A Fully Bayesian Extension to FEMU for Identification of Spatially Varying Elastic Properties from Digital Image and Volume Correlation Measurements

Armand Touminet, Sabine Cantournet, Victor Fabre, Pierre Kerfriden

A Fully Bayesian Extension to FEMU for Identification of Spatially Varying Elastic Properties from Digital Image and Volume Correlation Measurements

Armand Touminet[1,2], Sabine Cantournet[1], Victor Fabre[2], and Pierre Kerfriden[1]

[1]Mines Paris, Centre des Matériaux, 91100, France
[2]Hutchinson, Centre de Recherche et Innovation, Chalette-sur-Loing, 45120, France

Abstract

We present a fully Bayesian framework for identifying spatially varying elastic parameters and their covariance properties using noisy displacement observations obtained with DIC or DVC trials. Our method is a generalization of identification procedures such as FEMU or I-DIC to materials with spatially varying properties and stochastic mesostructures. The identified variables can be used to generate new samples with similar covariance properties. We formulate the approach as a hierarchical Bayesian PDE-constrained inverse problem and MAP estimates are obtained through gradient based optimization. We resort to an adjoint based formulation and leverage automatic differentiation to derive the parameter sensitivities. We show how modelling unknown parameters with Gaussian Random Fields leads to a natural Bayesian regularization and develop the use of Whittle-Matérn priors. Hyperparameter estimation is discussed, and we propose an empirical Bayes approach to avoid numerical shortcomings related to a standard hierarchical model. A set of numerical examples is presented to assess the performance of the proposed method, based on synthetic data generated through Matérn Random fields. In particular, we show how data noise is naturally modelled by the Bayesian formulation and impacts spatial covariance of identified parameters.

1 Introduction

Full field identification procedures from in-situ observations such as Digital Image Correlation (DIC), stereo-DIC, or Digital Volume Correlation (DVC) have widely been developed in the past decades to calibrate constitutive laws of a wide range of materials. Examples of such identification procedures include the Finite Elements Updating Method (FEMU), Integrated DIC (I-DIC) [Leclerc et al., 2009] or Mechanical Image Correlation (MIC) [Réthoré, 2010]. In these methods, a digital twin of a mechanical trial is developed, and the simulation to experiment gap is iteratively minimized by an optimization algorithm [Avril et al., 2008]. Usually, the parameters of interest are a set of scalar variables from a complex and non-linear constitutive law [Réthoré, 2010, Leclerc et al., 2009, Neggers et al., 2017]. Typically, the identified parameters are assumed to be constant across the specimen of interest, which is suitable to study homogeneous materials, or heterogeneous materials which mechanical properties present no significant variability at the macroscopic scale. In this work, we aim at extending those identification procedures to materials with a significant spatial variability of mechanical properties or with stochastic micro- or mesostructures. For instance, discontinuous composites include a large range of materials with such properties, including compression moulded Discontinuous Long Fiber composites (DLFs) with prepreg based Sheet Moulding Compounds (SMCs) or thermoplastic Tow Based Discontinuous Composites (TBDCs), and injection moulded short fiber composites. In these materials, a strong spatial variability is observed due to random fiber bundle layout, and fiber realignment during moulding [Martulli et al., 2019], which are of great importance for modelling such materials [Alves et al., 2023]. Other examples include concrete, where strength has been shown to be spatially variable [Geyer et al., 2023], or functionally graded materials [Watanabe et al., 2008], where mechanical parameters are purposely continuously varying across a given specimen. Although the authors aim at applying the present developments to DLF composites, this paper remains purely

numerical and material agnostic, and application to a specific material is deferred to subsequent works.

Extending identification procedures to spatially varying parameters introduces numerous numerical challenges to overcome. First, a proper spatial discretization of the parameters of interests is introduced, resulting in a high number of variables to optimize. In typical FEMU or I-DIC methods, the simulation to experiment gap is minimized with a quasi-Newton algorithm, and the gradient vector is usually approximated with finite differences, and requires solving multiple linear systems for each evaluation of the objective function. With a fine spatial discretization for the variable to optimize, this process becomes numerically intractable, and in this paper we leverage an adjoint based approach to compute the gradient vector efficiently, independently of the number of variables to optimize. Secondly, parameter identification from displacement observations is known to be an ill-posed inverse problem, and requires spatial regularization to avoid local minima of the objective function and to avoid overfitting observation noise. The use of spatial regularization, such as the well known Tikhonov regularizer family, introduces so-called regularization parameters that are often cumbersome to select, although some selection methods exist, such as those based on the Morozov principle [Nair, 2009], or the L-curve approach [Hansen, 1992]. The present framework is formulated as a Bayesian inverse problem, meaning that both parameters of interest and observation noise are modelled as random variables, and we derive Maximum A Posteriori (MAP) estimates for parameter identification. We use Gaussian priors to model the sought parameters and leverage covariance matrices derived from Matérn class kernels, which have the advantage of having statistically and geometrically interpretable parameters, namely the prior variance and length scale. More specifically, we use stationary Whittle-Matérn priors, which have been shown to be very numerically efficient to implement by following the SPDE approach [Lindgren et al., 2011], which leads to a sparse approximation of the precision matrix on unstructured grids. For hyperparameter selection, we propose a Bayesian approach by introducing a hierarchical model, which is in turn approximated following a Parametric Empirical Bayes (PEB) approach. Three choices of estimators are investigated for hyperparameter update, namely a maximum likelihood estimator, a maximum a posteriori estimator, and an empirical estimator based on variogram regression. We discuss the numerical issues and shortcomings of each approach and link numerical results to the existing literature. In particular, we show that a maximum likelihood update strategy alongside a Whittle-Matérn prior leads to an unstable numerical scheme, and propose a numerically robust alternative. We assess the performance of the PEB approach on synthetic samples and show the impact of data noise on estimated parameters and hyperparameters.
Although identification from surface (DIC) and volume (DVC) measurements typically constitute separate fields of research, we abstract the underlying method used to obtain displacement fields from image correlation, and treat noisy surface or volume displacement fields as input to our framework, that is presented hereafter to handle both types of measurements.
For implementation, we leverage automatic differentiation to derive the adjoint operator automatically and independently of the chosen stiffness model and show examples of parameter identification on synthetic data with a selected set of stiffness models. In the present paper, we develop the proposed method in a stiffness model agnostic way. However, for simplicity, we assume that the stiffness model is linear, though it could be adapted to non-linear constitutive laws with minor modifications only.

2 Related Works and Main Contributions

Following the developments of Digital Image Correlation (see, e.g., Sutton et al. [2009] for details about common image correlation techniques), many kinds of parameters identification methods were developed, among which we distinguish non-integrated procedures, such as FEMU, in which the simulation to experiment gap is formulated in terms of a displacement misfit, and integrated procedures, such as I-DIC, in which an image residual is considered to perform the identification, removing the need to obtain correlated displacement fields beforehand. Many other methods exist, and for a comprehensive review, we refer to [Avril et al., 2008]. Although the differences between all of these methods are not important for the present paper, it was shown that data noise can impact the identification results significantly, especially on setups with a low signal-to-noise ratio [Ruybalid et al.]. An interesting variation is the regularized FEMU (FEMU-R) method, in which the displacement gap

is weighted by a displacement noise correlation matrix [Leclerc et al., 2009, Réthoré, 2010], leading to the same data fidelity term as typically used in MAP estimates, hinting towards a Bayesian approach, although it was never formulated as such. In the aforementioned works, the cost function is commonly minimized by the means of a quasi-Newton algorithm with a Gauss-Newton approximation for the hessian, and the gradient vector is usually approximated by finite differences.
The adjoint method has been used to solve inverse problems in linear elasticity, in a context unrelated to DIC or DVC, in e.g. [Seidl et al., 2019], to identify a piecewise constant shear modulus, and also resorts to total variation (TV) regularization to enforce piecewise constant solutions. In [Bonnet and Constantinescu, 2005], the adjoint method is briefly presented to identify an unknown Poisson ratio distribution, however the questions of regularization and data noise are not investigated thoroughly. Also, in these last references, the spatial covariance structure of the unknown parameter fields are not investigated.
The use of Gaussian random fields for modelling spatial data has been developed and applied to a wide range of data, including soil properties [Ricketts et al., 2023], concrete failure parameters [Hai and Lyu, 2023], or even laminate composite tensile modulus [Sriramula and Chryssanthopoulos, 2009]. In such cases, random fields models can be used to generate new realisations of the underlying spatial process, and are a good alternative to Monte-Carlo simulations that can be cumbersome. For this purpose, a proper selection of covariance parameters is key, as it can highly influence the simulation results.
In the Bayesian inversion literature, identification of a spatially varying unknown parameter is a common problem, and the use of so-called Bayesian priors enforce statistical information about the solution. For a review of common methods and algorithms, we refer to [Dashti and Stuart, 2015]. A specific challenge arising when the unknown is discretized on an unstructured mesh, which is common when dealing with PDE-constrained inverse problems, is that evaluating the prior term typically involves assembling a dense matrix, which has a quadratic complexity in time and memory and is often intractable in large scale problems. This was partially solved in [Lindgren et al., 2011] by proposing a sparse approximation of precision matrices arising from some cases of Matérn class correlation kernels. In later works [Roininen et al., 2014], such priors were used to identify unknown PDE parameters through MAP estimates. In more recent works [Roininen et al., 2019], the case where Matérn prior hyperparameters are spatially varying was studied, and a hierarchical model based on Cauchy hyperpriors was proposed to estimate these hyperparameters through Markov Chain Monte Carlo (MCMC) sampling. However, papers tackling approximation of Bayesian priors on unstructured meshes are largely unrelated to parameter estimation from DIC or DVC measurements.
In the present paper, we bridge the gap between all the previously mentioned topics to construct a comprehensive framework to model spatially varying elastic properties with Gaussian random fields, and perform parameter and hyperparameter identification from DIC and DVC measurements:

- We extend the FEMU method to a fully Bayesian formulation that takes into account all sources of measurement noise,
- An optimization strategy based on the adjoint method and automatic differentiation is presented to derive maximum a posteriori estimates,
- We leverage Whittle-Matérn priors to model the covariance structure of unknown parameters and obtain an efficient discretization scheme,
- A hierarchical model is presented for hyperparameter estimation, and a numerical method based on the PEB approach is presented and thoroughly discussed.
- The results can be used to generate new realisations of the spatially varying parameter, with estimated covariance parameters.

3 Notations

Throughout this article, and unless specified otherwise, zeroth, first, second and fourth order tensors and tensor valued functions are denoted by a, $\underline{a}$, $\underline{\underline{a}}$ and $\overline{a}$ respectively. The simple and double contraction are written $\cdot$ and $:$ respectively, and we employ the Einstein notation to express them in a Cartesian basis:

$$\underline{a} \cdot \underline{b} = a_i b_j, \ \underline{\underline{a}} : \underline{\underline{b}} = a_{ij} b_{ij}, \ \overline{a} : \underline{\underline{b}} = a_{ijkl} b_{kl}.$$

If x is a continuous random variable, we note $D(x)$ the density of x. If x follows a normal or uniform law, we respectively note $x \sim \mathcal{N}(x^0, \sigma_x^2)$ where x^0 is the mean and σ_x^2 the variance, and $x \sim \mathcal{U}(I)$ where $I \subset \mathbb{R}$ is a bounded interval.
If U and V are two Banach spaces, and $F : U \to V$ is a differentiable function, the Gateaux derivative of F at point $x \in U$ along the direction $\delta x \in U$ is denoted by $\mathrm{d}F(x; \delta x)$. If $U_1, ..., U_n$ are Banach spaces, and $G : U_1 \times ... \times U_n \to V$ is differentiable function, we note $\partial_i G(x_1, ..., x_n; \delta x_i)$ its Gateaux derivative with the respect to its i-th variable, defined as

$$\partial_i G(x_1, ..., x_n; \delta x_i) = \lim_{h \to 0} \frac{G(x_1, ..., x_i + h\delta x_i, ..., x_n) - G(x_1, ..., x_n)}{h}.$$

If U_1, U_2 are of finite dimension, bilinear forms $a : U_1 \times U_2 \to \mathbb{R}$ are identified with their matrix representations, denoted with a bold upper case letter $\boldsymbol{A}$, and linear forms $a : U_1 \to \mathbb{R}$ with their vector representation, by a bold lower case letter $\boldsymbol{a}$. Given bases for U_1 and U_2, $\boldsymbol{a} \in \mathbb{R}^{\dim(U_1)}$ and $\boldsymbol{A} \in \mathbb{R}^{\dim(U_1) \times \dim(U_2)}$. Given an open and bounded domain $\Omega \subset \mathbb{R}^n$, $\partial\Omega = \overline{\Omega} \setminus \Omega$ denotes the boundary of Ω, and we introduce the common Hilbert and Sobolev spaces used in variational analysis, namely $L^\infty(\Omega)$ the space of bounded functions over Ω, $L^2(\Omega)$ the space of square integrable functions over Ω and $H^1(\Omega)$.

4 Stiffness Modelling with Gaussian Random Fields

Throughout this paper, we consider on open and bounded domain $\Omega \subset \mathbb{R}^3$ representing a specimen made of an abstract material with spatially varying elastic properties, in the sense that the stiffness tensor $\overline{C}(\underline{x})$ is spatially varying. We also assume that the stiffness tensor field can be parameterized with a set of N_p scalar valued parameter fields $(p_1(\underline{x}), ..., p_{N_p}(\underline{x})) \in \mathcal{D} \subset \mathbb{R}^{N_p}$ (e.g. isotropy, transverse isotropy, etc.), where $\mathcal{D}$ is called the *parameter space*. The mapping $(p_1, ..., p_{N_p}) \mapsto \overline{C}(p_1, ..., p_{N_p})$ is hereafter referred to as the *stiffness model*, and is assumed to be well-defined, in the sense that $\overline{C}(\mathcal{D})$ is a set containing tensors that satisfy the usual symmetry assumptions in linear elasticity. We further assume that spatial variations are due to some unknown spatial random process, and model the stiffness parameters as Gaussian random fields:

$$\forall 1 \leq i \leq N_p, \; p_i \sim \mathcal{N}(p_i^0, c_i), \tag{1}$$

where p_i^0 is known and c_i is an unknown covariance function. By combining the stiffness model with a specified covariance model for the stiffness parameters, it is possible to sample realizations of a synthetic specimen that accounts for spatial variability and randomness of the underlying material. The goal of this paper is to infer estimates for stiffness parameters and covariance models from displacement measurements only. The inferred model can then be used to generate numerous specimens for simulations. The method is formulated as a Bayesian inverse problem. Section 5 presents how the inverse problem is formulated and derives MAP estimates for the unknown stiffness parameters and covariance hyperparameters, calibration of the covariance model is detailed in section 6, section 8 presents how the adjoint method is used to solve the inverse problem, and section 9 illustrates the proposed method with two numerical examples.

