdot(\sigma - E\nabla u), \delta u\rangle + \beta\langle u - u^m, \delta u\rangle = 0, \tag{16}$$

$$D_\sigma \mathcal{L} \cdot \delta\sigma = \alpha\langle(\sigma - E\nabla u)E^{-1}, \delta\sigma\rangle + \gamma\langle(\sigma \cdot n - \tau^m) \cdot n, \delta\sigma\rangle_{D_\tau} - \langle\nabla\lambda, \delta\sigma\rangle + \langle\lambda \cdot n, \delta\sigma\rangle_{D_\tau} = 0, \tag{17}$$

$$\begin{aligned} D_E \mathcal{L} \cdot \delta E = \langle\nabla R(E), \delta E\rangle - \alpha\langle(\sigma - E\nabla u) \cdot (E^{-1}\nabla u), \delta E\rangle \\ - \frac{\alpha}{2}\langle[(\sigma - E\nabla u)E^{-1}] \cdot [E^{-1}(\sigma - E\nabla u)], \delta E\rangle = 0, \end{aligned} \tag{18}$$

$$D_\lambda \mathcal{L} \cdot \delta\lambda = \langle\nabla \cdot \sigma - f, \delta\lambda\rangle = 0. \tag{19}$$

From Eq. (17) we get

$$\sigma = E\left(\nabla u + \frac{\nabla\lambda}{\alpha}\right) \quad \text{in} \quad D,$$

$$\tau = \sigma \cdot n = \tau^m - \frac{\lambda}{\gamma} \quad \text{on} \quad D_\tau.$$

This allows us to reformulate the nonlinear programming problem in Eq. (14) into

$$\begin{aligned} \arg \underset{E\in\mathcal{E}}{\text{minimize}} \quad & \frac{\beta}{2}\langle u(E) - u^m, u(E) - u^m\rangle + R(E) \\ & + \frac{1}{2\alpha}\langle E\nabla\lambda(E), E^{-1}\left[E\nabla\lambda(E)\right]\rangle + \frac{1}{2\gamma}\langle\lambda(E), \lambda(E)\rangle_{D_\tau}, \end{aligned}$$

where the gradient operator is given by

$$D_E \mathcal{L} = \nabla R(E) - \langle\nabla\lambda(E), \nabla u(E)\rangle - \frac{1}{2\alpha}\langle\nabla\lambda(E), \nabla\lambda(E)\rangle.$$

Here, $u(E)$ and $\lambda(E)$ are solutions to the boundary value problem

$$\begin{aligned} \nabla \cdot \left[E\left(\nabla u + \frac{\nabla\lambda}{\alpha}\right)\right] = f \quad \text{in} \quad D, \\ \nabla \cdot (E\nabla\lambda) = -\beta(u - u^m) \quad \text{in} \quad D, \\ \tau^m = \frac{\lambda}{\gamma} \quad \text{on} \quad D_\tau. \end{aligned} \tag{20}$$

2.2.1 Discretization

We define finite-dimensional subspaces $\mathcal{U}_h \in \mathcal{U}$ and $\mathcal{E}_h \in \mathcal{E}$ with basis $\{\phi_1, \ldots, \phi_P\}$ and $\{\psi_1, \ldots, \psi_Q\}$, respectively. This leads to the Galerkin approximation of the optimality system: Find $\{u_h, E_h, \lambda_h\} \in \mathcal{U}_h \times \mathcal{E}_h \times \mathcal{U}_h$ such that $F(u_h, E_h, \lambda_h, w) = 0 \; \forall \; w \in U_h$. Here, $u_h = \sum_{i=1}^P u_i\phi_i$, $E_h = \sum_{i=1}^Q E_i\psi_i$, and $\lambda_h = \sum_{i=1}^P \lambda_i\phi_i$. Since $w = \sum_{i=1}^P w_i\phi_i$ for $w_i \in \mathbb{R}$, the Galerkin approximation of the optimality system is equivalent to

$$F(u_h, E_h, \lambda_h, w) = \left\{ \begin{array}{c} \langle w, \beta(u_h - u^m)\rangle - \langle\nabla w, E\nabla\lambda_h\rangle \\ \langle w, \nabla R(E_h)\rangle - \langle w, \frac{\alpha}{2}(\nabla\lambda_h \cdot \nabla\lambda_h)\rangle - \alpha\langle w, \nabla\lambda_h \cdot \nabla u_h\rangle \\ \left\langle w, \tau^m - \frac{\lambda_h}{\gamma}\right\rangle_{D_\tau} - \langle w, f\rangle - \left\langle\nabla w, E(\nabla u_h + \frac{\nabla\lambda_h}{\alpha})\right\rangle \end{array} \right\} = 0. \tag{21}$$

Thus, the discrete optimality system of Eq. (21) is given by

$$\left[\begin{array}{cc} \mathbf{K} & \widehat{\mathbf{K}} \\ -\beta\mathbf{M} & \mathbf{K} \end{array} \right] \left\{ \begin{array}{c} \mathbf{u} \\ \lambda \end{array} \right\} = \left\{ \begin{array}{c} \mathbf{f} \\ \bar{\mathbf{f}} \end{array} \right\}, \tag{22}$$

where

$$\mathbf{K} = \mathop{\mathbf{A}}_{e=1}^{n_e} \left(\int_{D^e} \sum_{i=1}^{P} \sum_{j=1}^{P} \nabla \phi_j \, E_h \nabla \phi_i \, dD \right),$$

$$\widehat{\mathbf{K}} = \mathop{\mathbf{A}}_{e=1}^{n_e} \left(\int_{D^e} \sum_{i=1}^{P} \sum_{j=1}^{P} \nabla \phi_j \, E_h \nabla \phi_i \, dD + \int_{D^e_\tau} \sum_{i=1}^{P} \sum_{j=1}^{P} \frac{1}{\gamma} \phi_j \phi_i \, dD_\tau \right),$$

$$\mathbf{M} = \mathop{\mathbf{A}}_{e=1}^{n_e} \left(\int_{D^e} \sum_{i=1}^{P} \sum_{j=1}^{P} \phi_j \, \phi_i \, dD \right),$$

$$\mathbf{f} = \mathop{\mathbf{A}}_{e=1}^{n_e} \left(\int_{D^e_\tau} \sum_{j=1}^{P} \phi^j \tau^m \, dD_\tau - \int_{D^e} \sum_{j=1}^{P} \phi^j f \, dD \right),$$

$$\bar{\mathbf{f}} = \mathop{\mathbf{A}}_{e=1}^{n_e} \left(-\beta \int_{D^e} \sum_{j=1}^{P} \phi^j \, u^m \, dD \right).$$

Finally, the discrete gradient is given by

$$D_E \mathcal{L} = \mathop{\mathbf{A}}_{e=1}^{n_e} \left(\int_{D^e} \sum_{j=1}^{P} \left(\phi_j \, \nabla R(E_h) - \phi_j \, (\nabla \lambda_h \cdot \nabla u_h) - \frac{1}{2\alpha} \phi_j \, (\nabla \lambda_h \cdot \nabla \lambda_h) \right) \, dD \right). \tag{23}$$

2.2.2 Identification Procedure

The following algorithm is used to compute a new estimate for the elastic modulus at every kth iteration of the minimization process:

Repeat until convergence:

Given an estimate for the elastic modulus E^k

1. Solve Eq. (22) to compute $u(E^k) \equiv u_h$ and $\lambda(E^k) \equiv \lambda_h$.
2. Use Eq. (23) to compute the discrete gradient $D_E \mathcal{L}$.
3. Compute a new estimate for the elastic modulus E^{k+1}.

2.3 ECE Formulation

In the ECE approach we define a cost functional based on the error in the constitutive equations that connect a set of kinematically admissible displacements and a set of statically admissible stresses. The parameter estimation problem is solved by finding material properties, along with admissible displacement and stress fields, such that the ECE functional is minimized.

The ECE functional can be defined as

$$\mathcal{J}(u_D, \sigma_N, E) = \frac{\beta}{2} \langle \sigma_N - E \nabla u_D, \sigma_N - E \nabla u_D \rangle,$$

where $\beta > 0$ is a given parameter, u_D is a kinematically admissible displacement field, σ_N is a statically admissible stress field, and E is the elastic modulus. A displacement field u_D is said to be kinematically admissible if it satisfies

the measured displacement field u^m and satisfies Dirichlet boundary conditions. A stress field σ_N is said to be statically admissible if it satisfies the equilibrium equations and Neumann boundary conditions. Mathematically, we can describe the collection of these quantities as follows:

$$\widetilde{\mathcal{U}} = \{u_D \colon u_D \in H^1(D), \nabla \cdot (E\nabla u_D) = f \text{ in } D, u_D = 0 \text{ on } D_u,\ u_D = u^m \text{ in } D^m\},$$

$$\widetilde{\mathcal{S}} = \{\sigma_N \colon \sigma_N \in L^2(D), \nabla \cdot \sigma_N = f \text{ in } D,\ u = 0 \text{ on } D_u,\ \sigma_N \cdot n = \tau \text{ on } D_\tau\},$$

where D^m is the domain where the measurements are taken, $\widetilde{\mathcal{U}}$ is the space of kinematically admissible field u_D, and $\widetilde{\mathcal{S}}$ is the space of statically admissible field σ_N.

2.3.1 Alternating Directions Approach for ECE

We consider the problem

$$\arg \quad \underset{E\in\mathcal{E}}{\text{minimize}} \quad \frac{\beta}{2}\langle \sigma_N - E\nabla u_D, \sigma_N - E\nabla u_D\rangle + R(E), \tag{24}$$

where the gradient operator is given by

$$D_E\mathcal{J} = \frac{\beta}{2}\left(\sigma_N - E\nabla u_D\right)\cdot \nabla u_D + \nabla R(E). \tag{25}$$

An alternating directions approach was employed to solve the parameter estimation problem. The approach consists of breaking the optimization process into two steps. First, given a current best guess for the material properties, we find a statically admissible stress field and a kinematically admissible displacement field. Second, we use these admissible fields to compute the gradient and proceed to update the material properties.

The alternating direction approach can be described as follows:

Repeat until convergence:

1. Find a statically admissible stress field σ_N such that:

$$\begin{aligned} \nabla \cdot \sigma_N &= f \quad \text{in} \quad D \\ u &= 0 \quad \text{on} \quad D_u \\ \sigma_N \cdot n &= \tau \quad \text{on} \quad D_\tau \end{aligned} \tag{26}$$

2. Find a kinematically admissible displacement field u_D such that:

$$\begin{aligned} \nabla \cdot (E\nabla u_D) &= f \quad \text{in} \quad D \\ u_D &= 0 \quad \text{on} \quad D_u \\ u_D &= u^m \quad \text{in} \quad D^m \\ (E\nabla u_D)\cdot n &= \tau \quad \text{on} \quad D_\tau \end{aligned} \tag{27}$$

3. Use Eq. (25) to compute the gradient operator $D_E\mathcal{J}$.
4. Compute a new estimate for the elastic modulus E.

2.3.2 Discretization

We define finite-dimensional subspaces $\mathcal{U}_h \in \mathcal{U}$, $\mathcal{E}_h \in \mathcal{E}$, and $\widehat{\mathcal{U}}_h \in \widehat{\mathcal{U}}$ with basis $\{\phi_1, \ldots, \phi_P\}$, $\{\psi_1, \ldots, \psi_Q\}$, and $\{\chi_1, \ldots, \chi_R\}$, respectively. Here, $\mathcal{U} = \{u \colon u \in H^1(D),\ u = 0 \text{ on } D_u\}$, $\widehat{\mathcal{U}} = \{u \colon u \in H^1(D),\ u = 0 \text{ on } D_u,\ u = u^m \text{ in } D^m\}$, and $\mathcal{E} = \{E \colon E \in L^2(D),\ E > 0\}$. This leads to the following Galerkin approximation of the

statically admissible problem: Find $u_N^h \in \mathcal{U}_h \,|\, \langle \nabla w, E \nabla u_N \rangle = \langle w, \tau \rangle_{D_\tau} - \langle w, f \rangle \quad \forall\, w \in \mathcal{U}_h$. Likewise, the Galerkin approximation of the kinematically admissible problem is given by: Find $u_D^h \in \widehat{\mathcal{U}}_h \,|\, \langle \nabla \widehat{w}, E \nabla u_D \rangle = \langle w, \tau \rangle_{D_\tau} - \langle w, f \rangle \quad \forall\, \widehat{w} \in \widehat{\mathcal{U}}_h$. Here, $u_N^h = \sum_{i=1}^{P} u_N^i \phi_i$, $E^h = \sum_{i=1}^{Q} E_i \psi_i$, and $u_D^h = \sum_{i=1}^{R} u_D^i \chi_i$. Finally, the discrete system of equations for the kinematically and statically admissible problems is given by Eq. (8) and the discrete gradient is given by

$$D_E \mathcal{J} = \mathop{\mathbf{A}}_{e=1}^{n_e} \left(\int_{D^e} \sum_{j=1}^{Q} \left[\psi_j \, \nabla R(E_h) - \beta \psi_j \left((\sigma_N - E_h \nabla u_D) \cdot \nabla u_D \right) \right] dD \right). \tag{28}$$

2.4 Nonlinear Least-Squares

The nonlinear least-squares (NLLS) method finds a set of unknown parameters $\theta \in \mathbb{R}^{n_\theta}$ that best fit the set of observations $u \in \mathbb{R}^{n_u}$ with a deterministic model, $u = f(\mathbf{x}, \theta)$, $\mathbf{x} \in D$, that is nonlinear with the set of unknown parameters θ. Assume we have additive Gaussian noise $e \in \mathbb{R}^{n_u}$, which is mutually independent of the unknown parameter θ and identically distributed, allowing us to define the parameter-to-observable map as $u = f(\mathbf{x}, \theta) + e$. Finally, the optimal values of θ are obtained by minimizing the sum of squared errors (SSE) of a given cost function.