5 Problem Description and Bayesian Modelling

In this section, we present the steps to extend the FEMU method to identify a spatially varying material parameter, with a fully Bayesian approach. The developments presented throughout this paper can be adapted to the I-DIC method in the same way, and for completeness, details are given in appendix A.

5.1 PDE Modelling and Discretization

We consider the spatial domain Ω to represent a trial specimen subject to an (unknown) boundary displacement $\underline{u}_D$ over $\Omega_D \subset \partial\Omega$, modelling some mechanical trial. We also consider an *observation zone* $\Omega^{\text{obs}} \subset \Omega$ modelling a domain where in-situ imaging is performed. For DIC, $\Omega^{\text{obs}} \subset \partial\Omega$

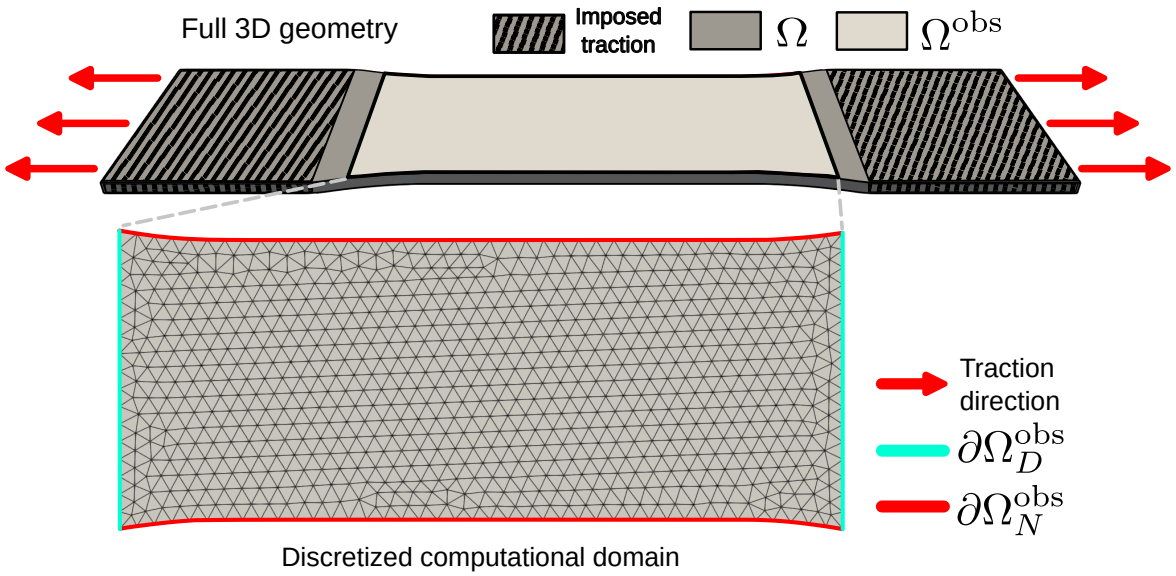


Figure 1: The two considered specimen: a thin dog bone geometry for uniaxial tensile test, and a thick biaxial geometry.

represents the specimen surface captured by the camera, as for DVC, Ω^{obs} represents a volume subregion imaged through, e.g., x-ray tomography. We represent in figure 1 and 11 two specimen geometries where the different regions are annotated, illustrating both surface and volume observation zones. We model the results of image correlation by a set of pointwise, noisy displacement observations

$$\boldsymbol{u}^{\text{obs}} = \left(\underline{u}^{\text{obs}}(\underline{x}_1), ..., \underline{u}^{\text{obs}}(\underline{x}_{N^{\text{obs}}})\right), \tag{2}$$

$$\underline{u}^{\text{obs}}(\underline{x}_i) = \underline{u}(\underline{x}_i) + \underline{\xi},\ 1 \leq i \leq N^{\text{obs}}, \tag{3}$$

where N^{obs} is the number of pointwise observations, and $\boldsymbol{u}^{\text{obs}}$ is a vector of size $3 \times N^{\text{obs}}$. Here, $\underline{u}$ denotes the true displacement field, and $\underline{\xi}$ some observation noise. The observed macroscopic force $F^{\text{obs}} \in \mathbb{R}$ models the measured force F, typically by a force sensor, that is also corrupt by some noise η, such that

$$F^{\text{obs}} = F + \eta. \tag{4}$$

Without loss of generality, we assume that there is only a single force sensor. In other cases (e.g. biaxial traction), F should be taken as vector valued. Let $\boldsymbol{y}^{\text{obs}}$ and $\boldsymbol{\epsilon}$ be vectors of size $3 \times N^{\text{obs}} + 1$ collecting all measured values and noise values respectively:

$$\boldsymbol{y}^{\text{obs}} = \left(\underline{u}^{\text{obs}}(\underline{x}_1), ..., \underline{u}^{\text{obs}}(\underline{x}_{N^{\text{obs}}}), F^{\text{obs}}\right), \tag{5}$$

$$\boldsymbol{\epsilon} = \left(\underline{\xi}(\underline{x}_1), ..., \underline{\xi}(\underline{x}_{N^{\text{obs}}}), \eta\right). \tag{6}$$

We assume that $\boldsymbol{y}^{\text{obs}}$ depends on a scalar valued material parameter field p, defined on some Hilbert space V, through some non-linear operator $\mathcal{A}$ defined on V as

$$\boldsymbol{y}^{\text{obs}} = \mathcal{A}(p) + \boldsymbol{\epsilon}. \tag{7}$$

Before going into the Bayesian formulation details, we introduce a digital twin of the experiment, which is modelled as the static equilibrium boundary value problem over Ω^{obs}
Find $\underline{u} \in H^1(\Omega^{obs})$ *such that*

$$\nabla \cdot \underline{\underline{\sigma}} = \underline{0} \text{ over } \Omega^{\text{obs}}, \tag{8}$$

$$\underline{\underline{\sigma}} = \overline{\overline{C}}(p) : \underline{\underline{\varepsilon}}, \tag{9}$$

$$\underline{\underline{\varepsilon}} = \frac{\nabla \underline{u} + (\nabla \underline{u})^T}{2}, \tag{10}$$

$$\underline{u} = \underline{u}_D^{\text{obs}} \text{ over } \partial\Omega_D^{\text{obs}}, \tag{11}$$

$$\underline{\underline{\sigma}} \cdot \underline{n} = \underline{0} \text{ over } \partial\Omega_N^{\text{obs}}, \tag{12}$$

where $\underline{u}$ is the displacement field, $\underline{\underline{\varepsilon}}$ and $\underline{\underline{\sigma}}$ are respectively the strain and stress tensor fields, $\overline{C}(p) \in L^\infty(\Omega^{\text{obs}})$ is the stiffness tensor, and $\underline{u}_D^{\text{obs}}$ is a prescribed displacement boundary condition over $\partial\Omega_D^{\text{obs}} \subset \partial\Omega^{\text{obs}}$, and $\underline{n}$ is a outward boundary normal unit vector. Here, we emphasize on the fact that $\overline{C}(p)$ is *not* assumed to be constant over Ω^{obs}. To keep notations concise, and without loss of generality, the stiffness model is assumed to depend on a single parameter p, and is assumed to be differentiable. The stiffness model is typically material specific, and to remain as general as possible, no further assumption is made in this section. The considered constitutive relation (9) is linear as we're only interested in identification of elastic properties, and extensions to other cases are out of scope of this paper. The measured force is then modelled as the average stress over the traction direction, which can be expressed as a volume integral as

$$F = \frac{|\partial\Omega_D^{\text{obs}}|}{|\Omega^{\text{obs}}|} \int_{\Omega^{\text{obs}}} \underline{n}_F \cdot \underline{\underline{\sigma}}(\underline{u}(p)) \cdot \underline{n}_F \, \mathrm{d}\underline{x}, \tag{13}$$

where

$$|\partial\Omega_D^{\text{obs}}| = \int_{\partial\Omega_D^{\text{obs}}} \mathrm{d}\underline{S}, \tag{14}$$

$$|\Omega^{\text{obs}}| = \int_{\Omega^{\text{obs}}} \mathrm{d}\underline{x}, \tag{15}$$

and $\underline{n}_F$ is a unit vector aligned with the loading direction. As usual, we introduce the following Sobolev function spaces

$$H_D^1(\Omega^{\text{obs}}) = \left\{ \underline{v} \in H^1(\Omega^{\text{obs}}) \;\middle|\; \underline{v}|_{\partial\Omega_D^{\text{obs}}} = \underline{u}_D \right\}, \tag{16}$$

$$H_0^1(\Omega^{\text{obs}}) = \left\{ \underline{v} \in H^1(\Omega^{\text{obs}}) \;\middle|\; \underline{v}|_{\partial\Omega_D^{\text{obs}}} = \underline{0} \right\}, \tag{17}$$

and reformulate equations 8-11 in their weak form:
Find $\underline{u} \in H_D^1(\Omega^{obs})$ *and* $\underline{v} \in H_0^1(\Omega^{obs})$ *such that*

$$a(\underline{u}, \underline{v}) = l(\underline{v}), \tag{18}$$

$$a(\underline{u}, \underline{v}) = \int_\Omega \underline{\underline{\varepsilon}}(\underline{u}) : \overline{C}(p) : \underline{\underline{\varepsilon}}(\underline{v}) \, \mathrm{d}\underline{x}, \tag{19}$$

$$l(\underline{v}) = 0. \tag{20}$$

The computational domain Ω is discretized using a standard finite element mesh, which we enforce to be conforming with the boundary of the observation zone $\partial\Omega_{\text{obs}}$. We consider a mesh of Ω^{obs} made of N_T cells $(\mathcal{T}_k)_{0 \le k < N_T}$, and introduce the first-order polynomial discretization space,

$$V(\Omega^{\text{obs}}) = \left\{ u \in \mathcal{C}^0(\Omega^{\text{obs}}) \;\middle|\; \forall 0 \le k < N_T,\ u|_{\mathcal{T}_k} \in P^1 \right\}, \tag{21}$$

where P^1 denotes the space of first order polynomials. To keep notations simple, the P^1 discretization space for vector valued functions is noted $\underline{V}(\Omega^{\text{obs}})$. We introduce finite elements bases for $V(\Omega^{\text{obs}})$ and $\underline{V}(\Omega^{\text{obs}})$:

$$V(\Omega^{\text{obs}}) = \text{span}\left(\varphi_0, ..., \varphi_{N_V}\right), \tag{22}$$

$$\underline{V}(\Omega^{\text{obs}}) = \text{span}\left(\underline{\varphi}_0, ..., \underline{\varphi}_{N_{\underline{V}}}\right). \tag{23}$$

With these notations, the simulated displacement and unknown parameter are discretized in $\underline{V}(\Omega^{\text{obs}})$ and $V(\Omega^{\text{obs}})$ respectively:

$$\underline{u} = \sum_{k=0}^{N_{\underline{V}}} u_k \underline{\varphi}_k, \;\; p = \sum_{k=0}^{N_V} p_k \varphi_k. \tag{24}$$

We note $\boldsymbol{u} = \left(u_0, ..., u_{N_{\underline{V}}}\right)$ and $\boldsymbol{p} = (p_0, ..., p_{N_V})$. For simplicity, we assume that observation points coincide with mesh nodes, i.e. degrees of freedom coordinates of space $\underline{V}(\Omega^{\text{obs}})$ are $\underline{x}_1, ..., \underline{x}_{N^{\text{obs}}}$, such that the notation $\boldsymbol{u}^{\text{obs}} - \boldsymbol{u}$ makes sense. With regard to the previously introduced operator, we pose $\boldsymbol{y} = (u_0, ..., u_{N_V}, F) = \mathcal{A}(p)$.

5.2 Bayesian Formulation

In a Bayesian inference setting, the problem of parameter identification is to estimate $\boldsymbol{p}$ given observations $\boldsymbol{y}^{\text{obs}}$. Following the Bayesian approach, we model both $\boldsymbol{p}$ and $\boldsymbol{\epsilon}$ as random variables. First, we assume that the observation noise ξ and η are Gaussian:

$$\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_u}), \tag{25}$$

$$\eta \sim \mathcal{N}(0, \sigma_F^2), \tag{26}$$

and we note

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_y}), \tag{27}$$

where $\boldsymbol{\Sigma_u}$ is the displacement noise covariance matrix, σ_F^2 is the measured force uncertainty variance, and $\boldsymbol{\Sigma_y}$ is the global noise covariance matrix, defined blockwise as

$$\boldsymbol{\Sigma_y} = \begin{pmatrix} \boldsymbol{\Sigma_u} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_F^2 \end{pmatrix}. \tag{28}$$

We assume $\boldsymbol{\Sigma_y}$ to be fully known, and its exact expression depends on the experimental setup and to the algorithm used to perform the correlation. For instance, when performing global DIC, the displacement noise is not a white noise, but instead is spatially correlated, and closed form expressions can be obtained in some cases. We give details and point to the relevant literature in appendix B. Following the Bayesian approach, we are interested in the posterior distribution of parameter $\boldsymbol{p}$, given observations $\boldsymbol{y}^{\text{obs}}$, which, using the Bayes theorem, can be expressed as follows:

$$D(\boldsymbol{p}|\boldsymbol{y}^{\text{obs}}) \propto D(\boldsymbol{y}^{\text{obs}}|\boldsymbol{p}) D(\boldsymbol{p}), \tag{29}$$

where $D(\boldsymbol{y}^{\text{obs}}|\boldsymbol{p})$ is the likelihood, and $D(\boldsymbol{p})$ is the prior density, which remains to be specified. The posterior density can be sampled from using a Markov chain Monte Carlo (MCMC) algorithm, which is useful to perform uncertainty quantification. However, in the present paper we are only interested in point estimates of the unknown parameter, which are usually obtained by the means of a *Maximum A Posteriori* (MAP) estimator, which maximises the posterior density, and can be computed by the following optimization problem

$$\boldsymbol{p}_{\text{MAP}} = \underset{\boldsymbol{p}}{\text{argmin}} \; -\log D(\boldsymbol{y}^{\text{obs}}|\boldsymbol{p}) - \log D(\boldsymbol{p}). \tag{30}$$

Using assumptions (7) and (27), the likelihood can be expressed as

$$D(\boldsymbol{y}^{\text{obs}}|\boldsymbol{p}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{y}^{\text{obs}} - \boldsymbol{y})^T \boldsymbol{\Sigma_y}^{-1} (\boldsymbol{y}^{\text{obs}} - \boldsymbol{y})\right]. \tag{31}$$

As explained in the introduction, in order to model spatial variability of elastic properties, we model p as a stationary Gaussian random field, which translates to a multivariate Gaussian prior on $\boldsymbol{p}$:

$$\boldsymbol{p} \sim \mathcal{N}(\boldsymbol{p}^0, \boldsymbol{\Sigma_p}(\boldsymbol{\mu})), \tag{32}$$

where the mean $\boldsymbol{p}^0$ is assumed to be known, $\boldsymbol{\Sigma_p}$ is the prior covariance matrix, and $\boldsymbol{\mu}$ is a set of covariance parameters, referred to as hyperparameters for the rest of this paper. In this section, $\boldsymbol{\mu}$ is assumed to be known. Formula (30) now writes

$$\boldsymbol{p}_{\text{MAP}} = \underset{\boldsymbol{p}}{\text{argmin}} \; \frac{1}{2}(\boldsymbol{y}^{\text{obs}} - \boldsymbol{y})^T \boldsymbol{\Sigma}_y^{-1} (\boldsymbol{y}^{\text{obs}} - \boldsymbol{y}) + \frac{1}{2}(\boldsymbol{p} - \boldsymbol{p}^0)^T \boldsymbol{\Sigma_p}^{-1} (\boldsymbol{p} - \boldsymbol{p}^0). \tag{33}$$

This formulation for the MAP estimate leads to a scalar valued function J to minimize. For the subsequent explanations, we split J into three terms:

$$J(\boldsymbol{p}) = J_{\boldsymbol{u}}(\boldsymbol{p}) + J_F(\boldsymbol{p}) + J_{\boldsymbol{p}}(\boldsymbol{p}), \tag{34}$$

$$J_{\boldsymbol{u}}(\boldsymbol{p}) = \frac{1}{2}(\boldsymbol{u}^{\text{obs}} - \boldsymbol{u})^T \boldsymbol{\Sigma_u}^{-1} (\boldsymbol{u}^{\text{obs}} - \boldsymbol{u}), \tag{35}$$

$$J_F(\boldsymbol{p}) = \frac{(F - F^{\text{obs}})^2}{2\sigma_F^2}, \tag{36}$$

$$J_{\boldsymbol{p}}(\boldsymbol{p}) = \frac{1}{2}(\boldsymbol{p} - \boldsymbol{p}^0)^T \boldsymbol{\Sigma_p}^{-1} (\boldsymbol{p} - \boldsymbol{p}^0). \tag{37}$$

Using a gradient based optimization algorithm to minimize J is not straightforward for two reasons:

1. The prior precision matrix $\boldsymbol{\Sigma}_{\boldsymbol{p}}^{-1}$ is dense and thus intractable on larger meshes,
2. Computing ∇J involves computing $\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}\boldsymbol{p}}$ and a finite difference approximation (which is the common approach in the FEMU literature) is intractable on larger meshes.

We address both points in sections 6.1 and 8.1 respectively.

5.3 Parametric Empirical Bayes Formulation

In this section, we no longer assume the hyperparameters $\boldsymbol{\mu} = (\mu_k)_{0 \leq k < N_\mu}$ to be known and propose a Bayesian approach for hyperparameter estimation. More precisely, we model $\boldsymbol{\mu}$ as a random vector, and formulate a log-normal hyperprior on each component of $\boldsymbol{\mu}$:

$$\forall\ 0 \leq i < N_\mu,\ \log \mu_i \sim \mathcal{N}\left(\tau_i^0, \sigma_{\tau_i}^2\right). \tag{38}$$

The choice of the log-normal hyperprior is due to the fact that covariance hyperparameters are often positive. In the rest of this section, we pose $\boldsymbol{\tau} = (\log \mu_k)_{i \leq k < N_\mu}$. The posterior distribution (29) is then updated as follows:

$$D(\boldsymbol{p}, \boldsymbol{\tau} | \boldsymbol{y}^{\mathrm{obs}}) \propto D(\boldsymbol{y}^{\mathrm{obs}} | \boldsymbol{p}) D(\boldsymbol{p} | \boldsymbol{\tau}) D(\boldsymbol{\tau}). \tag{39}$$

The expression for the prior $D(\boldsymbol{p}|\boldsymbol{\tau})$ in (37) must be updated to take into account the normalization term that depends on $\boldsymbol{\mu}$:

$$J_{\boldsymbol{p}}(\boldsymbol{p}) = \frac{1}{2}(\boldsymbol{p} - \boldsymbol{p}^0)^T \boldsymbol{\Sigma}_{\boldsymbol{p}}^{-1}(\boldsymbol{p} - \boldsymbol{p}^0) - \frac{1}{2}\log\det \boldsymbol{\Sigma}_{\boldsymbol{p}}^{-1} + \frac{N_V}{2}\log(2\pi). \tag{40}$$

In the resulting hierarchical Bayesian formulation, the posterior distribution can be approximated with MCMC sampling (e.g. Gibbs sampling, see section 11.3 of [Bishop, 2006]). As we are only interested in the MAP estimate of the unknown parameter, like in section 5.2, it is natural to approximate the hierarchical model with a *Parametric Empirical Bayes* approach (PEB), where the hyperparameters are iteratively updated from the parameter MAP estimator, following the Bayes formula at the hypermodel level:

$$D(\boldsymbol{\tau} | \boldsymbol{p}_{\mathrm{MAP}}) \propto D(\boldsymbol{p}_{\mathrm{MAP}} | \boldsymbol{\tau}) D(\boldsymbol{\tau}). \tag{41}$$

The formulated hypermodel 38 suggests updating the hyperparameters with a MAP estimator $\boldsymbol{\tau}_{\mathrm{MAP}}$ of $\boldsymbol{\tau}$:

$$\boldsymbol{\tau}_{\mathrm{MAP}} = \underset{\boldsymbol{\tau}}{\operatorname{argmin}}\ J_{\boldsymbol{p}}(\boldsymbol{p}_{\mathrm{MAP}}, \boldsymbol{\tau}) + J_{\boldsymbol{\tau}}(\boldsymbol{\tau}), \tag{42}$$

where $J_{\boldsymbol{\tau}}(\boldsymbol{\tau})$ is a new term corresponding to the log-normal hyperprior:

$$J_{\boldsymbol{\tau}}(\boldsymbol{\tau}) = \sum_{i=0}^{N_\mu - 1} \frac{(\tau_i - \tau_i^0)^2}{2\sigma_{\tau_i}}. \tag{43}$$

The resulting formulation in quite inconvenient as we now have $2N_\mu$ hyper-hyperparameters to specify, which are typically also unknown. A common way to tackle this issue is to specify an improper hyperprior as opposed to a log-normal distribution like in equation (38):

$$\forall\ 0 \leq i < N_\mu,\ \log \mu_i \sim \mathcal{U}(\mathbb{R}), \tag{44}$$

which leads to $J_{\boldsymbol{\tau}}(\boldsymbol{\tau}) = 0$. From an empirical Bayes point of view, this is equivalent to updating the hyperparameters with a maximum likelihood (ML) estimator as opposed to a MAP estimator:

$$\boldsymbol{\tau}_{\mathrm{ML}} = \underset{\boldsymbol{\tau}}{\operatorname{argmin}}\ J_{\boldsymbol{p}}(\boldsymbol{p}_{\mathrm{MAP}}, \boldsymbol{\tau}). \tag{45}$$

In section 6.2, we show the numerical shortcomings of using the ML estimator for updating the hyperparameters, and in particular we explain why the resulting optimization problem is unstable when the Matérn covariance is used for prior modelling. To address this issue, we propose in section 6.2 an alternative updating scheme, where the hyperparameter are iteratively identified through variogram regression, another variation of the PEB approach.

6 Spatial Covariance Modelling Through Whittle-Matérn Priors

In the previous section, a multivariate normal prior was formulated on $\boldsymbol{p}$, however a covariance model remains to be specified. In this section, we present and discuss the use of the Matérn covariance model.

6.1 Covariance Precision Discretization

Our choice is motivated by the requirement to obtain an efficient discretization on unstructured grids. In this regard, the Matérn covariance is a suitable choice, as it can lead to a sparse approximation of the precision matrix [Lindgren et al., 2011]. More precisely, the Matérn covariance is defined as

$$M(r;\sigma^2,\ell,\nu) = \frac{\sigma^2 2^{1-\nu}}{\Gamma(\nu)}\left(\frac{r}{\ell}\right)^{\nu} K_\nu\left(\frac{r}{\ell}\right), \tag{46}$$

where σ^2 is the marginal variance, ℓ is the length scale and ν is the smoothness parameter. With regard to the notations introduced in section 5.2, $\boldsymbol{\mu} = (\sigma^2, \ell, \nu)$ is the hyperparameter vector. We also introduce the *correlation length* $\rho = \ell\sqrt{8\nu}$, which is the distance at which the correlations are near 0.1 [Lindgren et al., 2011]. It was shown in [Lindgren et al., 2011, Roininen et al., 2014] that Gaussian random fields with Matérn covariance are the solutions of the following stochastic PDE:

$$(\mathrm{Id} - \ell^2\Delta^2)^{(\nu+d/2)/2} x = \sigma\sqrt{\frac{\ell^d}{\beta}}\mathcal{W}, \tag{47}$$

where

$$\beta = \frac{\Gamma(\nu)}{2^d\pi^{d/2}\Gamma(\nu+d/2)}, \tag{48}$$

and $d = 2, 3$ is the spatial dimension. We also pose $\alpha = \nu + d/2$. For a broad overview of the different applications related to this link between the Matérn covariance and the SPDE (47), we refer to [Lindgren et al., 2022]. When considering an unstructured mesh discretization, the precision matrix of x can have a sparse approximation. We refer to [Lindgren et al., 2011] for a proof. More precisely, let $\boldsymbol{x} = (x_i)_{0\leq i<N_V}$ be the degrees of freedom of a solution of equation (47) discretized in $V(\Omega^{\text{obs}})$, and let $\boldsymbol{Q}_{\boldsymbol{x}}(\alpha,\ell)$ be the precision matrix of $\boldsymbol{x}$. We introduce the usual FEM mass and stiffness matrices for a scalar valued P^1 discretization:

$$\boldsymbol{M} = \left(\int_{\Omega^{\text{obs}}} \varphi_i\varphi_j \,\mathrm{d}\underline{x}\right)_{0\leq i,j<N_V}, \tag{49}$$

$$\boldsymbol{K} = \left(\int_{\Omega^{\text{obs}}} \nabla\varphi_i\cdot\nabla\varphi_j \,\mathrm{d}\underline{x}\right)_{0\leq i,j<N_V}. \tag{50}$$

We have the following sparse approximations:

$$\boldsymbol{Q}_{\boldsymbol{x}}(1,\ell) = \frac{\beta}{\sigma^2\ell^d}(\boldsymbol{M}+\ell^2\boldsymbol{K}), \tag{51}$$

$$\boldsymbol{Q}_{\boldsymbol{x}}(2,\ell) = \frac{\beta}{\sigma^2\ell^d}(\boldsymbol{M}+\ell^2\boldsymbol{K})\boldsymbol{M}^{-1}(\boldsymbol{M}+\ell^2\boldsymbol{K}), \tag{52}$$

$$\forall n\in\mathbb{N},\ n\geq 3,\ \boldsymbol{Q}_{\boldsymbol{x}}(n,\ell) = (\boldsymbol{M}+\ell^2\boldsymbol{K})\boldsymbol{M}^{-1}\boldsymbol{Q}_{\boldsymbol{x}}(n-2,\ell)\boldsymbol{M}^{-1}(\boldsymbol{M}+\ell^2\boldsymbol{K}). \tag{53}$$

Formulas (51)-(53) correspond to $\nu = 0, 1, 2, \ldots$ for $d = 2$ and $\nu = 0.5, 1.5, 2.5, \ldots$ for $d = 3$. For other values $\alpha > 0$ that are not integers, more recent papers give another approximation of $\boldsymbol{Q}_{\boldsymbol{x}}(\alpha,\ell)$ based on, e.g., a truncated Taylor expansion of the underlying pseudodifferential operator or a quadrature of its Brochner integral, and we refer to, e.g. [Bolin et al., 2018, Roininen et al., 2018] for details. Note that formulas (51)-(52) are slightly different to those in [Lindgren et al., 2011], as the base SPDE (47) was taken from [Roininen et al., 2014] which defines different SPDE coefficients. In equation (47), the marginal variance and length scale of $\boldsymbol{x}$ are controlled by separate hyperparameters, as opposed

to the range parameter in [Lindgren et al., 2011] which affects both the correlation length and the marginal variance. In (52)-(53), the inverse of the mass matrix can be approximated by the inverse of its lumped approximation, defined as

$$\forall 0 \leq i < N_V,\ \widetilde{\boldsymbol{M}}_{ii} = \sum_{j=0}^{N_V-1} M_{ij}, \tag{54}$$

$$\forall i \neq j,\ \widetilde{\boldsymbol{M}}_{ij} = 0. \tag{55}$$

In regard to the previous Bayesian formulation for the material parameter $\boldsymbol{p}$, fixing $\alpha \geq 2$ to an integer value leads to a sparse expression for the prior (37), that can be implemented efficiently. For instance, with $\alpha = 2$, we can plug (52) into the MAP estimate expression (37), resulting in a so-called Whittle-Matérn prior:

$$J_{\boldsymbol{p}}(\boldsymbol{p}) = \frac{\beta}{2\sigma^2\ell^2}(\boldsymbol{p}-\boldsymbol{p}^0)^T(\boldsymbol{M}+\ell^2\boldsymbol{K})\widetilde{\boldsymbol{M}}^{-1}(\boldsymbol{M}+\ell^2\boldsymbol{K})(\boldsymbol{p}-\boldsymbol{p}^0). \tag{56}$$

6.2 Whittle-Matérn Hyperparameters Estimation

As discussed in section 5.3, hyperparameter estimation using the PEB method essentially consists in iteratively estimating $\boldsymbol{\mu} = (\sigma^2, \ell, \nu)$ from parameter estimates $\boldsymbol{p}_{\text{MAP}}$. In this section, we discuss several issues related to the use of the Matérn covariance, and in particular, the use of a sparse approximation with a fixed smoothness parameter ν.

First, given the sparse approximation $\boldsymbol{Q}_{\boldsymbol{x}}$ of $\boldsymbol{\Sigma}_{\boldsymbol{p}}^{-1}$ in formulas (51)-(53), the mapping $\nu \mapsto \boldsymbol{Q}_{\boldsymbol{x}}(\alpha(\nu), \ell)$ is not defined for all $\nu > 0$, and thus not differentiable. A gradient based optimization procedure for computing $\boldsymbol{\mu}_{\text{MAP}}$ is thus not possible. A possible workaround would be to set $\alpha = 2$ and estimate only (σ^2, ℓ). Surprisingly, many authors employ this solution, claiming that the data does not contain much information about ν, which is often poorly identified [Diggle and Ribiero, 2007, Bose et al., 2018]. However, this claim is questionable, and for a detailed discussion on this matter, we refer to [Oliveira and Han, 2022].