A typical formulation for an unconstrained NLLS problem is to find the optimal values of θ to minimize the following SSE cost function:

$$\mathcal{J}(\theta) = \sum_{i=1}^{n_u} \left[f(\mathbf{x}_i, \theta) - u_i \right]^2 = \sum_{i=1}^{n_u} \left[r_i(\theta) \right]^2.$$

The unconstrained NLLS approach requires an optimization algorithm to find the least-squares approximation $\widehat{\theta}$ of a minimum θ^*. Specialized optimization algorithms have been designed to exploit the structure of the SSE cost function. If $\mathcal{J}(\theta)$ is twice differentiable, terms of residual $r_i(\theta)$, $r_i''(\theta)$, and $[r_i'(\theta)]^2$ result. By assuming that the residuals $r_i(\theta)$ are close to zero near the solution, the Hessian matrix of second derivatives of $S(\theta)$ can be approximated using only first derivatives of $r_i(\theta)$. An algorithm that is particularly well-suited to the small-residual case and the above formulation is the Gauss-Newton algorithm. This formulation and algorithm combination typically requires the user to explicitly formulate each term in the least-squares (e.g., n_u terms for n_u data points) along with the gradients for each term.

The computational expense of the NLLS method implemented in this work increases as the number of parameters n_θ increases due to the forward difference approximation of the gradient, e.g., $\mathcal{O}(n_\theta)$. For this reason, a finite representation was used to approximate the spatially varying parameter of interest. The quality of a basis for the problem at hand can be judged based on the dimensionality of the subspace needed to approximate the desired parameter. The basis should capture the nature of the possible solution. In this work we use Gaussian RBF to approximate the elastic modulus in the region of interest since it can represent gradual or sharp changes of the material properties in localized regions (see Section 1.2).

2.4.1 Standard Formulation

For this case, the deterministic model is given by Eq. (1) and the SSE is defined as

$$\mathcal{J}(\theta) = \sum_{i=1}^{n_u} [u_i^m - u_i(\theta)]^2,$$

where u^m denotes the measured data and $u(\theta)$ is the finite element solution to Eq. (8).

2.4.2 MECE Formulation

For this case, the deterministic model is given by Eq. (20) and the SSE is defined as

$$\mathcal{J}(\theta) = \sum_{i=1}^{n_u} [u_i^m - u_i(\theta)]^2 + \sum_{j=1}^{n_\lambda} [\lambda_j(\theta)]^2,$$

where u^m denotes the measured data and n_λ denotes the number of Lagrange multipliers, $u(\theta)$ is the finite element solution, and $\lambda(\theta)$ is the Lagrange multiplier. These quantities are computed by solving Eq. (22). Finally, notice that $\lambda(\theta) \to 0$ as optimality is reached.

2.4.3 ECE Formulation

The alternating directions approach described in Section 2.3.1 was employed to solve the NLLS problem. Therefore, at each iteration of the NLLS problem we find a statically admissible stress field and a kinematically admissible displacement field by solving Eq. (26) and Eq. (27), respectively. The SSE for this problem is given by

$$\mathcal{J}(\theta) = \sum_{i=1}^{n_\sigma} [\sigma_N^i(\theta) - E(\theta)\nabla u_D^i)]^2,$$

where n_σ denotes the number of stress components, σ_N is the statically admissible stress field, and u_D is the kinematically admmissible displacement field. Finally, notice that $\mathcal{J}(\theta) = 0 \iff \sigma_N(\theta) = E(\theta)\nabla u_D$.

3. STOCHASTIC PARAMETER ESTIMATION METHODS

3.1 Bayesian Calibration

Nonuniqueness is a central feature of ill-posed inverse problems; hence, multiple solutions of the unknown may be consistent with the observations. Statistical inverse problems attempt to remove the ill-posedness by restating the inverse problem as a well-posed extension in a larger space of probability distributions [35]. Furthermore, statistical methods allow us to exploit prior information that is hidden in the regularization schemes often used in deterministic inverse problems.

Assume we measure a quantity $y \in \mathbb{R}^{n_y}$ in order to get information about parameter $x \in \mathbb{R}^{n_x}$. These quantities are related through a model which may be inaccurate and can contain parameters that are unknown. Moreover, the measured quantity is usually poorly known and contains errors $e \in \mathbb{R}^{n_y}$. In a Bayesian framework, quantities y, x, and e are viewed as random variables and the parameter-to-observable map $f\colon \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ is defined as

$$Y(\omega) = f[X(\omega)],$$

where $X(\omega)\colon \Omega \to \mathbb{R}^{n_x}$ and $Y(\omega)\colon \Omega \to \mathbb{R}^{n_y}$ are random variables and the sample space Ω is the set of all possible outcomes ω. In this work we denote random variables with uppercase letters and their realization with lowercase letters. Furthermore, for ease of presentation we will omit the explicit dependency of the random variables to ω in the subsequent derivations.

Assume now that X has a known prior probability density $\pi_{pr}(x)\colon \mathbb{R}^{n_x} \to \mathbb{R}$ and the known data consist of the observed quantity Y such that the marginal probability $\pi(y)\colon \mathbb{R}^{n_y} \to \mathbb{R}$. Then, the posterior probability $\pi_{post}(x)\colon \mathbb{R}^{n_x} \to \mathbb{R}$ is given by

$$\pi_{post}(x) := \pi(x \mid y) = \frac{\pi_{prior}(x)\,\pi(y \mid x)}{\pi(y)} \propto \pi_{prior}(x)\,\pi(y \mid x).$$

In the subsequent Bayesian formulation we use a scaled identity as the covariance matrix for both the prior and likelihood pdf's.

3.1.1 Standard Likelihood

Assume we have a noisy measurement $u \in \mathbb{R}^{n_u}$ and a deterministic model of the form $A(x)u = f$, as defined in Eq. (8), where $x \in \mathbb{R}^{n_x}$ is an unknown parameter. Furthermore, consider that the actual measurement is corrupted by additive noise e, which is mutually independent with measurement u and unknown parameter x. Then, if we have a solution operator

$$\widehat{u}\,(x) := A\,(x)^{-1}\,f \mid \widehat{u}\,(x) : \mathbb{R}^{n_x} \to \mathbb{R}^{n_u}, \tag{29}$$

the standard stochastic model is defined as

$$U = \widehat{U}(X) + \Sigma. \tag{30}$$

Here, $U\colon \Omega \to \mathbb{R}^{n_u}$, $X\colon \Omega \to \mathbb{R}^{n_x}$, $\widehat{U}\colon \Omega \to \mathbb{R}^{n_x} \to \mathbb{R}^{n_u}$, and $\Sigma\colon \Omega \to \mathbb{R}^{n_u}$ reflects both the modeling and observation errors.

If we denote $\pi_{error}\left[u - \widehat{u}(x)\right] : \mathbb{R}^{n_u} \to \mathbb{R}$ and $\pi_{prior}\left(x\right) : \mathbb{R}^{n_x} \to \mathbb{R}$ as the probability density function (pdf) of Σ and X, respectively. The posterior pdf is given by

$$\pi_{post}(x) \propto \pi_{error}\left[u - \widehat{u}(x)\right] \pi_{prior}(x).$$

For normally distributed prior and likelihood pdfs we have

$$\pi_{prior}(x) \propto \exp\left[-\frac{1}{2}\left(x - \bar{x}\right)^{\intercal} \Gamma_{prior}^{-1}\left(x - \bar{x}\right)\right], \tag{31}$$

$$\pi(u - \widehat{u}(x)) \propto \exp\left(-\frac{1}{2}\left[u - \widehat{u}(x) - \bar{e}\right]^{\intercal} \Gamma_{error}^{-1}\left[u - \widehat{u}(x) - \bar{e}\right]\right). \tag{32}$$

Similarly, we assume that the posterior pdf is normally distributed and is given by

$$\pi_{post}(x) \propto \exp\left(-\frac{1}{2}\left(x - \bar{x}\right)^{\intercal} \Gamma_{prior}^{-1}\left(x - \bar{x}\right) - \frac{1}{2}\left[u - \widehat{u}(x) - \bar{e}\right]^{\intercal} \Gamma_{error}^{-1}\left[u - \widehat{u}(x) - \bar{e}\right]\right), \tag{33}$$

where $\bar{x} \in \mathbb{R}^{n_x}$ is the mean of the prior pdf, $\bar{e} \in \mathbb{R}^{n_u}$ is mean of the likelihood pdf, $\Gamma_{prior} \in \mathbb{R}^{n_x \times n_x}$ is the prior covariance matrix, and $\Gamma_{error} \in \mathbb{R}^{n_u \times n_u}$ is the likelihood covariance.

3.1.2 MECE Likelihood

Assume we have a measurement $y \in \mathbb{R}^{n_u + n_\lambda}$ and a deterministic model of the form $A(x)(u + \lambda) = f$, as defined in Eq. (22), where $x \in \mathbb{R}^{n_x}$ is an unknown parameter, $u \in \mathbb{R}^{n_u}$ is the state solution , and $\lambda \in \mathbb{R}^{n_\lambda}$ are the Lagrange multipliers. Notice that measurement y takes into consideration both the state solution and the Lagrange multipliers. One can physically measure the state solution but not the Lagrange multipliers. However, we know that $\lambda \to 0$ as both optimality and feasibility are reached, allowing us to assume that the measured Lagrange multipliers are equal to zero. Finally, the modeling and observation errors are modeled as additive noise e, which is mutually independent of unknown parameters x and measurements y. Thus, if we have a solution operator

$$\widehat{h}(x) := A^{-1}(x) f \mid \widehat{h}\colon \mathbb{R}^{n_x} \to \mathbb{R}^{n_u + n_\lambda}, \tag{34}$$

the MECE stochastic model is defined as

$$Y = \widehat{H}(X) + \Sigma. \tag{35}$$

Here, $X\colon \Omega \to \mathbb{R}^{n_x}$, $Y\colon \Omega \to \mathbb{R}^{n_u + n_\lambda}$, $H\colon \Omega \to \mathbb{R}^{n_x} \to \mathbb{R}^{n_u + n_\lambda}$, and $\Sigma\colon \Omega \to \mathbb{R}^{n_u + n_\lambda}$ reflects both the modeling and observation errors.

Σ and X have pdf $\pi_{error}[u - \widehat{h}(x)]\colon \mathbb{R}^{n_u + n_\lambda} \to \mathbb{R}$ and $\pi_{prior}(x)\colon \mathbb{R}^{n_x} \to \mathbb{R}$, respectively. The posterior pdf is given by

$$\pi_{post}(x) \propto \pi_{prior}(x)\pi_{error}[y - \widehat{h}(x)].$$

For normally distributed prior and likelihood pdf's we have

$$\pi_{prior}(x) \propto \exp\left[-\frac{1}{2}\,(x-\bar{x})^{\intercal}\,\Gamma_{prior}^{-1}\,(x-\bar{x})\right], \tag{36}$$

$$\pi_{error}[y-\widehat{h}(x)] \propto \exp\left[-\frac{1}{2}[y-\widehat{h}(x)-\bar{e}]^{\intercal}\Gamma_{error}^{-1}[y-\widehat{h}(x)-\bar{e}]\right], \tag{37}$$

Similarly, we assume that the posterior pdf is normally distributed and is given by

$$\pi_{post}(x) \propto \exp\left[(x-\bar{x})^{\intercal}\,\Gamma_{prior}^{-1}\,(x-\bar{x}) - \frac{1}{2}[y-\widehat{h}(x)-\bar{e}]^{\intercal}\Gamma_{error}^{-1}[y-\widehat{h}(x)-\bar{e}]\right] \tag{38}$$

where $\bar{x} \in \mathbb{R}^{n_x}$ is the mean of the prior pdf, $\bar{e} \in \mathbb{R}^{n_u+n_\lambda}$ is the mean of likelihood pdf, $\Gamma_{prior} \in \mathbb{R}^{n_x \times n_x}$ is the prior covariance matrix, and $\Gamma_{error} \in \mathbb{R}^{(n_u+n_\lambda)\times(n_u+n_\lambda)}$ is the likelihood covariance matrix.

3.1.3 ECE Likelihood

Assume we have a measurement $y \in \mathbb{R}^{n_y}$. If an alternating directions approach is used, the parameter-to-observable map becomes linear with respect to the unknown parameter $x \in \mathbb{R}^{n_x}$. This results in a deterministic model of the form $y - vx = 0$, where $v \in \mathbb{R}^{n_v}$. In the context of the ECE formulation, y denotes the Neumann stress field σ_N and v denotes the Dirichlet strain field ∇u_D.

Consider that the modeling and observation errors are modeled as additive noise e, which is mutually independent of the unknown parameter x and measurements y. This leads to an ECE stochastic model of the form

$$Y = VX + \Sigma, \tag{39}$$

where $X\colon \Omega \to \mathbb{R}^{n_x}$, $Y\colon \Omega \to \mathbb{R}^{n_y}$, $V\colon \Omega \to \mathbb{R}^{n_v}$, and $\Sigma\colon \Omega \to \mathbb{R}^{n_y}$ reflects both modeling and observation errors.

$\pi_{error}(y-vx)\colon \mathbb{R}^{n_y} \to \mathbb{R}$ and $\pi_{prior}(x)\colon \mathbb{R}^{n_x} \to \mathbb{R}$ denote the pdf of Σ and X, respectively. The posterior pdf is given by

$$\pi_{post}(x) \propto \pi_{prior}(x)\pi_{error}(y-vx).$$

For normally distributed prior and likelihood pdf's we have

$$\pi(x) \propto \exp\left(-\frac{1}{2}(x-\bar{x})^{\intercal}\Gamma_{prior}^{-1}(x-\bar{x})\right), \tag{40}$$

$$\pi_{error}(y-vx) \propto \exp\left(-\frac{1}{2}(y-vx-\bar{e})^{\intercal}\Gamma_{error}^{-1}(y-vx-\bar{e})\right). \tag{41}$$

Similarly, assuming a normally distibuted posterior pdf we get

$$\pi_{post}(x) \propto \exp\left(-\frac{1}{2}(x-\bar{x})^{\intercal}\Gamma_{prior}^{-1}(x-\bar{x}) - \frac{1}{2}(y-vx-\bar{e})^{\intercal}\Gamma_{error}^{-1}(y-vx-\bar{e})\right), \tag{42}$$

where $\bar{x} \in \mathbb{R}^{n_x}$ is the mean of the prior pdf, $\bar{e} \in \mathbb{R}^{n_y}$ is the mean of likelihood pdf, $\Gamma_{prior} \in \mathbb{R}^{n_x \times n_x}$ is the prior covariance matrix, $\Gamma_{error} \in \mathbb{R}^{n_y \times n_y}$ is the likelihood covariance matrix.

3.2 MAP Estimate

In large-scale inverse problems, the posterior distribution lives in a high-dimensional space and sampling methods become expensive and often unfeasible. For problems where the parameter-to-observable map is linear of the form $AX = Y\colon \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ and the prior and likelihood pdf's are normally distributed, the posterior pdf can be taken

as normally distributed [35]. In this case, point estimates and standard deviations of the unknown parameters can be calculated using an MAP estimate approach.

Mathematically this approach can be described as follows: Given a linear parameter-to-observable map and normally distirbuted prior and likelihood pdf's, the MAP estimate can be found by solving the following deterministic optimization problem:

$$x_{MAP} = \arg\underset{x\in\mathbb{R}^{n_x}}{\text{maximize}} \quad \pi(x \mid y). \tag{43}$$

The possible nonexistence and nonuniqueness of such problems indicate that the MAP estimate may be unsatisfactory. However, if an efficient deterministic optimization strategy is used, the MAP estimate can yield information about the mean and standard deviation of the parameter of interest when the posterior distribution lives in a high-dimensional space.

This approach can be extended to problems where the parameter-to-observable map is nonlinear of the form $A(X)U = Y : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$. Similar to the linear case, the MAP estimate is found by solving Eq. (43). However, in order to find an expression to calculate the posterior covariance, we linearize the operator $A(X)$ at the equilibrium point x_{MAP}. This linearization allows us to take advantage of the local stability of the linear approximation to the operator $A(X)$ [28, 36].

3.2.1 Standard Likelihood

The standard stochastic model for this problem is given by Eq. (30). The respective normally distributed prior and likelihood pdf's are given by Eq. (31) and Eq. (32). Then, the MAP estimate is obtained by maximizing the exponent on Eq. (33), or by minimizing its negative as follows:

$$x_{MAP} = \arg\underset{x\in\mathbb{R}^{n}}{\text{minimize}} \quad \frac{1}{2}\left(x-\bar{x}\right)^{\intercal}\Gamma_{prior}^{-1}\left(x-\bar{x}\right) + \frac{1}{2}\left[u-\widehat{u}(x)-\bar{e}\right]^{\intercal}\Gamma_{error}^{-1}\left[u-\widehat{u}(x)-\bar{e}\right]. \tag{44}$$

The optimization scheme presented in Section 2.1 is utilized to solve the resulting deterministic optimization problem.

An approximation to the posterior covariance matrix is derived by linearizing the cost function in Eq. (44) around the equilibrium point or MAP estimate. This results in an expression of the form

$$\Gamma_{post} \simeq \left[\Gamma_{prior}^{-1} + \widehat{u}_x(x)^{\intercal}\Gamma_{error}^{-1}\widehat{u}_x(x)\right]^{-1}.$$

where $\Gamma_{post} \in \mathbb{R}^{n_x\times n_x}$ is the posterior covariance matrix, subscript x denotes partial differentiation with respect to the unknown parameter x, and $\widehat{u}_x$ is the sensitivity of the solution operator defined in Eq. (29) with respect to the unknown parameter. Notice that the computation of the sensitivities requires the solution of n_x linear system of equations.

3.2.2 MECE Likelihood

The stochastic model for this problem is given by Eq. (35). The respective normally distributed prior and likelihood pdf's are given by Eq. (36) and Eq. (37). Then, the MAP estimate is given by maximizing the exponent on Eq. (38), or by minimizing its negative as follows:

$$x_{MAP} = \arg\underset{x\in\mathbb{R}^{n}}{\text{minimize}} \quad \frac{1}{2}\left(x-\bar{x}\right)^{\intercal}\Gamma_{prior}^{-1}\left(x-\bar{x}\right) + \frac{1}{2}[y-\widehat{h}(x)-\bar{e}]^{\intercal}\Gamma_{error}^{-1}[y-\widehat{h}(x)-\bar{e}]. \tag{45}$$

The optimization scheme presented in Section 2.2 is utilized to solve the resulting deterministic optimization problem.

An approximation to the posterior covariance matrix is derived by linearizing the cost function in Eq. (45) around the MAP estimate. This results in an expression of the form

$$\Gamma_{post} \simeq \left[\Gamma_{prior}^{-1} + \widehat{h}_x(x)^{\intercal}\Gamma_{error}^{-1}\widehat{h}_x(x)\right]^{-1},$$

where $\Gamma_{post} \in \mathbb{R}^{n_x \times n_x}$ is the posterior covariance matrix, subscript x denotes partial differentiation with respect to the unknown parameter x, and $\widehat{h}_x$ is the sensitivity of the solution operator defined in Eq. (34) with respect to the unknown parameter x. Similar to the standard likelihood case, the computation of the sensitivities requires the solution of n_x linear system of equations.

3.2.3 ECE Likelihood

The stochastic model for this problem is given by Eq. (39). The normally distributed prior and likelihood pdf's are given by Eq. (40) and Eq. (41). Then, the MAP estimate is given by maximizing the exponent on Eq. (42), or by minimizing its negative as follows:

$$x_{MAP} = \arg\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}(x-\bar{x})^\intercal \Gamma_{prior}^{-1}(x-\bar{x}) + \frac{1}{2}(y-vx-\bar{e})^\intercal \Gamma_{error}^{-1}(y-vx-\bar{e}). \tag{46}$$

Notice that we have employed an alternating direction approach in order to use the stochastic model defined in Eq. (39). Thus, the optimization scheme presented in Section 2.3 is used to solve the resulting deterministic optimization problem.

The parameter-to-observable map for this problem is linear. Thus, the operator $A(x)$ is not linearized around the MAP estimate. This results in an expression for the posterior covariance matrix of the form

$$\Gamma_{post} = \left(\Gamma_{prior}^{-1} + v^\intercal \Gamma_{error}^{-1} v\right)^{-1},$$

where $\Gamma_{post} \in \mathbb{R}^{n_x \times n_x}$ is the posterior covariance matrix.

4. RESULTS

A numerical study was performed to test the proposed methodologies for deterministic and stochastic parameter estimation. The numerical model consisted of a one-dimensional finite element model. Figure 1 shows a pictorial description of the domain and boundary conditions used for the numerical study. The domain $D = [0, L]$, where $L = 1\ m$. On $x = L,\ \tau = 1\ Pa$ and on $x = 0,\ u = 0$. For the purpose of this case study, the surface tractions were assumed to be known exactly, the body forces were taken as $f = 0$, and the penalty parameters α and β were each set to 1.

The observed displacement field for the parameter estimation problem was generated by solving one forward finite element problem using the benchmark material properties. The numerical model used to generate the 'observed' displacement field was built using a 2000 finite element mesh with fully integrated two-node line elements. The numerical model used for the inversion was built using a 1000 finite element mesh with fully integrated two-node line elements. Different levels of random Gaussian noise were considered, $\Delta \sim N(0, 1)$ with standard deviation $\zeta = 0.0,\ 0.01$, and 0.05, to examine the accuracy of the proposed methodologies to corrupt data. The synthetic noisy data are generated as follows:

$$\widehat{u}_i^m = u_i^m * (1 + \zeta * \Delta_i), \quad i = 1, \ldots, N_{obs},$$

where $\widehat{u}^m$ is the perturbed measured data and N_{obs} is the total number of observations.

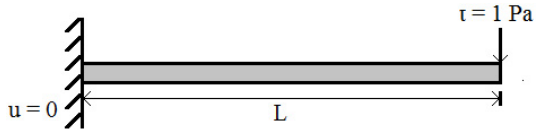


FIG. 1: Simple schematic of the beam model used for the numerical studies.

The misfit between the benchmark elastic modulus E_b and the optimal elastic modulus E_{opt} was quantified as follows:

$$\epsilon_E = \frac{\left[\sum_{i=1}^{N_p} \left(E_b^i - E_{opt}^i\right)^2\right]^{1/2}}{\left[\sum_{i=1}^{n} \left(E_b^i\right)^2\right]^{1/2}},$$

where N_p denotes the total number of parameters.

The following definition defined in [37] was used for the regularization operator:

$$R(E) = \frac{\kappa}{2\eta} \int_\Omega \left(\nabla E \cdot \nabla E + \varepsilon\right)^\eta \, d\Omega,$$

where $\kappa > 0$ is the regularization coefficient. This expression gives the flexibility of imposing Tikhonov regularization when ($\eta = 1$, $\varepsilon = 0$) and total variation regularization when ($\eta = 1/2$, $0 < \varepsilon \leq 1$).

We compared many different variations of calibration methods (specifically, optimization of a cost functional, nonlinear least-squares, Bayesian MCMC calibration, and MAP estimate), regularization type (Tikhonov vs total variation), and problem formulation (specifically L^2, MECE, and ECE). Table 1 shows the different combinations of options that we examined and present in this study. Finally, all calculations were performed on a Linux workstation with a 2.13 GHz Intel Core i7 processor and 8 GB of RAM.