More precisely, fixing ν to an arbitrary leads to estimability issues of σ^2 and ℓ: there does not exist an estimator for the pair (σ^2, ℓ) that is weakly consistent under infill asymptotics, i.e. when considering a sequence of increasing number of pointwise observations $((p_k)_{0\leq k<n})_{n\geq 1}$, and a sequence of estimators $(\widehat{\boldsymbol{\mu}}_n)_{n\geq 1}$, $\widehat{\boldsymbol{\mu}}_k = (\widehat{\sigma}_k^2, \widehat{\ell}_k)$, then $\widehat{\boldsymbol{\mu}}_n$ does not converge when $n \to \infty$. This result is proven in [Zhang, 2004] and thoroughly discussed in [Zhang, 2004, Tang et al., 2021]. Another result of [Zhang, 2004] is that the so-called *microergodic* parameter $c = \sigma^2/\ell$ is consistently estimable. In the light of these results, in the hierarchical Bayesian model of section 5.3, when setting ν to an arbitrary value, and considering an improper hyperprior for (σ^2, ℓ), the minimization problem for the maximum likelihood estimator $\boldsymbol{\mu}_{\text{ML}}$ used in the PEB hyperparameters update is unstable, in the sense that $\boldsymbol{\mu}_{\text{ML}}$ does not converge when the computational mesh is refined, and that the result of the optimization process is highly sensitive to initialization values due to ridges in the likelihood profile (see, e.g., figure 1 of [Lalchand et al., 2022]). In order to stabilize the hierarchical model, a proper hyperprior (or a fixed value), e.g. a log-normal hyperprior as presented in equation (38), must be set on either σ^2 or ℓ, which results in a hybrid approach between ML and MAP estimator for PEB hyperparameter updating, where is referred to as *hybrid-MAP* for the rest of this paper.

Another shortcoming of this approach is that, assuming σ^2 is known, ν set to an arbitrary value and ℓ estimated using an improper hyperprior, then there is no guarantee that the resulting correlation length ρ matches the statistics of the underlying random process. To illustrate this problem, let's consider a set of one dimensional Matérn field realizations $\boldsymbol{f}_\alpha$ with fixed hyperparameters $\sigma_{\boldsymbol{f}}^2, \rho_{\boldsymbol{f}}$ and increasing values for α (i.e. increasing ν and decreasing $\ell = \rho_{\boldsymbol{f}}/\sqrt{\nu}$). We wish to identify the correlation length $\rho_{\boldsymbol{f}}$ using a maximum likelihood estimator, computed by using formula (56) with a smoothness parameter α_0 set to an arbitrary value ($\alpha_0 = 2$ in this case). The estimator is then

computed with

$$\rho_\alpha^{\mathrm{ML}} = \sqrt{1.5}\ \underset{\ell}{\operatorname{argmin}}\ [-\log L_\ell(\boldsymbol{f})] = \frac{\beta}{2\sigma^2\ell}(\boldsymbol{f}_\alpha - \boldsymbol{f}_\alpha^0)^T(\boldsymbol{M}+\ell^2\boldsymbol{K})\widetilde{\boldsymbol{M}}^{-1}(\boldsymbol{M}+\ell^2\boldsymbol{K})(\boldsymbol{f}_\alpha - \boldsymbol{f}_\alpha^0) \quad (57)$$
$$- \frac{1}{2}\log\det\left[\frac{\beta}{\sigma^2\ell}(\boldsymbol{M}+\ell^2\boldsymbol{K})\widetilde{\boldsymbol{M}}^{-1}(\boldsymbol{M}+\ell^2\boldsymbol{K})\right].$$

We represent in figure 2 realisations $\boldsymbol{f}_2$, $\boldsymbol{f}_4$, $\boldsymbol{f}_{16}$ and the profile of the objective function $-\log L_\ell(\boldsymbol{f}_\alpha)$

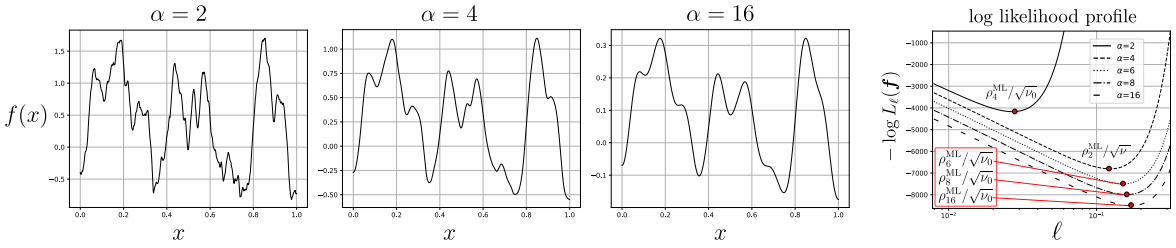


Figure 2: Realisations of a one dimensional Matérn field f with several smoothness parameters α, and fixed correlation length $\rho_{\boldsymbol{f}} = \ell\sqrt{\nu}$. On the right, the negative log-likelihood profile of each realisation w.r.t. ℓ for a fixed smoothness parameter $\alpha_0 = 2$ for discretization of the Matérn precision.

with the values of the estimated correlation length. The right correlation length is recovered only for $\alpha = 2$, which corresponds to the case where the precision is discretized with the correct smoothness parameter. For all other cases, the correlation length is overestimated by an order of magnitude. This simplified example shows that the hybrid-MAP approach highly overestimates the correlation length when the posterior mean $\boldsymbol{p}_{\mathrm{MAP}}$ is smoother than the arbitrarily chosen value set for ν.
In regard to this last issue, we also explore an alternative hyperparameter updating scheme that uses variogram regression. This idea closely follows the developments of [Brown et al., 2020]. More precisely, if $\boldsymbol{z} = (z(\underline{x}_k))_{0\le k<N_{\mathrm{obs}}}$ are pointwise observations of an underlying Gaussian process z with Matérn covariance $M(\cdot;\sigma^2,\ell,\nu)$, then the *empirical variogram* is defined as

$$\widehat{\gamma}(r\pm\delta;\boldsymbol{z}) = \frac{1}{2|\mathcal{N}(r\pm\delta)|}\sum_{(i,j)\in\mathcal{N}(r\pm\delta)}|z_i - z_j|^2, \quad (58)$$
$$\mathcal{N}(r\pm\delta) = \left\{0\le i,j<N_{\mathrm{obs}} : |\underline{x}_i - \underline{x}_j| = r\pm\delta\right\}, \quad (59)$$

where δ is the separation distance which is used for binning. The Matérn variogram [Brown et al., 2020] is defined as

$$\gamma_M(r;a_0,\sigma^2,\ell,\nu) = \begin{cases} 0 & \text{if } r = 0 \\ a_0 + (\sigma^2 - a_0)\left(1 - \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{r}{\ell}\right)^\nu K_\nu\left(\frac{r}{\ell}\right)\right) & \text{if } r > 0 \end{cases}, \quad (60)$$

where the *nugget* $a_0 \ge 0$ is the variance value at a distance just greater than 0, and the *sill* σ^2 is also the marginal variance can be fitted to the empirical variogram $\widehat{\gamma}$ through various methods, e.g. weighted least-squares [Cressie, 1985]:

$$\boldsymbol{\psi}_{\mathrm{opt}} = \underset{\boldsymbol{\psi}}{\operatorname{argmin}}\ J_{\mathrm{vario}}(\boldsymbol{\psi},\boldsymbol{z}) := \sum_r \frac{|\mathcal{N}(r\pm\delta)|}{\gamma_M(r;\boldsymbol{\psi})^2}\left(\widehat{\gamma}(r\pm\delta;\boldsymbol{z}) - \gamma_M(r;\boldsymbol{\psi})\right)^2, \quad (61)$$

where $\boldsymbol{\psi} = (a_0,\sigma^2,\ell,\nu)$. Although the nugget a_0 is estimated by variogram fitting, it is not useful for the considered problem, as it does not appear in the prior expression. A typical fitting method will yield non-integer values for ν. In order to simplify numerical implementation, it is useful to approximate it with the nearest integer values (in 2D) or the nearest half-integer (in 3D). In order to preserve the correlation length, it is then necessary to update the identified length scale accordingly. More precisely, if $\ell_{\mathrm{opt}}, \nu_{\mathrm{opt}}$ are identified hyperparameters, and $\lfloor\nu\rceil_d$, $d = 2,3$ denotes the nearest integer or half-integer, then choosing

$$\widetilde{\ell}_{\mathrm{opt}} = \ell_{\mathrm{opt}}\sqrt{\frac{\nu_{\mathrm{opt}}}{\lfloor\nu\rceil_d}} \quad (62)$$

will preserve the correlation length. Using the pair $(\widetilde{\ell}_{\text{opt}}, \lfloor\nu\rceil_d)$, it is then possible to use formula (53) for an efficient prior precision discretization. With the variogram fitting approach, the posterior for the hierarchical problem is then approximated with an alternate optimization approach described in algorithm 1.

Algorithm 1: Hyperparameter Identification Procedure.

Function `IdentifyHyperparameters`(σ_0^2, ℓ_0, $\boldsymbol{y}^{obs}$, $\boldsymbol{p}_0$, ε_{tol}, η_{tol})**:**
- $\boldsymbol{\mu}_{\text{opt}} \leftarrow (\sigma_0^2, \ell_0)$; $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu}_{\text{opt}}$ — `// Initial hyperparameters guess`
- $\boldsymbol{p}_{\text{opt}} \leftarrow \boldsymbol{p}_0$
- $r \leftarrow 2 * \varepsilon_{\text{tol}}$
- **while** $r \geq \varepsilon_{tol}$ **do**
  - $\boldsymbol{p}_0 \leftarrow \boldsymbol{p}_{\text{opt}}$
  - $\boldsymbol{p}_{\text{opt}} \leftarrow$ `IdentifyParameter`$(\boldsymbol{y}^{\text{obs}}, \boldsymbol{p}_0, \eta_{\text{tol}})$ — `// Algorithm 2`
  - $(a_0, \sigma^2, \ell, \nu) \leftarrow \underset{\boldsymbol{\psi}}{\text{argmin}}\ J_{\text{vario}}(\boldsymbol{\psi}, \boldsymbol{p}_k)$ — `// Variogram regression`
  - $\widetilde{\ell} \leftarrow \ell\sqrt{\nu/\lfloor\nu\rceil_d}$
  - $\boldsymbol{\mu}_{\text{opt}} \leftarrow (\sigma^2, \widetilde{\ell}, \lfloor\nu\rceil_d)$ — `// Rounding of smoothness parameter`
  - $r \leftarrow |\boldsymbol{\mu}_{\text{opt}} - \boldsymbol{\mu}|$
  - $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu}_{\text{opt}}$
- **return** $\boldsymbol{\mu}_{\text{opt}}$

For hyperparameter initialization, the authors of [Brown et al., 2020] recommend fitting the Matérn variogram on the observed field. However, in the present case, due to the non-linear relation between the stiffness parameter $\boldsymbol{p}$ and the displacement $\boldsymbol{u}$, the observed displacement field has a very different spatial correlation structure than the material parameter. We therefore recommend initializing $\boldsymbol{\mu}_0$ to an arbitrary value. Note that with the variogram regression approach, the issues related to the identifiability of (σ^2, ℓ) remain present, as the marginal variance estimated by fitting the empirical variogram is simply the empirical variance of the observed sample and does not converge to the true marginal variance of the underlying Gaussian process when refining the computational mesh. The only difference when compared to the hybrid-MAP approach with the empirical variance as the hyperprior mean is that the smoothness parameter ν is estimated from the data as opposed to set to an arbitrary value, and the discrete prior precision matrix is approximated for the nearest integer (or half integer) value of ν.

7 Comparison with a Classical Regularization Approach

Instead of a Bayesian formulation as developed in section 5.2, a classical regularization approach could have been used instead. In this case, a similar objective function could be defined:

$$J^{\text{c}}(\boldsymbol{p}; \gamma_R) = (\boldsymbol{u}^{\text{obs}} - \boldsymbol{u})^T(\boldsymbol{u}^{\text{obs}} - \boldsymbol{u}) + \gamma_F(F^{\text{obs}} - F)^2 + \gamma_R\|\boldsymbol{p}\|_L^2, \tag{63}$$

where $\|\cdot\|_L$ is a regularization norm, γ_F a weight between the two data terms, and γ_R a regularization parameter. The weight γ_F is usually fixed, and a common choice is to set $\gamma_F = \sigma_u^2/\sigma_F^2$, where σ_u^2 is the average pointwise measurement displacement noise variance, so that each data term is weighted with the corresponding measurement noise. A common choice for $\|\cdot\|$ is the so-called Tikhonov regularization which is simply the L^2 norm, that writes in a discrete setting:

$$\|\boldsymbol{p}\|_{L^2}^2 = \boldsymbol{p}^T\boldsymbol{M}\boldsymbol{p}. \tag{64}$$

However, in order to enforce a smooth solution, a more relevant choice is the $H^1(\Omega)$ semi-norm:

$$\|p\|_{H^1(\Omega)}^2 = \int_\Omega (\nabla p)^2 \, \mathrm{d}x, \tag{65}$$

which takes the following form in a discrete setting:

$$\|\boldsymbol{p}\|_{H^1(\Omega)}^2 = \boldsymbol{p}^T\boldsymbol{K}\boldsymbol{p}. \tag{66}$$

Compared to the Bayesian formulation, there is only a single regularization parameter, which must be tuned. A common approach is to select γ_R with an approach based on the L-curve, a parametric curve defined as:

$$\gamma_R \mapsto (J^c_{\text{data}}(\gamma_R), J^c_{\text{reg}}(\gamma_R)), \tag{67}$$

where

$$J^c_{\text{data}}(\gamma_R) = (\boldsymbol{u}^{\text{obs}} - \boldsymbol{u}(\boldsymbol{p}_{\gamma_R}))^T(\boldsymbol{u}^{\text{obs}} - \boldsymbol{u}(\boldsymbol{p}_{\gamma_R})) + \gamma_F(F^{\text{obs}} - F(\boldsymbol{p}_{\gamma_R}))^2, \tag{68}$$

$$J^c_{\text{reg}}(\gamma_R) = \|\boldsymbol{p}_{\gamma_R}\|_L, \tag{69}$$

$$\boldsymbol{p}_{\gamma_R} = \underset{\boldsymbol{p}}{\text{argmin}}\ J^c(\boldsymbol{p}; \gamma_R). \tag{70}$$

The L-curve represents the tradeoff between the regularized solution norm and the residual norm. For more background and details, we refer to [Hansen, 1992] (and [Gulliksson, 1998] for application to nonlinear problems). A common method to select the regularization parameter is to choose the value that corresponds to the L-curve point of maximum curvature, which yields the value satisfying the Morozov discrepancy principle [Engl and Grever, 1994]. In practice, the point of maximum curvature can be obtained graphically, or by the means of more sophisticated algorithms that aim at reducing the number of L-curve points to construct (see, e.g., [Cultrera and Callegaro, 2020]).

We note that even though an analogy between classical and Bayesian inverse problems can be formulated [Polson and Sokolov, 2019], the Whittle-Matérn prior formulated in section 6 is not equivalent to a standard Tikhonov regularization. In particular, the single regularization parameter γ_R does not allow controlling the amplitude and length scale of the solution separately, as it is the case with the Matérn covariance parameters (σ^2, ρ).

In section 9, we show an example of identified parameter and regularization parameter through the classical approach. Although it leads to satisfactory results for parameter estimation, we lose the statistical interpretation of the regularization parameter and the sampling capabilities from the tuned covariance matrix. This is why the presented framework based on Gaussian random fields modelling and the Bayesian inverse problem formulation is considered more relevant for modelling spatially varying material parameters.

8 Numerical Optimization : Adjoint Formulation and Numerical Scheme

In this section, we discuss the numerical approximation of the MAP estimator of the unknown parameter defined in section 5. We first present how to efficiently compute the loss function gradient through the adjoint method, and then discuss the optimization algorithm used to derive MAP estimates.

8.1 Efficient Gradient Computation with the Adjoint Method

In typical parameter identification procedures such as FEMU or I-DIC, the parameter sensitivities are approximated with a finite difference scheme. This is tractable in the case where there are few parameters to identify. In the present case, as we introduced a finite elements discretization of the unknown parameter p, there is a total number of N_V degrees of freedom to identify, which scales with mesh size. In this case a finite differences approximation of ∇J is usually intractable for larger meshes. The adjoint method introduces a so-called *adjoint problem* which can be solved using the finite elements method and provides an efficient way to compute the gradient. The adjoint method has been used in a wide variety of settings to compute gradients in PDE-constrained optimization problems [Richter, 2021, Giannakoglou and Papadimitriou, 2008, Asch et al., 2016], however the method is often presented in a problem specific way and is not always straightforward to adapt to another problem. For clarity, we derive the equations in the specific case of problem (8)-(11).
Reusing the notations introduced in section 5.1, given a displacement field $\underline{u} \in H^1_D(\Omega^{\text{obs}})$ and a stiffness parameter field $p \in L^\infty(\Omega^{\text{obs}})$, we define the *weak residual* as the linear form

$$\underline{v} \mapsto R(\underline{u}, \underline{v}, p) = l(\underline{v}) - a(\underline{u}, \underline{v}, p), \tag{71}$$

where we made the dependency with respect to p appear explicitly in the notation. We consider an arbitrary scalar valued loss function $J(\underline{u}, p) : H^1_D(\Omega^{\text{obs}}) \times L^\infty(\Omega^{\text{obs}}) \to \mathbb{R}$ that is assumed to be Gateaux differentiable. With these notations, the parameter identification problem can be written as follows:

$$(\underline{u}^{\text{opt}}, p^{\text{opt}}) = \underset{(\underline{u},p)\in H^1(\Omega^{\text{obs}})\times L^\infty(\Omega^{\text{obs}})}{\operatorname{argmin}} J(\underline{u}, p) \tag{72}$$

$$\text{such that } \forall \underline{v} \in H^1_0(\Omega^{\text{obs}}),\ R(\underline{u}^{\text{opt}}, \underline{v}, p^{\text{opt}}) = 0. \tag{73}$$

The PDE constraint (73) introduces a dependency between $\underline{u}$ and p and J can thus be seen as a composed function $p \mapsto J(\underline{u}(p), p)$. We now consider $p, \delta p \in L^\infty(\Omega^{\text{obs}})$ and $\underline{u} \in H^1_D(\Omega^{\text{obs}})$, and expand the Gateaux derivative of J and R at point p in the direction δp, using the chain rule:

$$\mathrm{d}J(\underline{u}(p), p; \delta p) = \partial_1 J(\underline{u}(p), p; \mathrm{d}\underline{u}(p; \delta p)) + \partial_2 J(\underline{u}(p), p; \delta p), \tag{74}$$

$$\forall \underline{v} \in H^1_0(\Omega^{\text{obs}}),\ \mathrm{d}R(\underline{u}(p), \underline{v}, p; \delta p) = \partial_1 R(\underline{u}(p), \underline{v}, p; \mathrm{d}\underline{u}(p; \delta p)) + \partial_3 R(\underline{u}(p), \underline{v}, p; \delta p). \tag{75}$$

Using the constraint (73), we can write $\mathrm{d}R(\underline{u}(p), \underline{v}, p, \delta p) = 0$ for all $\underline{v} \in H^1_0(\Omega^{\text{obs}})$, and then

$$\partial_3 R(\underline{u}(p), \underline{v}, p; \delta p) = -\partial_1 R(\underline{u}(p), \underline{v}, p; \mathrm{d}\underline{u}(p; \delta p)). \tag{76}$$

Now, let's consider $\underline{\lambda}$, the unique solution to the variational *adjoint* problem:
Find $\underline{\lambda} \in H^1_0(\Omega^{obs})$ *such that*

$$\forall \delta\underline{u} \in H^1(\Omega^{\text{obs}}),\ \partial_1 R(\underline{u}(p), \underline{\lambda}, p; \delta u) = \partial_1 J(\underline{u}(p), p; \delta u). \tag{77}$$

Combining (77) with $\delta\underline{u} = \mathrm{d}\underline{u}(p; \delta p)$ and (76) with $\underline{v} = \underline{\lambda}$ inside (74) yields

$$\mathrm{d}J(\underline{u}(p), p; \delta p) = \partial_2 J(\underline{u}(p), p; \delta p) - \partial_3 R(\underline{u}, \underline{\lambda}, p; \delta p). \tag{78}$$

Typically, both derivatives in (78) can be computed analytically, provided the adjoint solution $\underline{\lambda}$ is available. We also note that in the present case, as the operator $(\underline{u}, \underline{v}) \mapsto a(\underline{u}, \underline{v}, p)$ is bilinear and self-adjoint, the adjoint equation (77) reduces to

$$a(\delta\underline{u}, \underline{\lambda}, p) = \partial_1 J(\underline{u}(p), p; \delta u). \tag{79}$$

Now, let's apply the parameter sensitivity formula (78) to the discretized Bayesian formulation introduced in section 5.2. We recall that the Bayesian loss is composed of three terms $J(\boldsymbol{p}) = J_{\boldsymbol{u}}(\boldsymbol{p}) + J_F(\boldsymbol{p}) + J_{\boldsymbol{p}}(\boldsymbol{p})$. For the third term, computing the gradient is straightforward:

$$\nabla J_{\boldsymbol{p}}(\boldsymbol{p}) = (\boldsymbol{p} - \boldsymbol{p}^0)^T \boldsymbol{\Sigma}_{\boldsymbol{p}}^{-1}. \tag{80}$$

For $\nabla J_{\boldsymbol{u}}$ and ∇J_F, we apply the adjoint method. First, we note $\boldsymbol{A}$ the matrix of operator a in basis $(\underline{\varphi}_k)_{0\le k<N_{\underline{V}}}$:

$$\boldsymbol{A} = \Big(a(\underline{\varphi}_i, \underline{\varphi}_j)\Big)_{0\le i,j<N_{\underline{V}}}. \tag{81}$$

To keep notations simple, we assume that basis functions are ordered in such a way that $\boldsymbol{A}$ can be written in the following blocked form

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{DD} & \boldsymbol{A}_{DU} \\ \boldsymbol{A}_{DU} & \boldsymbol{A}_{UU} \end{bmatrix}, \tag{82}$$

where D is the set of DOF indices located inside $\partial\Omega^{\text{obs}}_D$ and U is the set of DOF indices located inside $\Omega^{\text{obs}} \setminus \partial\Omega^{\text{obs}}_D$. The linear system satisfied by $\boldsymbol{u}$ given boundary conditions (11) writes

$$\boldsymbol{A}_0 \boldsymbol{u} = \boldsymbol{b}, \tag{83}$$

where

$$\boldsymbol{A}_0 = \begin{bmatrix} \boldsymbol{I}_{DD} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{UU} \end{bmatrix},\ \boldsymbol{b} = \begin{bmatrix} \boldsymbol{u}^{\text{obs}}_D \\ -\boldsymbol{A}_{UD}\boldsymbol{u}^{\text{obs}}_D \end{bmatrix}, \tag{84}$$

and $\boldsymbol{u}_D^{\text{obs}}$ is a vector collecting degrees of freedom of $\underline{u}^{\text{obs}}$ over $\partial\Omega_D^{\text{pbs}}$. Constructing $\boldsymbol{A}_0$ and $\boldsymbol{b}$ is a common way to apply non-homogeneous Dirichlet boundary conditions to a linear system. As specified in problem (77), the adjoint solution $\underline{\lambda}$ is sought in $H_0^1(\Omega^{\text{obs}})$ and therefore satisfies homogeneous Dirichlet boundary conditions. The linear system for the discretized adjoint solution $\boldsymbol{\lambda}$ then writes

$$\boldsymbol{A}_0\boldsymbol{\lambda} = -\boldsymbol{b}_{\text{adj}}, \tag{85}$$

where

$$\boldsymbol{b}_{\text{adj}} = \begin{bmatrix} \boldsymbol{0}_D \\ \left(\dfrac{\partial J}{\partial \boldsymbol{u}}\right)_U \end{bmatrix}, \quad \frac{\partial J}{\partial \boldsymbol{u}} = -(\boldsymbol{u}^{\text{obs}} - \boldsymbol{u})^T \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} + \frac{\partial J_F}{\partial \boldsymbol{u}}. \tag{86}$$

The expression for the gradient is then derived from (78), provided $\boldsymbol{u}$ and $\boldsymbol{\lambda}$ have already been computed:

$$\nabla J(\boldsymbol{p}) = \frac{\partial J}{\partial \boldsymbol{p}} - \frac{\partial}{\partial \boldsymbol{p}}\left(\boldsymbol{\lambda}^T \boldsymbol{R}\boldsymbol{u}\right) \tag{87}$$

$$= (\boldsymbol{p} - \boldsymbol{p}^0)^T \boldsymbol{\Sigma}_{\boldsymbol{p}}^{-1} - \frac{\partial}{\partial \boldsymbol{p}}\left(\boldsymbol{\lambda}^T \boldsymbol{R}\boldsymbol{u}\right), \tag{88}$$

where

$$\boldsymbol{R} = \left(-a(\underline{\varphi}_i, \underline{\varphi}_j, p)\right)_{0 \leq i,j < N_{\underline{v}}}. \tag{89}$$

Note that the second term of (88) is not straightforward to compute and depends on the stiffness model $p \mapsto \overline{C}(p)$. We show in section 8.2 how to leverage automatic differentiation to compute this term easily.

Remark. In the case where multiple stiffness parameter fields $\boldsymbol{p}_1, ..., \boldsymbol{p}_{N_{\boldsymbol{p}}}$ are sought, as derivatives with respect to $\boldsymbol{p}_i$ are not required to construct the adjoint system, only two linear systems must be solved to compute $\nabla J(\boldsymbol{p}_1, ..., \boldsymbol{p}_{N_{\boldsymbol{p}}})$, regardless of the number of stiffness parameters.

8.2 Implementation

The FEMU loss function is minimized with a quasi-Newton method. As opposed to most FEMU implementations, e.g. [Leclerc et al., 2009, Neggers et al., 2017], the L-BFGS method [Liu and Nocedal, 1989] is used instead of Gauss-Newton, as the latter requires computing the Jacobian matrix of the residual vector to approximate the hessian, and in our case storing the full Jacobian has a memory complexity of $\mathcal{O}(N_V^2)$, which can make it intractable for larger meshes. Instead, L-BFGS requires only the loss gradient, and stores only a few columns of the hessian approximation, which keeps the overall memory complexity to $\mathcal{O}(N_V)$. The iterative optimization procedure for parameter identification is summarized in algorithm 2, and the global hyperparameter identification procedure in algorithm 1.

Our implementation of the proposed method is written in Python using free and open source software only. For finite elements operators assembly, we use the DOLFINx library [Habera et al., 2020]. Using the FEniCSx ecosystem, both variational problems (18) and (77), as well as the stiffness model $p \mapsto \overline{C}(p)$ and its gradient can be modelled using the UFL language [Alnæs et al., 2014]. We provide in figure 3 a simplified version of the code to compute the PDE residual gradient and the gradient of the macroscopic force misfit ∇J_F, which are both dependent on the stiffness model and thus not straightforward to differentiate analytically. Instead, by leveraging the automatic differentiation capabilities of UFL and FFCX, the required derivatives do not need to be expressed analytically. As the optimization loop involves solving two linear systems, we leverage the PETSc library [Balay et al., 2022] for all solves. A stabilized conjugate gradient solver is used alongside an algebraic multigrid preconditioner. For hyperparameter estimation, the GSTools library is used [Müller et al., 2022, Müller and Schüler, 2023], which provides routines for empirical variogram estimation and variogram fitting.