4.1 Parameter Estimation: Deterministic Approach

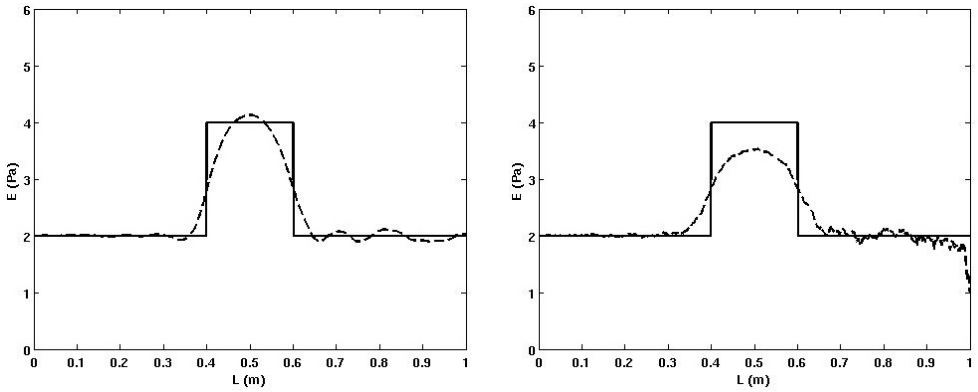
For the constrained optimization approaches study herein we set the bounds to 1 Pa $\leq E \leq$ 10 Pa; the initial guess E_{init} was taken as E_{init} = 5 Pa. The constrained optimization problem was solved using the npsol_sqp method in DAKOTA [38]. The npsol_sqp method is a sequential quadratic programming algorithm, which uses an augmented Lagrangian merit function and a Broyden-Fletcher-Goldfarb-Shanno (BFGS) approximation to the Hessian of the Lagrangian, and will not necessarily satisfy the constraints until the final solution. It uses a sufficient-decrease line

TABLE 1: Combinations of methods examined in this study

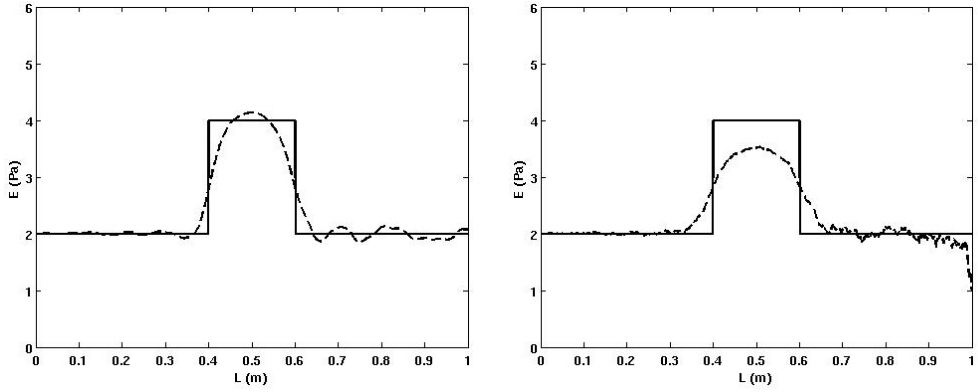
Calibration method	Problem formulation	Regularization
Constrained optimization of a cost function	L^2 MECE ECE	Tikhonov
	L^2 MECE ECE	Total variation
Unconstrained optimization of a cost function: Nonlinear least squares	Standard MECE ECE	None
Bayesian MCMC	Standard likelihood MECE likelihood ECE likelihood	Gaussian prior
Bayesian MAP	Standard likelihood MECE likelihood ECE likelihood	Total variation

search approach, which is a gradient-based line search for analytic or DAKOTA-supplied numerical gradients and is a value-based line search in the vendor numerical case.

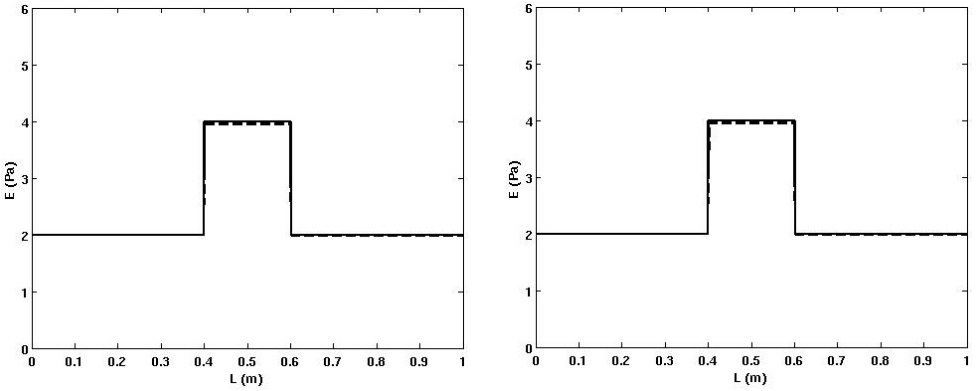
Figures 2 and 3 show the optimal elastic modulus field obtained for each of the constrained optimization approaches investigated in this work. For the case with 1% noise in the data, the elastic modulus field predicted with



(a) L^2 Formulation

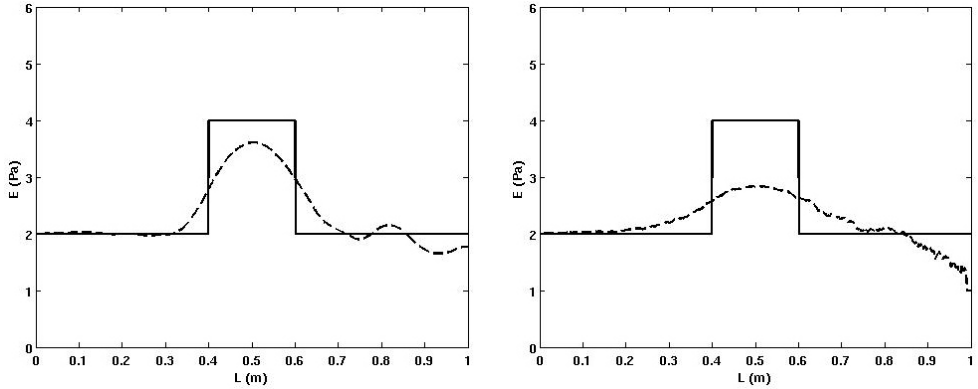


(b) MECE Formulation

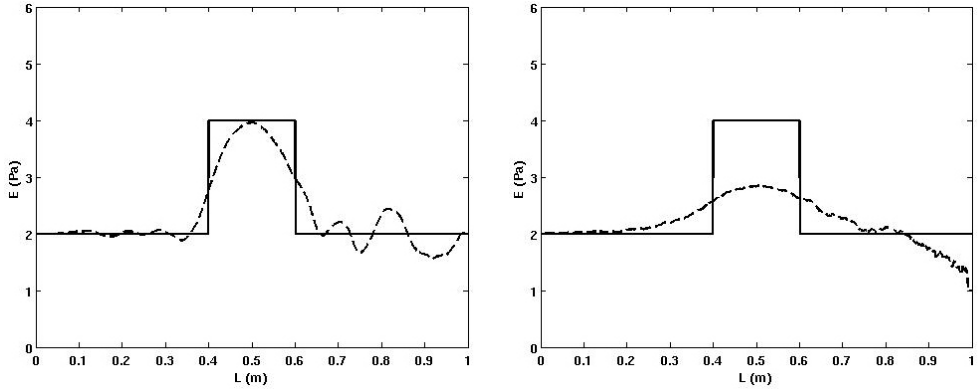


(c) ECE Formulation

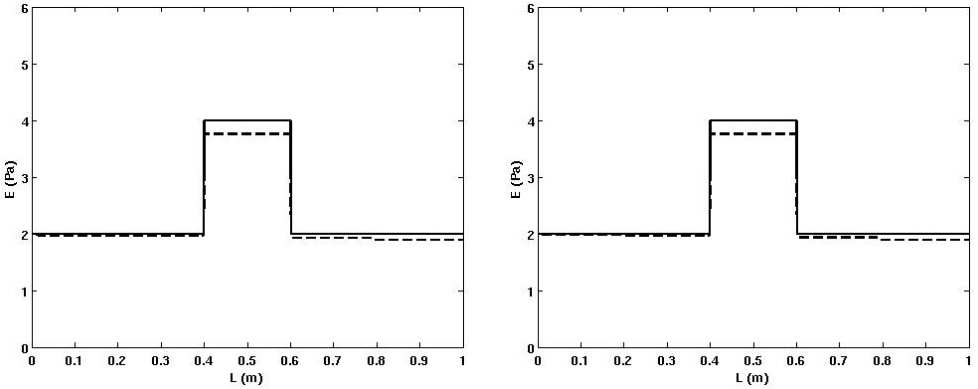
FIG. 2: Predicted elastic modulus E with 1% noise in the data for each constrained optimization approach investigated. Left column: Predicted E with TV regularization. Right column: Predicted E with Tikhonov regularization. Here, the solid black line denotes the target elastic modulus field, the dashed line denotes the predicted elastic modulus.



(a) L^2 Formulation



(b) MECE Formulation



(c) ECE Formulation

FIG. 3: Predicted elastic modulus field E with 5% noise in the data for each constrained optimization approach investigated. Left column: Predicted E with TV regularization. Right column: Predicted E with Tikhonov regularization. Here, the solid black line denotes the target elastic modulus field, the dashed line denotes the predicted elastic modulus.

each approach was close to the benchmark elastic modulus field, regardless of which regularization operator was implemented. However, Tables 2 and 3 show that the reconstruction errors were lower when TV regularization was used instead of Tikhonov regularization as the corruption in the observed data increased. For the case with 5% noise in the data, each optimization approach predicted an elastic modulus field that was close to the benchmark elastic modulus

TABLE 2: Reconstruction errors ϵ_E for each constrained optimization approaches study herein. The results correspond to the case where Tikhonov regularization was implemented

Noise ζ	**Formulation**	κ	ϵ_E	**Number of iterations** N_{iter}
0.0	L^2	1.0×10^{-6}	0.0269	557
	MECE	1.0×10^{-6}	0.0267	557
	ECE	1.0×10^{-4}	0.0167	6
0.01	L^2	1.0×10^{-3}	0.1468	117
	MECE	1.0×10^{-3}	0.1460	117
	ECE	1.0×10^{-4}	0.0201	6
0.05	L^2	1.0×10^{-2}	0.2443	53
	MECE	1.0×10^{-2}	0.2443	53
	ECE	1.0×10^{-4}	0.0594	6

TABLE 3: Reconstruction errors ϵ_E for each constrained optimization approaches study herein. The results correspond to the case where TV regularization was implemented

Noise ζ	**Formulation**	κ	ε	ϵ_E	**Number of iterations** N_{iter}
0.0	L^2	1.0×10^{-8}	1.0×10^{-8}	0.0368	222
	MECE	1.0×10^{-8}	1.0×10^{-8}	0.0366	321
	ECE	1.0×10^{-6}	1.0×10^{-6}	0.0167	6
0.01	L^2	1.0×10^{-5}	1.0×10^{-6}	0.0999	192
	MECE	1.0×10^{-5}	1.0×10^{-6}	0.0954	162
	ECE	1.0×10^{-6}	1.0×10^{-6}	0.0206	6
0.05	L^2	1.0×10^{-4}	1.0×10^{-6}	0.1505	319
	MECE	1.0×10^{-4}	1.0×10^{-6}	0.1321	217
	ECE	1.0×10^{-5}	1.0×10^{-6}	0.0542	5

field only when TV regularization was implemented. Moreover, the optimization approach based on the ECE formulation produced an elastic modulus field close to the benchmark solution with the least number of iterations, regardless of which regularization operator was implemented.

The computational savings obtained with an approach based on the ECE formulation can be due to the convexity property of the ECE cost functional, which has been shown to be convex for elliptic boundary value problems [39, 40]. To test the accuracy and precision of the constrained optimization approaches, several numerical experiments with different initial guesses were done. The approach based on the ECE formulation consistently predicted accurate elastic modulus fields with the fewest iterations. In some cases, the optimization approach based on the MECE formulation provided accurate results with fewer iterations than the optimization approach based on the L^2 formulation. However, the results were not as consistent as those obtained when the ECE formulation was implemented. The authors are not aware of an existing formal study that shows that the MECE cost functional is convex. However, Allix et al. [41] performed a numerical study in which they showed that the MECE cost functional was convex for a linear elastodynamic problem.

In the context of the regularization methods, notice that the reconstruction errors obtained with TV regularization were less than those obtained with Tikhonov regularization as the corruption in the observed data increased. Indeed, TV regularization has the ability to capture sharp discontinuities (see, e.g., [42]). Thus, one can effectively reconstruct functions with jump discontinuities, which is the case in this numerical study. The results could be improved if an

adequate method (e.g., discrepancy principle method, L-curve method, and generalized cross-validation method) was used to optimally select the regularization coefficient κ. The selection of κ becomes crucial in most applications as the corruption in the observed data increases. However, in this numerical study, κ was selected through numerical experiments.

The unconstrained nonlinear least-square optimization was performed using the nl2sol optimization method in DAKOTA. The nl2sol algorithm is a secant-based least-squares algorithm that is q-superlinearly convergent, where q denotes the quotient between two successive terms. It adaptively chooses between the Gauss-Newton Hessian approximation and this approximation augmented by a correction term from a secant update. nl2sol is more robust than many Gauss-Newton-based least-squares solvers, which experience difficulty when the residuals at the solution are significant. nl2sol is appropriate for large residual problems, i.e., least-squares problems for which the residuals do not tend toward zero at the solution. In this numerical study we have 15 residuals from 15 experimental points. The elastic modulus field was represented as an expansion of N_r radially symmetric RBF. The number of radial basis functions was set to 5 *a priori*. The bounds for the RBF parameters were set to 1 Pa $\leq E_c \leq$ 10 Pa, 1 Pa $\leq E_i \leq$ 10 Pa, and $-2 \leq z_i \leq 0$. The initial guess was set to E_c = 5 Pa, $\widehat{E}_i$ = 5 Pa, and $z_i = -1$. Finally, the maximum number of nl2sol iterations was set to 100.

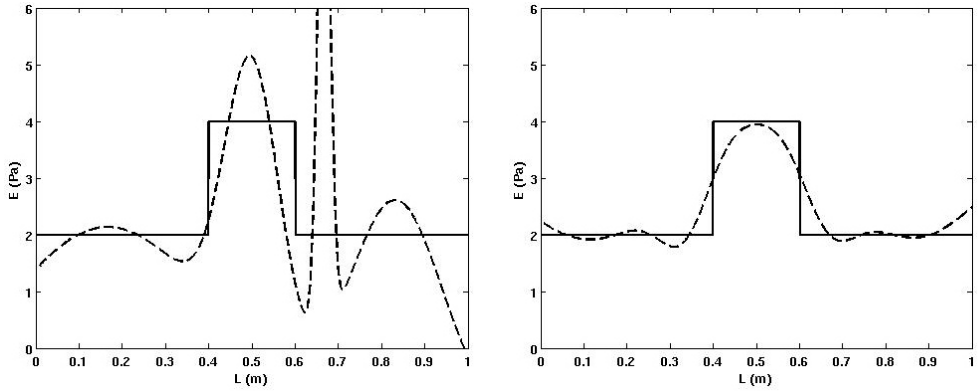
Figure 4 shows the predicted elastic modulus field using an unconstrained NLLS approach with 5% and 1% noise in the data. For the case with 1% noise in the data, each of the formulations investigated predicted elastic modulus fields close to the benchmark solution. However, for the case with 5% noise in the data, only the NLLS approaches based on the MECE and ECE formulations predicted elastic modulus fields close to the benchmark solution. Indeed, the reconstruction errors shown in Table 4 support this claim. The MECE and ECE formulations seem to provide a regularization behavior, which improves the quality of the reconstructions. Furthermore, the NLLS approach based on the ECE formulation required the fewest iterations to predict an optimal elastic modulus field. However, the reconstruction errors obtained with the MECE formulations were the lowest out of the three formulations investigated in this study. Finally, several numerical experiments with different initial guesses were performed to test the accuracy and precision of each NLLS formulation. The NLLS approach based on the ECE formulation consistently predicted elastic modulus fields close to the benchmark elastic modulus field with the fewest iterations. Furthermore, the standard NLLS formulation consistently produced the highest reconstruction errors as the corruption in the observed data increased.

4.2 Parameter Estimation Under Uncertainty: MCMC Approach

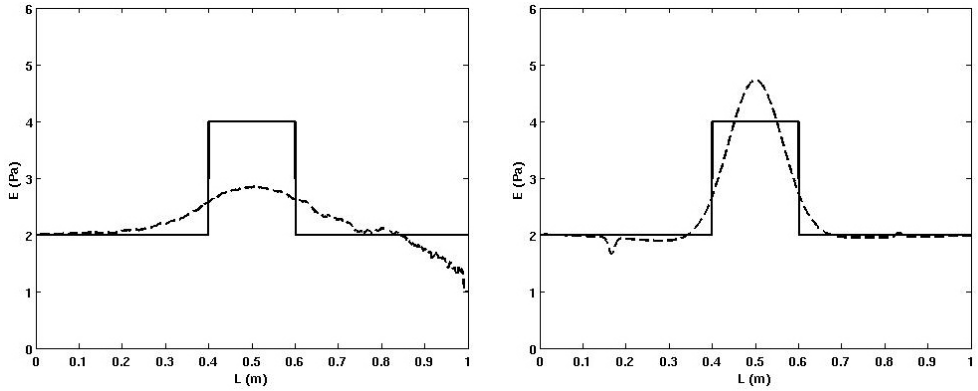
This section presents the results of Bayesian calibration. A delayed rejection adaptive metropolis (DRAM) algorithm was used to generate the posterior distribution. The DRAM algorithm used in this work can be found at `http://www.helsinki.fi/~mjlaine/dram/`. DRAM is a combination of two ideas, delayed rejection and adaptive Metropolis, for improving the efficiency of the Metropolis-Hastings MCMC algortihm [43]. The main idea of Delayed

TABLE 4: Reconstruction errors ϵ_E for each NLLS formulation study herein

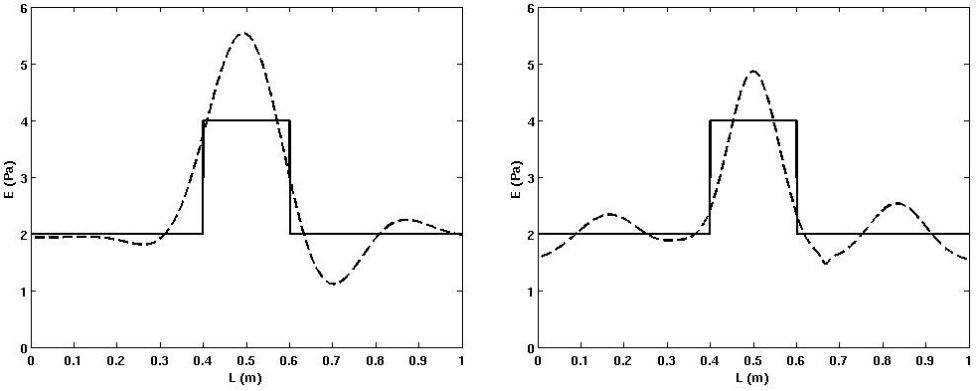
Noise ζ	**Formulation**	ϵ_E	**Number of iterations** N_{iter}
0.0	Standard	0.1275	56
	MECE	0.1333	27
	ECE	0.1401	4
0.01	Standard	0.1174	21
	MECE	0.1288	14
	ECE	0.1782	6
0.05	Standard	0.5591	16
	MECE	0.1317	28
	ECE	0.2348	7



(a) Standard Formulation



(b) MECE Formulation



(c) ECE Formulation

FIG. 4: Predicted elastic modulus field E obtained using the NLLS method. Left column: 5% noise in the data. Right column: 1% noise in the data. Here the solid black line denotes the target elastic modulus field and the dashed line denotes the predicted elastic modulus.

Rejection is to allow partial local adaptation of proposal within each time step of the Markov chain, while retaining its property and reversibility [44]. In adaptive Metropolis, the covariance matrix of the Gaussian proposal distribution is adapted on-the-fly using information from the previous chain. This adaptation destroys the Markovian property of the chain. However, it can be formally shown that the ergodicity properties of the generated sample remain [45].

The DRAM algorithm was used to generate 20,000 samples of the posterior parameter estimates and only the last 15,000 realizations were used to compute the relevant statistics. Due to the dimensionality of the problem, MCMC methods can be difficult to implement and accurate posterior parameter estimates are difficult to attain. To achieve a reduction in the dimension of the problem, the elastic modulus field was represented as an expansion of N_r radially symmetric basis functions. Thus, posterior estimates of the RBF parameters were generated instead of posterior estimates of the elastic modulus for each point in space. In this case study, five RBF were selected *a priori* to represent the elastic modulus field. Finally, the bounds for each RBF parameter were set to those used in the NLLS case study.

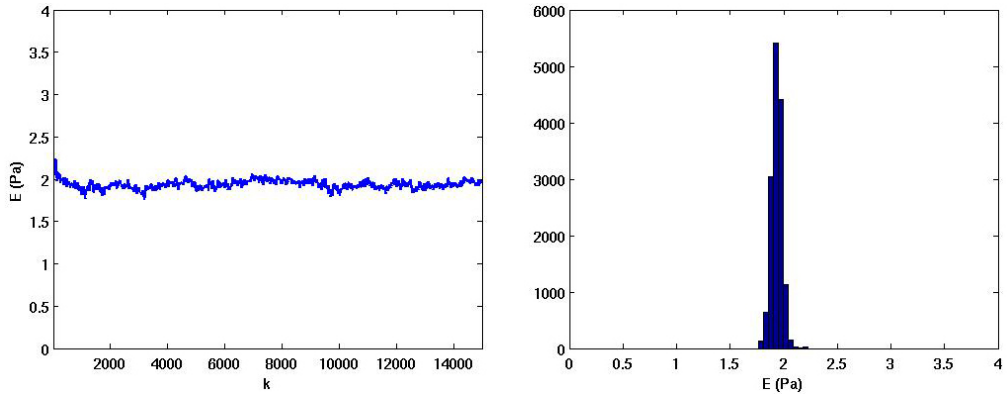
It is important to check convergence and mixing of the chain before analyzing the results. The simplest way is by visualizing the trace plots of the chain. Figures 5–7 show the one-dimensional (1D) trace plots of the Markov chain k for the parameters corresponding to one RBF. Visual inspection of the 1D trace plots suggests that the chain mixes well for the last 15,000 samples, regardless of the likelihood model. This implies that the samples are indeed from the stationary distribution of the Markov chain.

The autocorrelation function (ACF) is a useful tool to assess the convergence of a chain. If the standard deviation is too large, the chain easily moves out of the posterior support. Most of the proposal samples will be rejected, resulting in a long correlation. On the contrary, if the standard deviation is small, many of the proposal samples will be accepted and the chain can only move around a small portion of the posterior space and mix poorly. An appropriate standard deviation should result in a fast decay of the ACF with a lag along the chain [46, 47]. Figure 8 shows the ACF for each of the likelihood models investigated in this work. Notice that the ACF for the ECE likelihood model decays to zero at a very small lag for each of the radial basis function parameters. This is consistent with the good mixing observed in Figs. 5–7. However, the ACF plot shows that the standard and MECE likelihood models suffer from a slower rate of convergence.

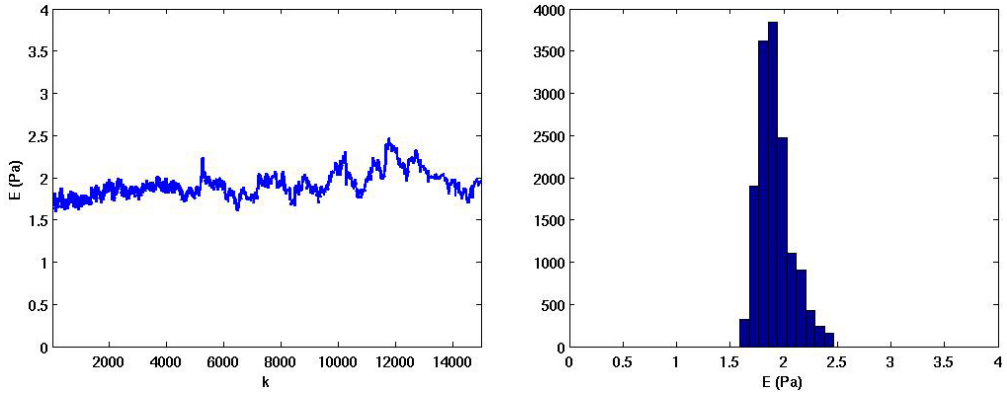
The posterior MCMC samples obtained for the parameters corresponding to one RBF are shown in Figs. 5–7. Notice that the posterior pdf obtained with the ECE likelihood model displays less variability from the mean. Several numerical experiments with different initial states were performed to verify that these results were consistently reproduced. Indeed, from these numerical experiments it was observed that the posterior pdf obtained with the ECE likelihood model consistently displayed less variability from its mean. Furthermore, from the numerical experiments we noticed that the ECE likelihood model required fewer burn-in samples to reach a stationary distribution of the Markov chain. However, the standard and MECE likelihood models resulted in better reconstructions as the corruption in the observed data increased; see Table 5. Overall, each of the likelihood models studied herein were capable of producing accurate elastic modulus field estimates as the corruption in the observed data increased; see Fig. 9. The only advantage seen was that the ECE likelihood model seems to required fewer burn-in samples to reach a stationary distribution of the Markov chain, providing a regularization behavior.