Algorithm 2: Parameter Identification Procedure

Function `IdentifyParameter`($\boldsymbol{y}^{obs}$, $\boldsymbol{p}_0$, η_{tol})**:**
- $\boldsymbol{p} \leftarrow \boldsymbol{p}_0$ `// Initial stiffness parameter guess`
- **while** $J(\boldsymbol{p}, \boldsymbol{y}^{obs}) > \eta_{tol}$ **do**
  - Assemble $\boldsymbol{A}_0,\ \boldsymbol{b}$;
  - Solve $\boldsymbol{A}_0\boldsymbol{u} = \boldsymbol{b}$ `// Forwad problem`
  - Assemble $\frac{\mathrm{d}J}{\mathrm{d}\boldsymbol{u}}$ and $\boldsymbol{b}_{\mathrm{adj}}$;
  - Solve $\boldsymbol{A}_0\boldsymbol{\lambda} = \boldsymbol{b}_{\mathrm{adj}}$ `// Adjoint problem`
  - Assemble $\frac{\partial J}{\partial \boldsymbol{p}}$ and $\frac{\partial}{\partial \boldsymbol{p}}\left(\boldsymbol{\lambda}^T \boldsymbol{R}\boldsymbol{u}\right)$;
  - $\frac{\mathrm{d}J}{\mathrm{d}\boldsymbol{p}} \leftarrow \frac{\partial J}{\partial \boldsymbol{p}} - \frac{\partial}{\partial \boldsymbol{p}}\left(\boldsymbol{\lambda}^T \boldsymbol{R}\boldsymbol{u}\right)$;
  - $\delta\boldsymbol{p} \leftarrow \mathrm{BGFS}\left(\frac{\mathrm{d}J}{\mathrm{d}\boldsymbol{p}}\right)$ `// L-BFGS update formula`
  - $\boldsymbol{p} \leftarrow \boldsymbol{p} + \delta\boldsymbol{p}$;
- **return** p

```python
from typing import Callable
from ufl import *
from dolfinx.fem import Function, form, assemble_vector, assemble_scalar

def epsilon(u):
    """strain"""
    return sym(grad(u))

def sigma(u, C):
    """stress"""
    return inner(C, epsilon(u))

def dRdp(u_h: Function, lmbda_h: Function, p: Function, stiffness_model: Callable):
    """
    Evaluate PDE residual sensitivity
        u_h: forward problem solution
        lmbda_h: adjoint problem solution
        p: current model parameter values
        stiffness_model: parameter to stiffness mapping in UFL
    """
    R = -dot(sigma(epsilon(u_h), stiffness_model(p)), epsilon(lmbda_h)) * dx
    dRdp = derivative(R, p)

    # call FFCX JIT compiler and generate optimized C code
    dRdp_form = form(dRdp)

    # assemble DOF array
    return assemble_vector(dRdp_form)

def dJFdp(u_h: Function, p: Function, F_obs: float, sigma_F: float, n_F, stiffness_model: Callable):
    """Evaluate macroscopic force missfit"""
    F = dot(n_F, dot(sigma(epsilon(u_h), stiffness_model(p)), n_F)) * dx
    dFdp = derivative(F, p)

    F_eval = assemble_scalar(form(F))
    dFdp_eval = assemble_vector(form(dFdp))

    return (F_eval - F_obs) / sigma_F**2
```

Figure 3: Simplified implementation of residual and loss sensitivities using UFL and DOLFINx assemblers

9 Results and Discussion

To showcase our method, we consider two numerical experiments, one modelling standard DIC with surface observations, and one with volume observations.

9.1 Identification of a Young Modulus from Surface Displacement

The first model problem is posed on the thin dog-bone specimen represented in figure 1. We consider an isotropic stiffness model, with a spatially varying Young modulus E, and a constant Poisson ratio ν_r (we employ the r subscript to avoid confusion with ν that denotes the Matérn smoothness hyperparameter). The constitutive law 9 writes:

$$\underline{\underline{\sigma}} = \frac{E\nu_r}{(1+\nu_r)(1-2\nu_r)}\,\mathrm{tr}(\underline{\underline{\varepsilon}})\underline{\underline{I}}_3 + \frac{E}{1+\nu_r}\underline{\underline{\varepsilon}}. \tag{90}$$

The considered Young modulus is a realisation of a log-Gaussian random field with a Matérn covariance

$$\boldsymbol{p} := \ln \boldsymbol{E} \sim \mathcal{N}(\ln E^0, \boldsymbol{\Sigma}(\sigma^2, \ell, \nu)), \tag{91}$$

with fixed smoothness parameter $\nu = 1$. A tensile test is simulated on Ω, where the traction direction is represented by the unit vector $\underline{n}_F$. A uniform stress boundary condition $\underline{\underline{\sigma}} \cdot \underline{n} = s_0 \underline{n}_F$ is applied on $\partial\Omega_D^{\mathrm{obs}}$ (in this case, $\underline{n} = \underline{n}_F$), and the surface displacement $\underline{u}$ is simulated with the finite element method. The observed DIC displacement is simulated by adding some noise $\boldsymbol{\xi}$ to the simulated displacement degrees of freedom $\boldsymbol{u}^{\mathrm{obs}} = \boldsymbol{u} + \boldsymbol{\xi}$. For simplicity, we consider an uncorrelated Gaussian white noise

$$\boldsymbol{\xi} \sim \mathcal{N}(0, \sigma_{\boldsymbol{u}}^2 \boldsymbol{I}), \tag{92}$$

and the measured force is calculated from the simulated stress field with formula (13) and corrupt with Gaussian white noise $\eta \sim \mathcal{N}(0, \sigma_F^2)$. The generated data is represented in figure 4. In order to study the impact of measurement noise on the identified parameter and hyperparameters, we define the observed signal-to-noise ratio (SNR) for the displacement and macroscopic force as

$$\mathrm{SNR}_{\boldsymbol{u}}^{\mathrm{dB}}(\boldsymbol{u}, \sigma_u) = 20\log_{10}\left(\frac{1}{N_{\mathrm{obs}}\sigma_u}\sum_{k=1}^{N_{\mathrm{obs}}} m_k\right), \tag{93}$$

$$m_k = \sqrt{(u_{3k}^{\mathrm{obs}})^2 + (u_{3k+1}^{\mathrm{obs}})^2 + (u_{3k+2}^{\mathrm{obs}})^2}, \tag{94}$$

$$\mathrm{SNR}_F^{\mathrm{dB}}(F, \sigma_F) = 20\log_{10}\left(\frac{F^{obs}}{\sigma_F}\right). \tag{95}$$

In the following results, we choose $\sigma_{\boldsymbol{u}}$ and σ_F such that $\mathrm{SNR}_{\boldsymbol{u}}^{\mathrm{dB}} \approx \mathrm{SNR}_F^{\mathrm{dB}}$. For a wide range of noise levels, we identify the MAP estimate for the posterior density, and the Matérn hyperparameters with the various PEB approaches presented in section 6.2: the ML estimator approach, the hybrid-MAP approach and the variogram regression approach. We recall that for the ML estimator and hybrid-MAP approaches, considering the approximations formulated in section 6.2, the discrete Whittle-Matérn prior is constructed by fixing the smoothness parameter to $\nu = 1$ (so that $\alpha = 2$). For the variogram regression approach, formula (53) is applied, using the identified smoothness parameter truncated to the nearest integer, as reported in algorithm 1. The MAP estimates E_{MAP} and associated kinematic fields are showcased in figure 5 for decreasing noise levels. For high SNR values, the identified Young modulus field approximates very well the true field. For decreasing SNR values, $\boldsymbol{p}_{\mathrm{MAP}}$ appears more smoothed out than the true field. Note that even though the true and simulated strain fields are represented in figures 4 and 5, no strain data is used for parameter or hyperparameter estimation. The reason is that the observed strain is computed from the noisy observed displacement field:

$$\underline{\underline{\varepsilon}}^{\mathrm{obs}} = \nabla\left(\underline{u} + \underline{\xi}\right). \tag{96}$$

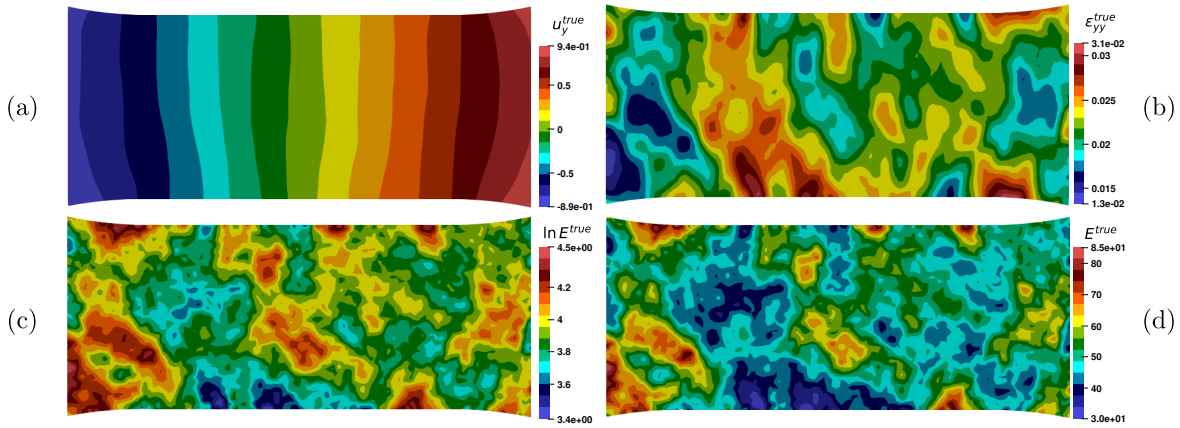


Figure 4: Generated data for model problem. Simulated displacement (a), simulated strain field (b), generated Matérn field $p = \ln E$ (c), corresponding Young modulus field $E = \exp(p)$ (d).

Differentiating noisy data is an ill-posed problem [Knowles and Renka, 2014] that involves some kind of underlying regularization (e.g. Tikhonov regularization). In typical DIC software, a regularization parameter is usually set to an arbitrary value in order to compute and visualize the strain field. Performing covariance estimation from the observed strain fields would yield more information about that regularization parameter than the covariance properties of the underlying material parameter. Even in a low noise setting, as the strain field is fully computed from the displacement field, including strain data in the Bayesian modelling would not be more informative than using the displacement field only.

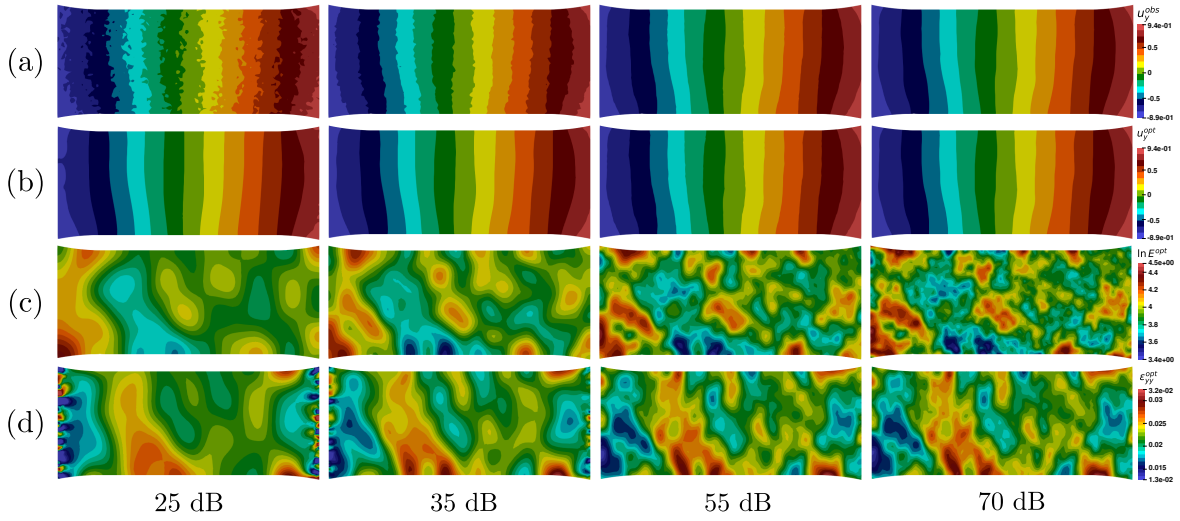


Figure 5: Plots of observed displacement (a), identified displacement (b), identified stiffness parameter $\ln E$ (c), identified strain (d) for decreasing noise levels.

In figure 10, we compare the identified hyperparameters with respect to noise level for the three proposed PEB updating approaches. For the variogram regression and hybrid-MAP approaches, the identified marginal variance and correlation length stabilize close to the true values, whereas for the ML approach, the identified hyperparameters are erroneous even with a high SNR, despite the identified Young modulus field being close to the other PEB methods. These results can be interpreted as an illustration of the results of [Zhang, 2004]: when fixing ν (here, ν is set to 1), it is not possible to differentiate two hyperparameter pairs (σ_0^2, ℓ_0) and (σ_1^2, ℓ_1) if $\sigma_0^2/\ell_0 = \sigma_1^2/\ell_1$ with a ML estimator. In the hybrid-MAP approach, a hyperprior is specified on σ^2 only. Numerically, this is equivalent to estimate the microergodic hyperparameter σ^2/ℓ, which is shown to be consistently estimable in [Zhang, 2004]. Numerically, the instabilities with the ML estimator approach translate to ridges

in the log-likelihood profile $\boldsymbol{\tau} \mapsto J_{\boldsymbol{p}}(\boldsymbol{p}_{\text{MAP}}, \boldsymbol{\tau})$ (see, e.g., figure 1 of [Lalchand et al., 2022]), which means that the identified hyperparameters are also highly dependent on initialization values. In this example, as the true smoothness parameter is set to $\nu = 1$, the discretized prior (56) constructed with $\alpha = 2$ models the correct smoothness. In that case, given the results of figure 10, the hybrid-MAP and variogram regression approaches are equivalent.

As an illustration of the classical approach detailed in section 7, the Young modulus field is also identified through gradient penalty regularization, with a regularization parameter selected from the constructed L-curve. For a given SNR level, identified parameters for various values of γ_R are represented in figure 6, alongside with the associated empirical L-curve. This figure illustrates that the solution is smoother as the value of γ_R increases. Figure 7 illustrates selection of γ_R through the

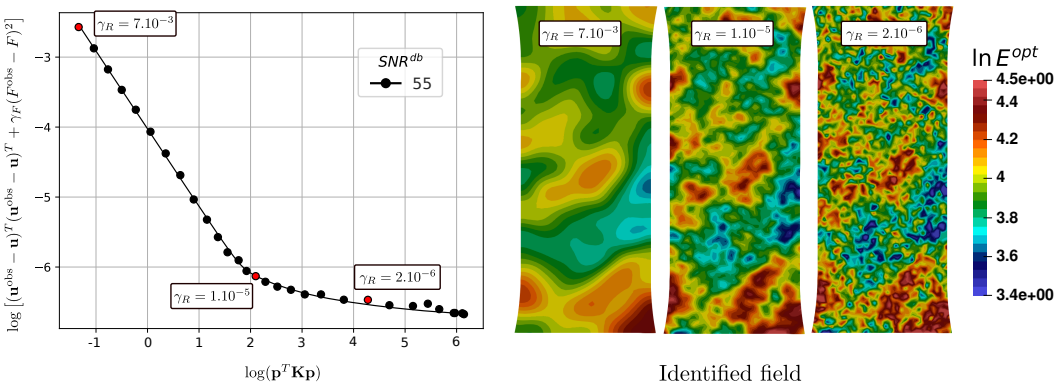


Figure 6: Representation of the L-curve for a fixed SNR level, and identified parameters for different values of γ_R.

L-curve and associated identified parameter, for two different noise levels. As expected, the magnitude of regularization is increased when the noise level is increased. We can also compare the identified parameters of figure 5 and figure 7 to conclude that both the classical and Bayesian approach yield similar results for parameter identification, and that the main difference between the two approaches lies in the statistical interpretation of covariance parameters, and the sampling capabilities of the Bayesian approach.

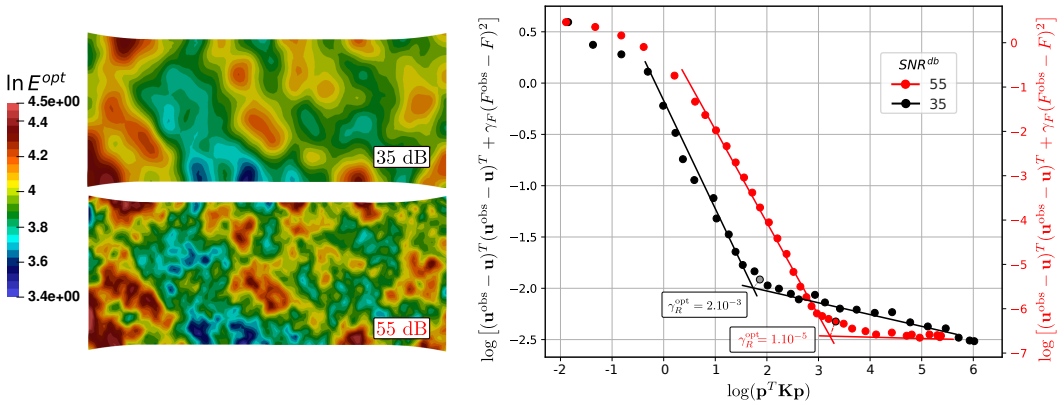


Figure 7: Identified classical regularization parameters from two L-curves associated with two different SNR levels, and Young modulus field with optimal values for γ_R.

From the identified hyperparameters, Matérn random fields with similar covariance properties can then be sampled. We illustrate this in figure 8, where new Matérn fields are sampled from the prior distribution with tuned covariance. The samples were generated using the Python library GSTools [Müller et al., 2022, Müller and Schüler, 2023]. These samples illustrate the impact of measurement noise on the spatial statistics of the resulting generative model.

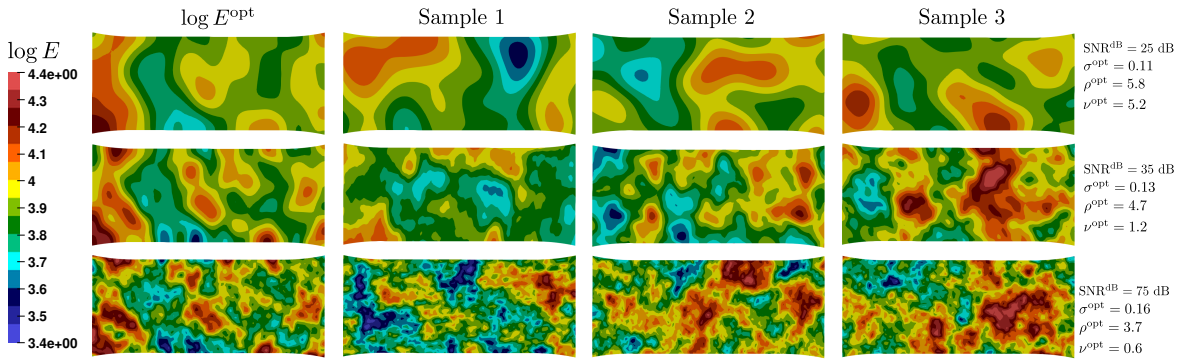


Figure 8: Identified stiffness parameter field $p^{\text{opt}} = \ln E^{\text{opt}}$ and random samples from the prior distribution with tuned Matérn covariance through the variogram regression approach, for increasing SNR levels.

A more interesting result is presented in figure 9, where a tensile test was simulated for the same specimen, using the stiffness tensor distribution resulting from the sampled Matérn fields with tuned hyperparameters using the variogram regression approach. For each simulation, the boundary conditions presented at the beginning of section 9.1 were used. The resulting simulated displacement and strain fields illustrate the expressiveness of the generative model obtained when combining the tuned covariance model with the selected stiffness model: our method was able to learn the spatial statistics of kinematic fields that cannot simply be modelled as Gaussian random fields.

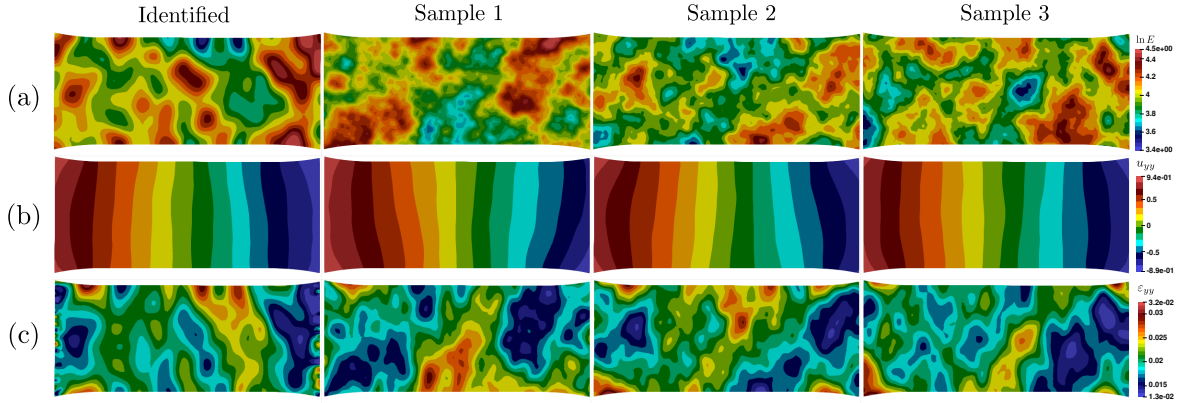


Figure 9: Identified fields and generated fields from the tuned hyperparameters and stiffness model: log Young modulus (a), tensile displacement (b), tensile strain (c). The identified fields were obtained using the variogram regression approach with a SNR of 55 dB.

Note. In figure 10, even though the identified parameter field and hyperparameters appear to stabilize with increasing SNR (for two of the three approaches), there is no proper convergence due to the ill-posedness nature of the inverse elasticity problem. More precisely, for any fixed set of hyperparameters $\boldsymbol{\mu}$, when $\text{SNR}^{\text{dB}} \to \infty$, then $\sigma_F^2 \to 0$ and $\|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\| \to \infty$ for any norm, and consequently the loss function (34) is not well-defined. For very high SNR levels, the prior term of the loss becomes negligible and $J(\boldsymbol{p}) \approx J_{\boldsymbol{u}}(\boldsymbol{p}) + J_F(\boldsymbol{p})$. This is essentially equivalent to a classical noise free inverse problem without regularization, which is numerically unstable due to the non-invertible nature of the forward operator $\mathcal{A}$. A Bayesian way to modify the modelling to ensure convergence with respect to SNR would be to model the ill-posedness of operator $\mathcal{A}$ as artificial operator noise $\zeta_{\mathcal{A}} \sim \mathcal{N}(0, \sigma_{\mathcal{A}}^2)$ and update the likelihood accordingly:

$$D(\boldsymbol{y}^{\text{obs}}|\boldsymbol{p}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{y}^{\text{obs}} - \boldsymbol{y})^T(\boldsymbol{\Sigma}_{\boldsymbol{y}} + \sigma_{\mathcal{A}}^2\boldsymbol{I})^{-1}(\boldsymbol{y}^{\text{obs}} - \boldsymbol{y})\right]. \tag{97}$$

The added term acts as a classical regularization, and is equivalent to artificially increasing the prior variance and thus the marginal variance hyperparameter no longer has a statistical interpretation. The noise free case can also be treated following the classical approach developed in section 7 with γ_F set to an arbitrary value. However, this asymptotic noise free case is not very important for actual DIC parameter identification, as it can only occur in numerical experiments.

9.2 Identification of a Spatial Orientation from Volume Observations

In this second example, we showcase our method on volume displacement observations, modelling a DVC experiment, in a case where the results for the hybrid-MAP and variogram fitting approaches differ. We consider the thick dog-bone specimen represented in figure 11, in which the observation zone is a volume subdomain $\Omega^{\text{obs}} \subset \mathbb{R}^3$. The considered stiffness model represents a transversely isotropic material, where the symmetry plane is orthogonal to the loading direction represented by the unit vector $\underline{n}_F$. The local orientation is defined in spherical coordinates (θ, φ) and represents the direction orthogonal to the symmetry plane in which the material is assumed to be stiffer. The local orientation is assumed to be spatially varying, which can be seen as a simplified model for waviness of fibers in a unidirectional composite. The transversely isotropic stiffness tensor is denoted by $\overline{C}_{\text{ref}}(E_1, E_2, G_{12}, \nu_{12}, \nu_{23})$ where $E_1, E_2, G_{12}, \nu_{12}, \nu_{23}$ are known constants. The stiffness of the wavy material is defined as

$$\overline{C}(\theta, \varphi) = \underline{\underline{\mathcal{R}}}(\theta, \varphi) : \underline{\underline{\mathcal{R}}}(\theta, \varphi) : \overline{C}_{\text{ref}}(E_1, E_2, G_{12}, \nu_{12}, \nu_{23}) : \underline{\underline{\mathcal{R}}}(\theta, \varphi) : \underline{\underline{\mathcal{R}}}(\theta, \varphi), \tag{98}$$

where $\underline{\underline{\mathcal{R}}}(\theta, \varphi)$ is a rotation matrix. For the generated data, we give the following closed form for θ^{true} and φ^{true}:

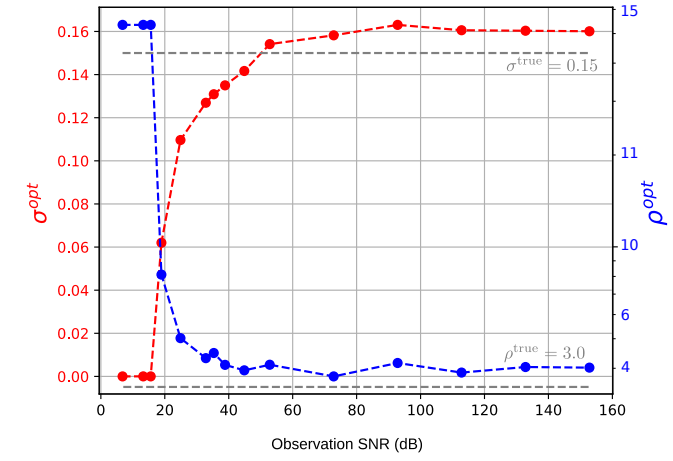
$$\theta^{\text{true}}(x, y, z) = a\left(\cos(x/b)\cos(y/b)\sin(z/b) - 1/2\right), \tag{99}$$

$$\varphi^{\text{true}}(x, y, z) = a\left(\sin(x/b)\sin(y/b)\cos(y/b) - 1/2\right), \tag{100}$$

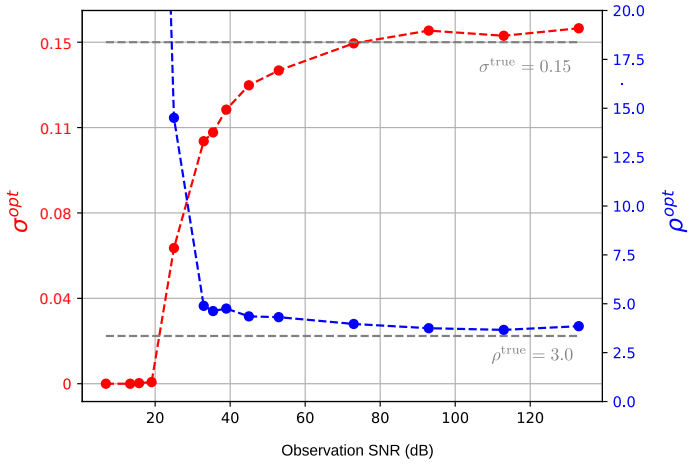
where $a, b > 0$ are fixed values. As with the previous example, the noisy displacement is observed in the volume subdomain Ω^{obs}, and the unknown discrete stiffness parameters $(\boldsymbol{p}_0, \boldsymbol{p}_1) = (\boldsymbol{\theta}, \boldsymbol{\varphi})$ are modelled by a stationary Gaussian random field with Matérn covariance. The unknown hyperparameters are assumed to be identical for both stiffness parameters ($\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$). The generated parameters ϕ_{true} and θ_{true}, and MAP estimates ϕ_{MAP} and θ_{MAP} are represented in figure 12, where the variogram regression approach was used for hyperparameter identification. The results show that both MAP estimates are close to the true parameters, even though the parameters were not initially sampled from the prior distribution. What's interesting is that, given the analytical expressions (99)-(100), ϕ_{true} and θ_{true} are of class $\mathcal{C}^\infty$ and thus cannot be modelled accurately with a rough Matérn covariance model (with $\nu = 1/2$). As discussed in section 6.2, the hybrid-MAP approach will overestimate the correlation length by an order of magnitude. This is represented in figure 13, where samples from the prior distribution are represented. In the first case, the hyperparameters are identified with the variogram regression approach, and in the other case the hybrid-MAP approach is used. A high SNR (60 dB) is used in both cases. The results show that the correlation length of the new samples is highly overestimated when the hybrid-MAP approach is used, whereas the variogram regression approach leads to the right correlation length. In this case, the variogram regression approach is thus better suited for hyperparameter identification than the hybrid-MAP approach.

9.3 Discussion

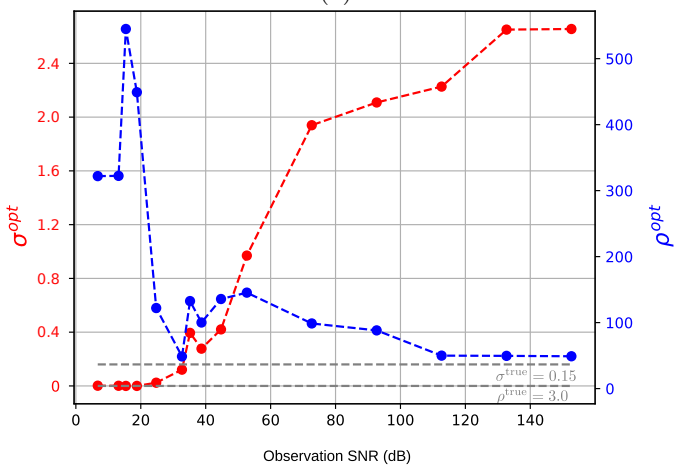
The numerical examples presented in sections 9.1 and 9.2 show that our method is able to estimate stiffness parameters from noisy surface or volume displacement observations. One of the main advantages of the proposed Bayesian formulation is that it accounts for measurement noise explicitly. The example shown in section 9.1 shows that the covariance structure of the identified stiffness parameters is highly dependent on the measurement noise level. More precisely, in a low SNR setting, the variance of the parameter of interest tends to be underestimated, and its correlation length overestimated, which is a common result in the Gaussian process regression literature (see, e.g. section 5.4.1 of Rasmussen [2004]). Unlike the classical regularization approach, the use of Whittle-Matérn priors models the variance and correlation length of the parameters of interest explicitly, which makes them



(a)



(b)



(c)

Figure 10: Identified Matérn hyperparameters with respect to observation SNR, using the three PEB approaches: variogram regression (10a), hybrid-MAP estimator with a hyperprior on σ^2 (10b), ML estimator (10c).