TABLE 5: Bayesian calibration: Reconstruction errors ϵ_E for each of the likelihood model study herein. The RBF parameters mean were used to compute the reconstruction errors

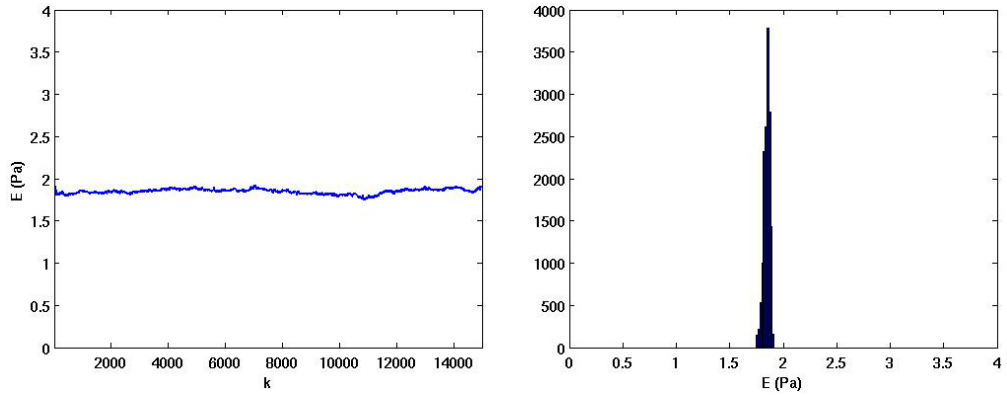
Noise ζ	**Likelihood model**	ϵ_E
0.0	Standard	0.1169
	MECE	0.1125
	ECE	0.1192
0.01	Standard	0.1489
	MECE	0.1173
	ECE	0.1173
0.05	Standard	0.1205
	MECE	0.1392
	ECE	0.2217



(a) Standard Likelihood



(b) MECE Likelihood

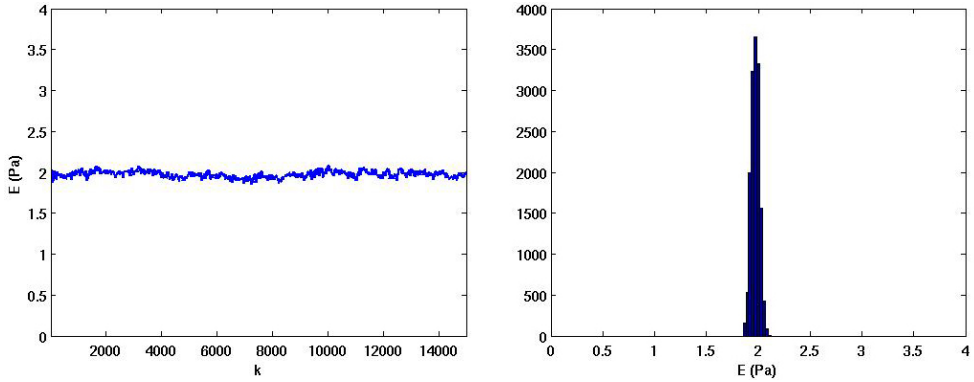


(c) ECE Likelihood

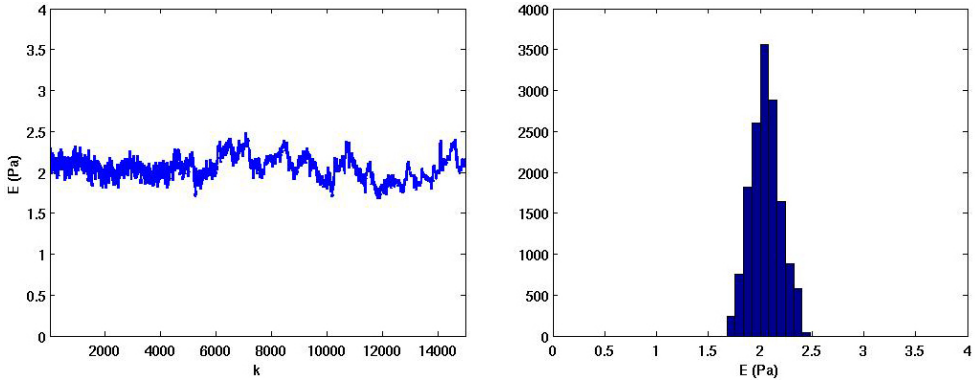
FIG. 5: Bayesian calibration with 5% noise in the data: Left column: Trace plot for the matrix elastic modulus E_c. Right column: Posterior pdf for E_c.

4.3 Parameter Estimation Under Uncertainty: MAP Estimate Approach

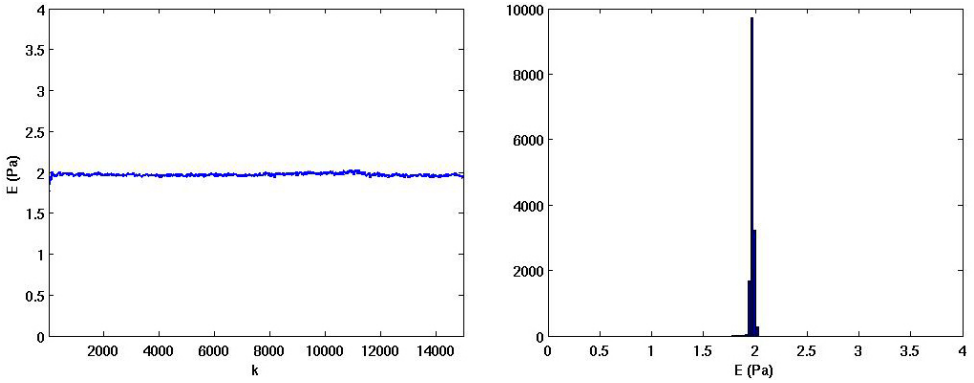
This section presents the results obtained with the MAP estimate approach. This approach produces the most likely value of the posterior parameter distribution. The resulting statistical parameter estimation problem is equivalent to solving a regularized deterministic optimization problem with Gaussian likelihood and prior. The problem was solved



(a) Standard Likelihood



(b) MECE Likelihood

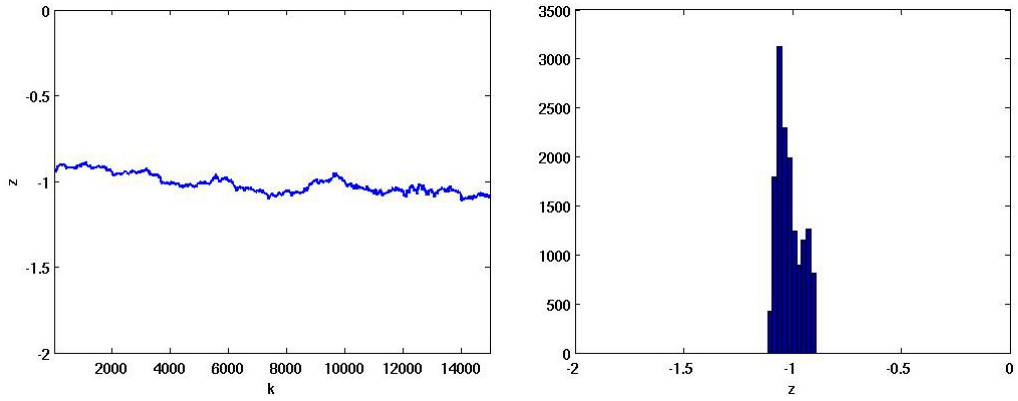


(c) ECE Likelihood

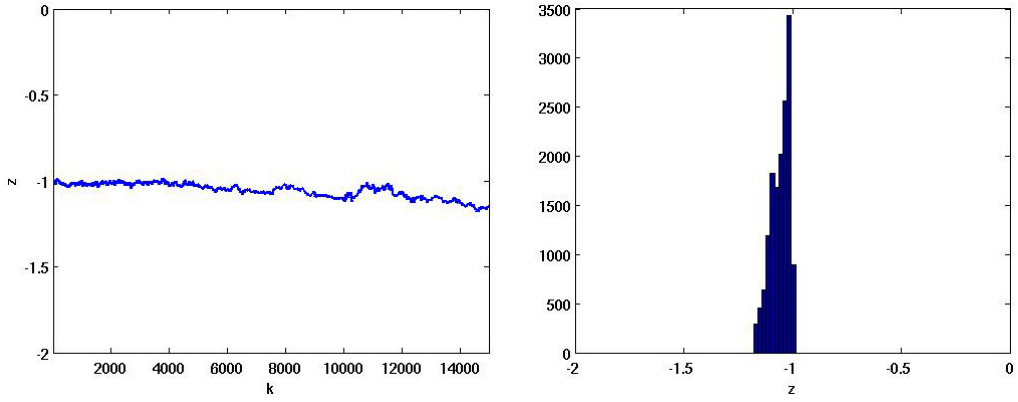
FIG. 6: Bayesian calibration with 5% noise in the data: Left column: Trace plot for the elastic modulus E_1 at the center of a radial basis function. Right column: Posterior pdf for E_1.

as a constrained optimization problem using the npsol_sqp method in DAKOTA. The bounds were set to 1 Pa $\leq E \leq$ 10 Pa and the initial guess was set to E_{init} = 5 Pa.

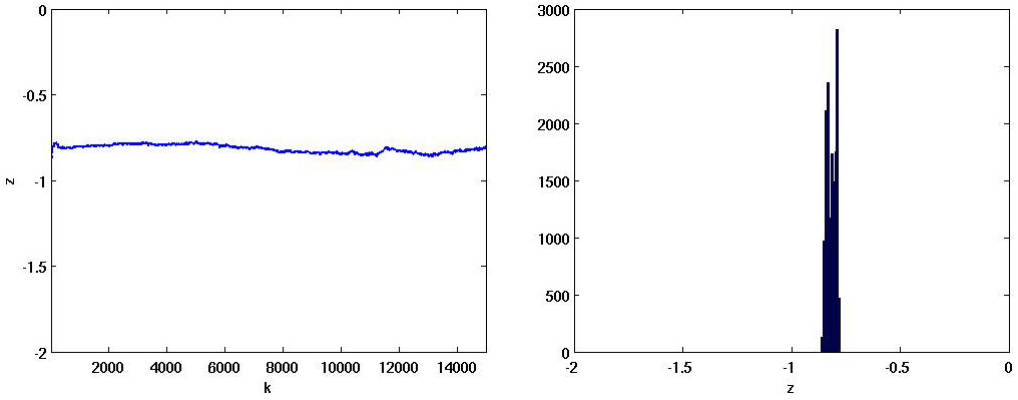
The MAP method was investigated in the context of the standard, MECE, and ECE Bayesian formulations presented in Section 3.2. The discretization of Eqs. (44)–(46) is equivalent to the choice of Gaussian pdf for the prior and likelihood functions



(a) Standard Likelihood



(b) MECE Likelihood



(c) ECE Likelihood

FIG. 7: Bayesian calibration with 5% noise in the data: Left column: Trace plot for the locality parameter z_1 of a radial basis function. Right column: Posterior pdf for z_1.

$$\pi_{prior} = \mathcal{N}(\bar{x}, \Gamma_{prior}),$$

$$\pi_{error} = \mathcal{N}(\bar{e}, \Gamma_{error}),$$

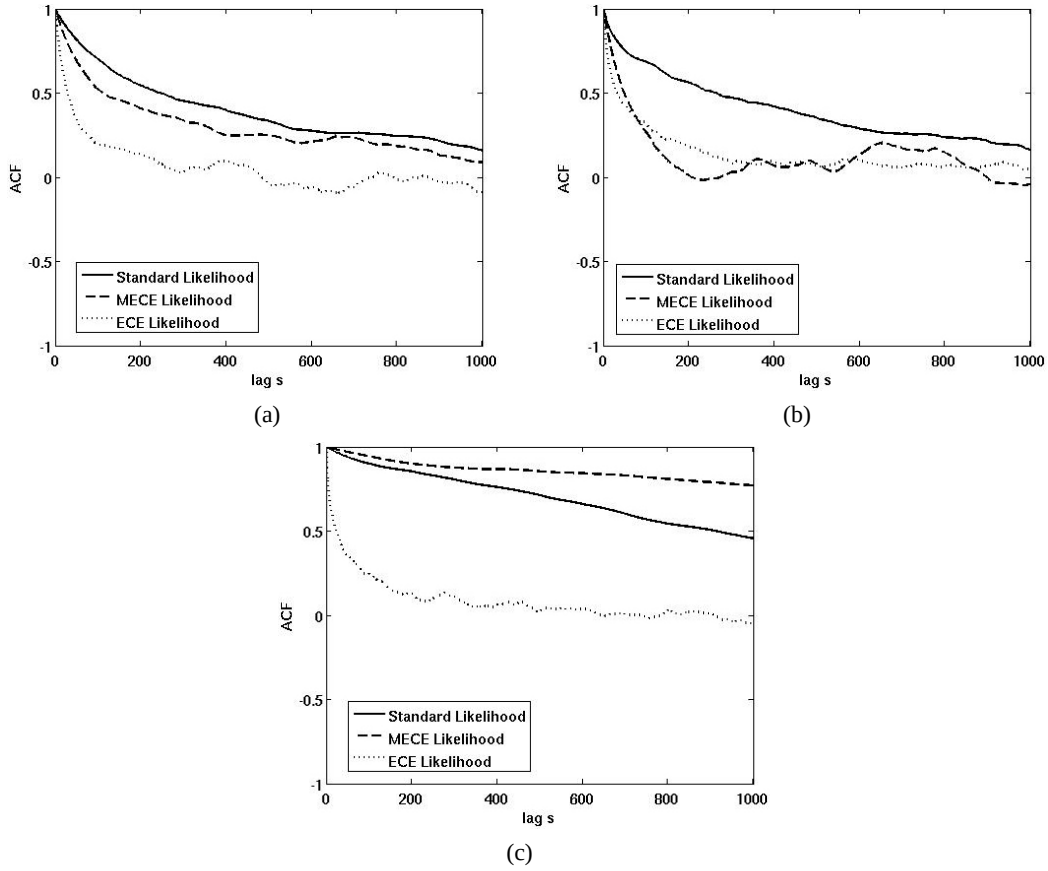
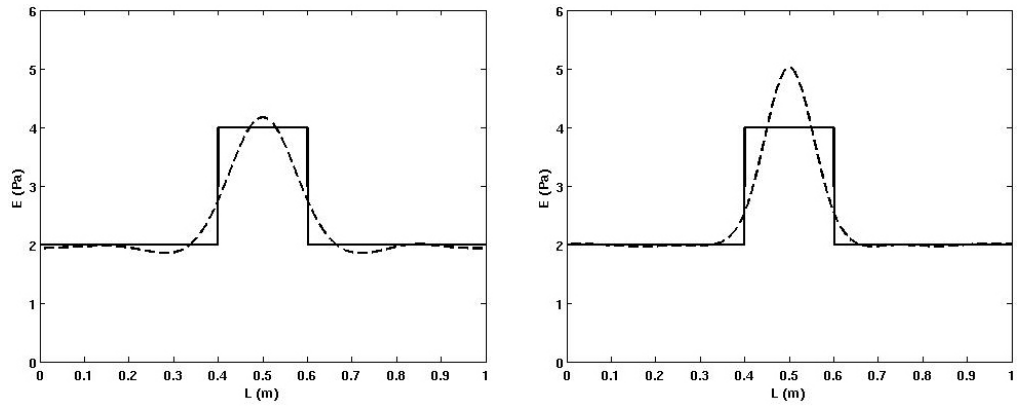


FIG. 8: The autocorrelation function (ACF) at lag s for the parameters of a radial basis function with 5% noise in the data. (a) ACF of the matrix elastic modulus E_c, (b) ACF for the elastic modulus of RBF 1 E_1, (c) ACF for the locality parameter of RBF 1 z_1.

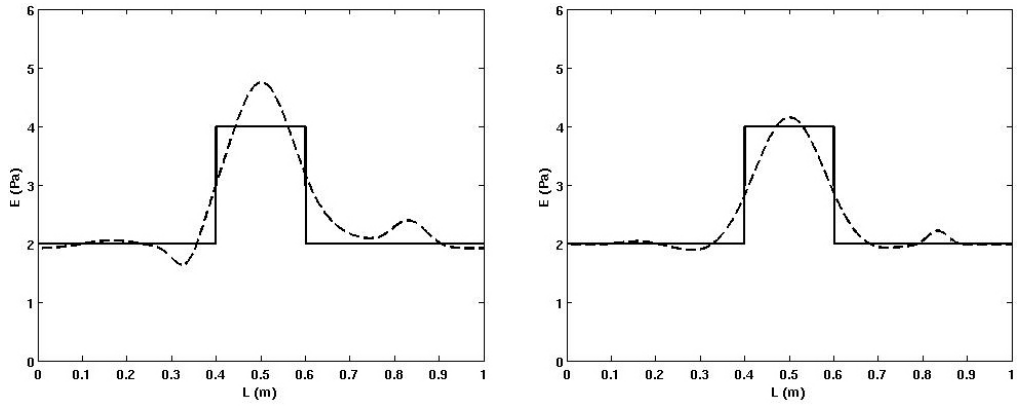
with prior mean $\bar{x} = 0$, likelihood mean $\bar{e} = 0$, $\Gamma_{error} = (\beta/h)\boldsymbol{I}$, and $\Gamma_{prior} = (\kappa/h)\boldsymbol{I}$. Here, $h = 1/1000$ denotes the element size and $\boldsymbol{I}$ is an identity matrix. Furthermore, we set $\beta = 1$ and the parameters κ and γ to the values shown in Table 3.

Figure 10 shows the posterior elastic modulus field predicted with the MAP estimate approach. The mean elastic modulus field E predicted with the standard, MECE, and ECE Bayesian formulations were close to the benchmark elastic modulus field. However, the accuracy of each prediction varied depending on which Bayesian formulation was utilized. In the case where the standard and MECE Bayesian formulations were used, the mean estimates and the standard deviations are almost similar. However, from the constrained optimization results shown in Table 2, we notice that the elastic modulus fields predicted with the MECE formulation were closer to the benchmark solution than the elastic modulus fields predicted with the L^2 formulation, especially as the corruption in the observed data increased. Furthermore, when the problem statistics were computed, the elastic modulus estimates predicted with the L^2 formulation displayed an increase in the parameters standard deviation. One can argue that these results are due to the weak enforcement of both the error in the observed data and the error in the constitutive equation into the variational boundary value problem when the MECE Bayesian formulation is implemented. This weak enforcement of both errors into the variational boundary value formulation seemed to provide a regularization behavior that reduced the uncertainty in the solution.

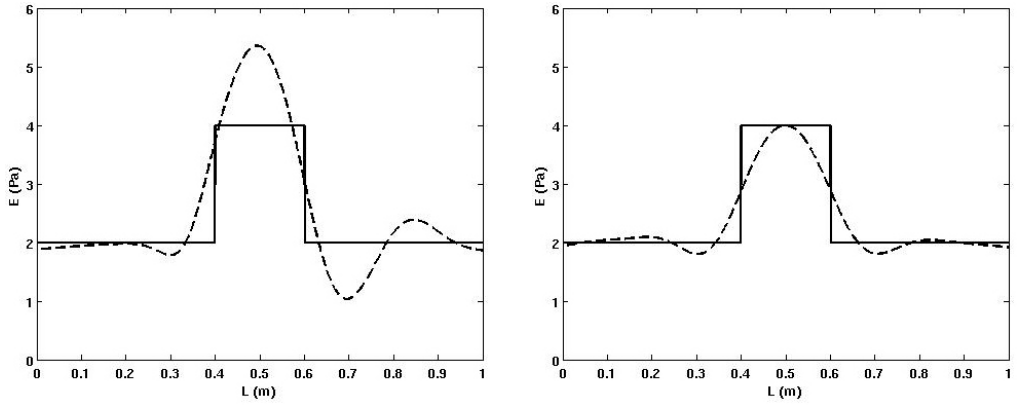
In the case where the ECE Bayesian formulation was used, the elastic modulus field estimates predicted with the ECE Bayesian approach were the most accurate of the three Bayesian MAP formulations studied herein. However,



(a) Standard Likelihood



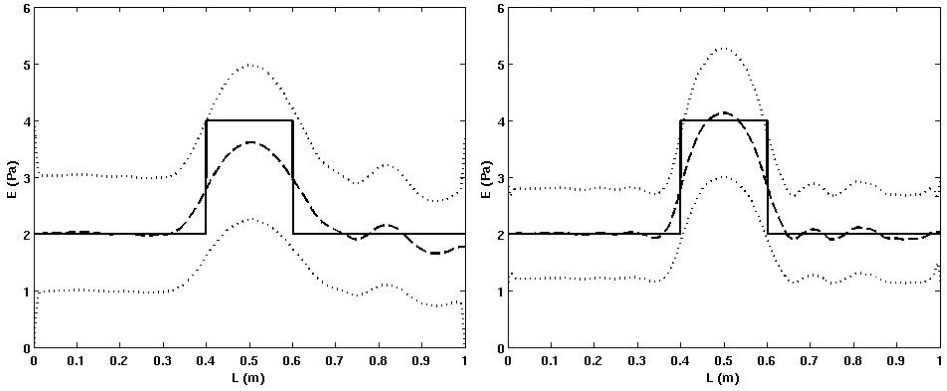
(b) MECE Likelihood



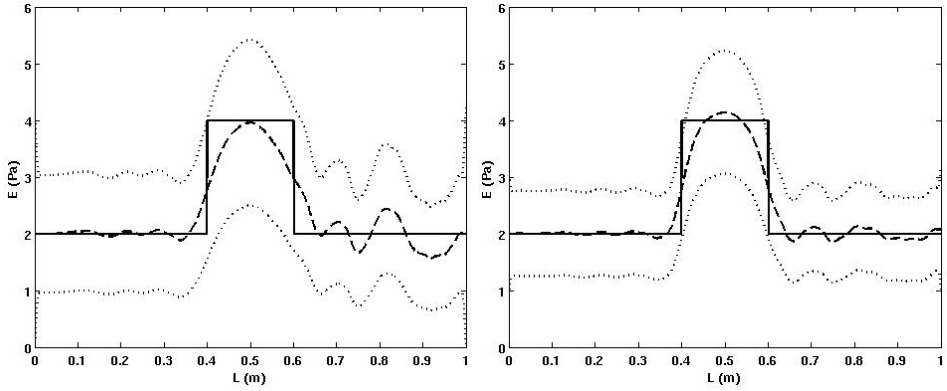
(c) ECE Likelihood

FIG. 9: Bayesian calibration: Elastic modulus field predicted with the RBF parameters mean. Left column: Predicted elastic modulus field with 5% noise in the data. Right column: Predicted elastic modulus field with 1% noise in the data. Here, the solid line denotes the target elastic modulus field, the dashed line denotes the predicted elastic modulus.

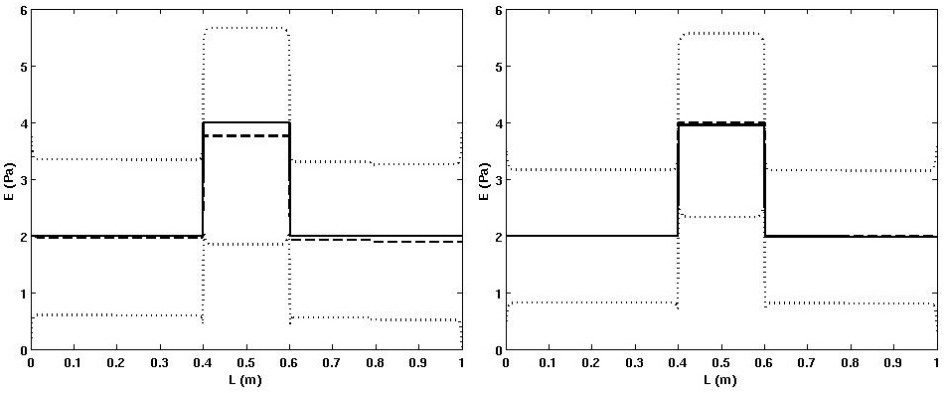
these elastic modulus field estimates displayed higher variance from its mean than the elastic modulus field estimates computed with the standard and MECE Bayesian formulations. One possible explanation for these high variances is the strong imposition of the observed data as Dirichlet conditions. Thus, any uncertainty in the observed data is



(a) Standard Likelihood



(b) MECE Likelihood



(c) ECE Likelihood

FIG. 10: MAP estimate results. Left column: Posterior elastic modulus field with 5% noise in the data. Right column: Posterior elastic modulus field with 1% noise in the data. Here, the solid line denotes the target elastic modulus field, the dashed line denotes the predicted elastic modulus, and the dotted lines denote the elastic modulus upper and lower bounds (i.e., mean ± 2 standard deviation).

directly passed into the deterministic model. Thus, for highly corrupted data, this approach can lead to inaccurate mean estimates as well as high uncertainty in the solution.

Finally, the authors observed that the elastic modulus field estimates were sensitive to the regularization coefficient κ. For instance, for high values of κ, the uncertainty in the mean estimates decreased and for low values of κ it

increased. The uncertainty in the mean estimates is related to the corruption in the observed data. High levels of corruption in the observed data will induce significant errors in the mean estimates. Regularization methods have been developed to deal with these errors. As the corruption in the observed data increases, more regularization is needed in order to prevent overfitting. The direct effect of increasing the regularization coefficient κ is the smoothing of the error surface; thus, the error surface becomes more convex. Hence, the uncertainty in the solution will decrease. However, the accuracy of the mean estimate will suffer due to the smoothing of the error surface. Meanwhile, less regularization is needed as the corruption in the observed data decreases. However, the possibility of multiple solutions satisfying the problem constraints increases for lower values of κ. Thus, the uncertainty in the elastic modulus field estimates will increase.

5. CONCLUSIONS

This paper investigated the problem of parameter estimation in the context of constrained regularized optimization, unconstrained optimization, and stochastic parameter estimation. For the constrained optimization approaches, we investigated three formulations: L^2, MECE, and ECE formulations. We found that the MECE and ECE formulations predicted accurate results in fewer iterations than the L^2 formulation, with the ECE formulation requiring the fewest iterations. However, it is possible that the constrained optimization approach based on the ECE formulation can break down for high levels of noise due to the strong imposition of the observed data as Dirichlet conditions. For this reason, we expect that the MECE formulation will result in more accurate predictions than the ECE formulation when the corruption in the observed data increases. In the context of the regularization methods, it was observed that lower reconstruction errors were obtained when the TV regularization was used instead of Tikhonov regularization. We expected to have such results given that TV regularization is an appropriate regularization method to use for reconstructing discontinuous functions. For the unconstrained nonlinear least-squares approach, it was observed that the elastic modulus field predicted with both the MECE and ECE formulations was more accurate than the elastic modulus field obtained with the standard NLLS formulation. Moreover, the MECE and ECE formulations were more tolerant to corruption in the observed data than the standard NLLS approach.

In the context of Bayesian calibration, the global convergence capability of MCMC algorithms lets all three stochastic models yield elastic modulus field estimates close to the benchmark elastic modulus field. However, the ECE stochastic model required fewer burn-in samples than the standard and MECE stochastic models to reach a stationary distribution of the Markov chain. Moreover, our study showed that the ECE likelihood formulation provides a regularization behavior and improves the convergence of the Markov chain posterior. Yet, as in the deterministic optimization methods case study, it is possible that the ECE likelihood formulation could break down for high levels of noise in the data due to the strong imposition of the observed data as Dirichlet conditions.

For the MAP estimate approach, it was observed that the MECE and ECE formulations produced lower reconstruction errors than the L^2 formulations. Thus, the elastic modulus fields predicted with the stochastic models based on the MECE and ECE formulations were closer to the benchmark solution. Furthermore, the results showed that the MAP estimate approach based on the ECE formulation required the fewest iterations to yield an optimal elastic modulus field. One main criticism of the MAP estimate approach is that it gives point estimates of the posterior parameter distribution, whereas the MCMC approach generates full posterior distribution via sampling. However, the MAP estimate approach can serve as an alternative to MCMC methods for quantifying uncertainty in statistical inverse problems that are governed by expensive numerical models and high-dimensional spaces.

ACKNOWLEDGMENTS

The authors would like to thank the Validation, Verification, and Uncertainty and Quantification (VU) program under the Department of Energy's Nuclear Energy Advanced Modeling and Simulation (NEAMS) program. We would also like to thank Bart Van Bloemen Waanders and Denis Ridzal for fruitful discussions related to this work. Sandia National Laboratories is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract no. DE-AC04-94AL85000.

REFERENCES

1. Liew, H. L. and Pinsky, P. M., Recovery of shear modulus in elastography using an adjoint method with b-spline representation, *Finite Elem. Anal. Design*, 41(7-8):778–799, 2005.
2. Sampath, R. and Zabaras, N., An object-oriented framework for the implementation of adjoint techniques in the design and control of complex continuum systems, *Int. J. Num. Methods Eng.*, 48(2):239–266, 2000.
3. Bonnet, M. and Constantinescu, A., Inverse problems in elasticity, *Inverse Probl.*, 21(2):R1–R50, 2005.
4. Gockenbach, M. S., Numerical analysis of elliptic inverse problems with interior data, *J. Phys. Conf. Ser.*, 124(1):1–12, 2008.
5. Gockenbach, M. S., Jadamba, B., and Khan, A. A., Numerical estimation of discontinuous coefficients by the method of equation error, *Math. Mech. Solids*, 1:343–359, 2006.
6. Gockenbach, M. S. and Khan, A. A., An abstract framework for elliptic inverse problems: Part 2. An augmented Lagrangian approach, *Math. Mech. Solids*, 14(6):517–539, 2009.
7. Kohn, R. V. and Lowe, B. D., A variational method for parameter identification, *Math. Model. Num. Anal.*, 22(1):119–158, 1998.
8. Kohn, R. V. and McKenney, A., Numerical implementation of a variational method for electrical impedance tomography, *Inverse Problems*, 6(3):389–414, 1990.
9. Kohn, R. V. and Vogelius, M., Determining conductivity by boundary measurements, *Commun. Pure Appl. Math.*, 37(3):289–298, 1984.
10. Kohn, R. V. and Vogelius, M., Determining conductivity by boundary measurements II. Interior results, *Commun. Pure Appl. Math.*, 38(5):643–667, 1985.
11. Chouaki, A. T., Ladevèze, P., and Proslier, L., Updating structural dynamic models with emphasis on the damping properties, *Am. Ins. Aeronaut. Astronaut. J.*, 36(6):1094–1099, 1998.
12. Ladevèze, P. and Leguillon, D., Error estimates procedures in the finite element method and applications, *SIAM J. Num. Anal.*, 20(3):485–509, 1983.
13. Deraemaeker, A., Ladevèze, P., and Leconte, P., Reduced bases for model updating in structural dynamics based on the constitutive relation error, *Comput. Methods Appl. Mech. Eng.*, 191(21-22):2427–2444, 2002.
14. Feissel, P. and Allix, O., Modified constitutive relation error identification strategy for transient dynamics with corrupted data: The elastic case, *Comput. Methods Appl. Mech. Eng.*, 196(13-16):1968–1983, 2007.
15. Ladevèze, P. and Leguillon, D., Validation of structural dynamics models containing uncertainty, *Comput. Methods Appl. Mech. Eng.*, 195(4-6):373–393, 2006.
16. Nocedal, J. and Wright, S. J., *Numerical Optimization*, Springer, New York, 2006.
17. Dennis, J. and Lewis, R., A comparison of nolinear programming approaches to an elliptic inverse problem and a new domain decomposition approach, Technical Report TR-9433, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1994.
18. Kunisch, K. and Sachs, E., Reduced SQP methods for parameter identification problems, *SIAM J. Num. Anal.*, 29:1793–1820, 1992.
19. Orozco, C. and Ghattas, O., Massively parallel aerodynamic shape optimization, *Comput. Syst. Eng.*, 3(1-4):311–320, 1992.
20. Biros, G. and Ghattas, O., Parallel Lagrange Newton-Krylov-Schur methods for PDE-constrained optimization. Part I: The Krylov-Schur solver, *SIAM J. Sci. Comput.*, 27(2):687–713, 2005.
21. Biegler, L., Nocedal, J., and Schmid, C., A reduced Hessian method for large-scale constrained optimization, *SIAM J. Optim.*, 5:314–347, 1995.
22. Biros, G. and Ghattas, O., Parallel Newton-Krylov Methods for PDE-constrained optimization, In *Proceedings of Supercomputing 99*, Portland, Oregon, USA, November 13–19, 1999.
23. Gill, P. and Wong, E., Sequential Quadratic Programming methods, Technical Report NA-10-03, Department of Mathematics, UCSD, San Diego, CA, 2010.
24. Prudencio, E., Byrd, R., and Cai, X.-C., Parallel full-space SQP Lagrange-Newton-Krylov-Schwarz algorithm for PDE-constrained optimization problems, *SIAM J. Sci. Comput.*, 27(4):1305–1328, 2006.
25. Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., *Markov Chain Monte Carlo in Practice*, Chapman & Hall London,

1998.

26. Robert, C. P. and Casella, G., *Monte Carlo Statistical Methods*, Springer-Verlag, Berlin, 1999.
27. Akçelik, V., Flath, H. P., Ghattas, O., Hill, J., Van Bloemen Waanders, B., and Wilcox, L. C., Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations, *SIAM J. Sci. Comput.*, 33:407–432, 2011.
28. Aster, R., Borchers, B., and Thurber, C., *Parameter Estimation and Inverse Problems*, Elsevier, New York, 2005.
29. Siltanen, S., Kolehmainen, V., Ärvenpää, S., Kaipio, J. P., Koistinen, P., Lassas, M., Pirttilä, J., and Somersalo, E., Statistical inversion for medical x-ray tomography with few radiographs: I. General theory, *Phys. Med. Biol.*, 48(10):1437–1463, 2003.
30. Carr, J. C., Fright, W. R., and Beatson, R. K., Surface interpolation with radial basis functions for medical imaging, *IEEE Trans. Med. Imaging*, 16(1):96–107, 1997.
31. Fornefett, M., Rohr, K., and Stiehl, H. S., Radial basis functions with compact support for elastic registration of medical images, *Image Vis. Comput.*, 19(1-2):87–96, 2001.
32. Prudencio, E. and Cai, X.-C., Parallel multilevel restricted Schwarz preconditioners with pollution removing for PDE-constrained optimization problems, *SIAM J. Sci. Comput.*, 29(3):964–985, 2007.
33. Hinze, M., Pinnau, R., Ulbrich, M., and Ulbrich, S., *Optimization with PDE Constraints*, Springer, Berlin, 2009.
34. Gunzburger, M. D., *Perspectives in Flow Control and Optimization*, SIAM, Philadelphia, 1st ed., 2003.
35. Kaipio, J. and Somersalo, E., *Statistical and Computational Inverse Probelms*, Springer, New York, 2005.
36. Tarantola, A., *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, 2005.
37. Epanomeritakis, I., Akcelik, V., Ghattas, O., and Bielak, J., A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion, *Inverse Problems*, 24(3):1–26, 2008.
38. Adams, B. M., Bohnhoff, W. J., Dalbey, K. R., Eddy, J. P., Eldred, M. S., Gay, D. M., Haskell, K., Hough, P. D., and Swiler, L. P., *DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.0 Users Manual*, December 2010.
39. Gockenbach, M. S. and Khan, A. A., In *Mathematical Models and Methods for Real World Systems*, A convex objective functional for elliptic inverse problems, Chapman & Hall CRC Taylor & Francis Group, London, 2006.
40. Gockenbach, M. S. and Khan, A. A., An abstract framework for elliptic inverse problems: Part 1. An output least-squares approach, *Math. Mech. Solids*, 12(3):259–276, 2007.
41. Allix, O., Feissel, P., and Nguyen, H. M., Identification strategy in the presence of corrupted measurements, *Eng. Comput.*, 22(5-6):487–504, 2005.
42. Vogel, C. R., *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.
43. Haario, H., Laine, M., Mira, A., and Saksman, E., DRAM: Efficient adaptive MCMC, *Stat. Comput.*, 16(4):339–354, 2006.
44. Haario, H., Saksman, E., and Tamminen, J., An adaptive metropolis algorithm, *Bernoulli*, 7(2):223–242, 2001.
45. Mira, A., On Metropolis-Hastings algorithm with delayed rejection, *Metron*, 59(3-4):231–241, 2001.
46. Ma, X. and Zabaras, N., An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method, *Inverse Problems*, 25(3), ID-035313, 2009.
47. Box, G. and Jenkins, G., *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1976.