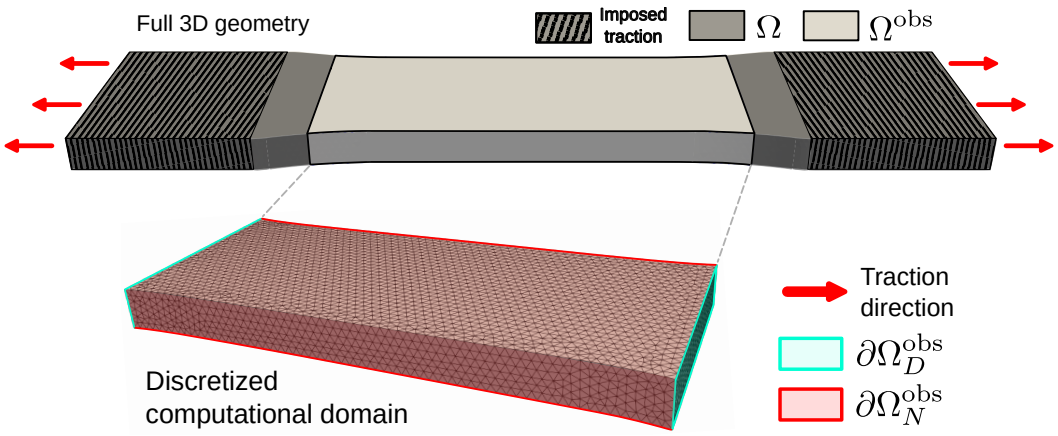


Figure 11: Thick dog-bone geometry for example with volume observations

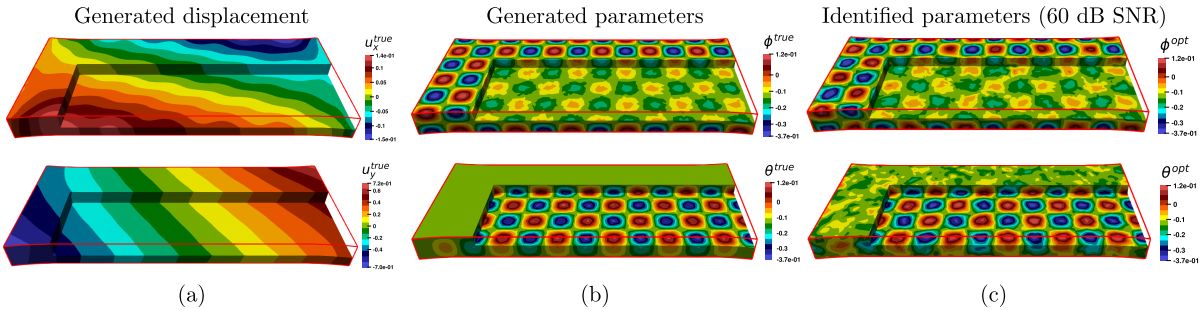
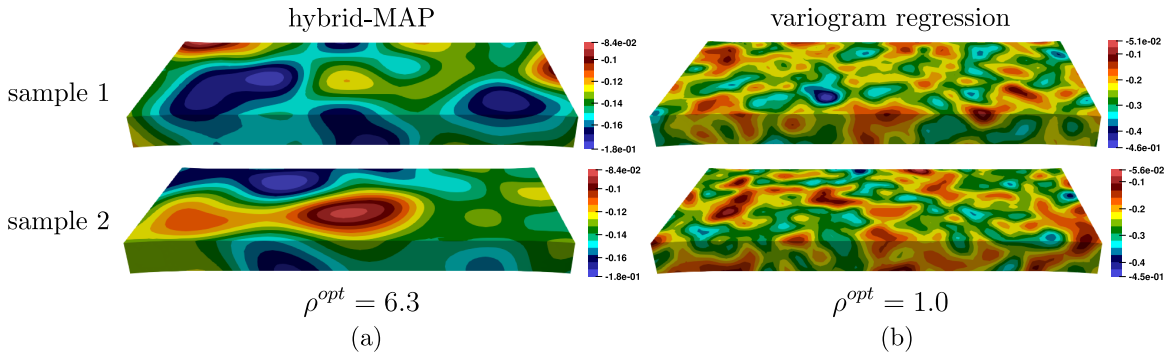


Figure 12: Generated displacement (a), generated parameters ϕ^{true} and θ^{true} (b), and identified parameters ϕ^{opt} and θ^{opt} with a 60 dB SNR (c). Hyperparameter selection was performed with the variogram regression approach.

interpretable with regard to microstructural features. As an ill-posed inverse problem, the Whittle-Matérn prior also acts as a regularizer, and we have shown the similarities with using a classical inverse problem formulation using a gradient penalty regularization. The main advantage of our approach is that formulating a Matérn random field prior on the unknown material parameter leads to a generative model after hyperparameter tuning: combining the identified correlation length ρ, the identified marginal variance σ^2 with a carefully chosen stiffness model $p \mapsto \overline{C}(p)$ enables modelling a wide range of random spatial elastic data $\overline{C}(\underline{x})$.

In this paper, we investigated various approaches for hyperparameters estimation, based on a Parametric Empirical Bayes approach. The various issues related to estimability of the Matérn covariance parameter pointed out in section 6.2 are due to several numerical approximations of the underlying hierarchical Bayesian model, namely iteratively estimating hyperparameters from the data, and the use of sparse approximation of the precision matrix with fixed smoothness parameters. A possible extension of this work would be to consider the full hierarchical model instead of the PEB approximation, and sample from the posterior distribution for $\boldsymbol{p}$ and $\boldsymbol{\mu}$ using MCMC sampling. Considering the recent papers that present numerical methods for a sparse approximation of the Matérn precision matrix for any $\nu > 0$ (see, e.g. [Roininen et al., 2018, Bolin et al., 2018]), MCMC sampling for all Matérn hyperparameters could be investigated, although the computational cost would be much higher than the presented framework. Under the PEB approximation, we've shown that the estimator for hyperparameters must be chosen carefully to avoid numerical instabilities. We've discussed the issues related to the use of a maximum likelihood estimator, and linked the numerical results to the existing literature. The conclusions of our investigations are that the variogram regression approach is the best choice for all cases. Using this approach, we were able to identify the parameter of interest and their covariance structure accurately in both numerical experiments.



$\rho^{opt} = 6.3$ (a) $\rho^{opt} = 1.0$ (b)

Figure 13: Random samples from the prior distribution with tuned Matérn covariance, for the hybrid-MAP approach (a) and the variogram regression approach (b).

In this work, we made the assumption that the covariance model structure was stationary. This may not always be relevant to model materials with non-stationary microstructural features, such as when edge effects are present. In this case, spatially varying hyperparameters could be considered. Some preliminary work is presented in [Roininen et al., 2019] where Whittle-Matérn priors with a spatially varying length scale is considered. The main challenge with this extension is that a correlation structure for the hyperparameters must be specified and tuned. Another limitation of our work is that we did not investigate the case where the stiffness model depends on multiple parameters, each having a different covariance structure. In this case, the number of hyperparameters to identify is increased, and the estimability of all hyperparameters should be investigated. As this paper was focused on identification of linear elastic properties, only a single displacement field observation was required to perform the identification. The Bayesian framework could be extended to time series observations in order to identify spatially varying parameters from a non-linear constitutive law. In order to achieve this, the loss function should account for the observed displacement for all time steps $\boldsymbol{u}_t^{\text{obs}}$, $t = 0, ..., N_t$, and the forward and adjoint problem solvers should be modified accordingly. However, the Bayesian modelling would remain unchanged.

10 Conclusion

The literature on numerical methods for materials with spatially varying, and random elastic properties is scarce, perhaps due to the fact that such materials are used in very specific contexts only. In this work, a Bayesian extension to parameter identification procedures from noisy displacement observations was presented, in which the unknown parameter is modelled as a Matérn random field. Our approach naturally leads to a generative model in which the identified covariance parameters can be used to sample new realisations from the random field prior, and generate spatially varying elasticity tensor fields. Also, the identified covariance hyperparameters are informative about the spatial statistics of underlying microstructural features. The numerical aspects of the proposed framework were presented and discussed. As shown, extending existing procedures to a spatially varying context involves quite a lot of specific challenges, and spatial covariance properties must be accounted for. The approach presented in this paper leverages numerical methods from fields of research largely unrelated to structural mechanics, which we believe to be very relevant for modelling such materials. As explained in the introduction, the authors aim at applying the presented framework to DLF composites, in which spatial covariance of elastic properties is of major importance, and is going to be the main topic of upcoming works.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Hutchinson SA, a company of TotalEnergies, is acknowledged for the financial support of this study. This work was also supported by the French Ministry of Research (ANRT).

A Extension to I-DIC

Adapting the I-DIC method to our framework is similar to the development for FEMU. Following [Leclerc et al., 2009, Réthoré, 2010, Neggers et al., 2017], we introduce the following notations:

- f is the reference image,
- g is the deformed image,
- $\tilde{g}$ is the back transformed image, defined for a given displacement field $\underline{u}$ as

$$\tilde{g}(\underline{x}) = g(\underline{x} + \underline{u}(\underline{x})),$$

- N is the number of pixels of all considered images.

In all this section, an image I of shape $(w, h) \in \mathbb{N}^2$ is represented by a smooth function $I : [0; w] \times [0; h] \to [0; 1]$ where $I(\underline{x})$ is defined by the bicubic interpolation of neighbour pixel values. When written with a bold letter, an image $\boldsymbol{I}$ denotes a $w \times h$ vector collecting pixel values.
The I-DIC loss is formulated in terms of a gray level misfit between the back transformed image and the reference image. In the Bayesian formulation, the gray level noise is assumed to be Gaussian and spatially uncorrelated. We note σ_f the variance of the white Gaussian noise affecting the captured gray level, and as commonly done in, e.g. [Réthoré, 2010], to simplify notations, only the reference image is assumed to be noisy with a variance $2\sigma_f^2$. Let $\underline{u}^{\text{true}}$ denote the true displacement field that maps f to g. Then

$$\forall 1 \leq i \leq N,\ g(\underline{x}_i + \underline{u}^{\text{true}}(\underline{x}_i))) - f(\underline{x}_i) \sim \mathcal{N}(0, 2\sigma_f^2). \tag{101}$$

The I-DIC loss can be written as the following likelihood:

$$J_I(\underline{u}(\boldsymbol{p})) = \frac{1}{2\sigma_f^2} \sum_{i=1}^{N} \left(g(\underline{x}_i + \underline{u}(\boldsymbol{p})(\underline{x}_i)) - f(\underline{x}_i)\right)^2. \tag{102}$$

Using (102) instead of (35) inside (34), the loss function for the MAP estimate using the I-DIC approach now writes

$$J_{\text{IDIC}}(\boldsymbol{p}) = J_I(\boldsymbol{p}) + J_F(\boldsymbol{p}) + J_{\boldsymbol{p}}(\boldsymbol{p}), \tag{103}$$

where J_F and $J_{\boldsymbol{p}}$ have the same expressions as for the FEMU method. Before deriving the parameter sensitivities for the I-DIC loss, it is important to note that the displacement field $\underline{u}$ present in J_I is not exactly the solution to the forward problem, but the simulated displacement mapped to image coordinates, as the simulated displacement $\underline{u}$ is defined on Ω and $\underline{u}$ is defined on $[0; w] \times [0; h]$. Hence, we introduce an invertible affine transformation $\underline{\phi}$ satisfying $\underline{\phi}(\Omega^{\text{obs}}) \subset [0; w] \times [0; h]$ that maps the computational domain to the reference image, and we pose $\underline{\psi} = \underline{\phi}^{-1}$. Also, after discretization, $\underline{u}$ is a displacement image that is not represented using finite elements shape functions. To avoid any confusion, we hereafter use the following notations:

- $\underline{u} : \Omega \to \mathbb{R}^3$ is the discrete forward problem solution, expressed in a finite element basis $(\underline{\varphi}_0,\ldots,\underline{\varphi}_{N_V})$ as

$$\underline{u} = \sum_{i=0}^{N_V} u_i \underline{\varphi}_i,$$

- $\underline{\tilde{u}} : [0; w] \times [0; h] \to \mathbb{R}^3$ is the forward problem solution in image coordinates, defined as

$$\underline{\tilde{u}}(\underline{x}) = \mathrm{d}\underline{\psi} \circ \underline{u} \circ \underline{\psi},$$

- $\underline{U} : [\![0; w]\!] \times [\![0; h]\!] \to \mathbb{R}^2$ is a discrete displacement image, that is linked to $\tilde{\underline{u}}$ through some image interpolation operator $\underline{\underline{\Pi}}$ such that $\underline{U}(\underline{x}_i) = (\underline{\underline{\Pi}} \circ \tilde{\underline{u}})(\underline{x}_i)$.

Also, in regard to the previously introduced discrete notations, $\boldsymbol{u} = (u_0, ..., u_n)$ denotes the coordinates of $\underline{u}$ in basis $\mathcal{B} = (\underline{\varphi}_0, ..., \underline{\varphi}_n)$, $\widetilde{\boldsymbol{u}} = (\widetilde{u}_0, ..., \widetilde{u}_n)$ denotes the coordinates of $\underline{u}$ in basis $\widetilde{\mathcal{B}} = (\widetilde{\underline{\varphi}}_0, ..., \widetilde{\underline{\varphi}}_n)$ where

$$\forall 0 \leq i \leq N_{\underline{V}},\ \widetilde{\underline{\varphi}}_i = \mathrm{d}\underline{\psi} \circ \underline{\varphi}_i \circ \underline{\psi},$$

and $\boldsymbol{U}$ is a $2 \times w \times h$ tensor such that $\boldsymbol{U}_{ikl} = \underline{U}(k, l)_i$. With these notations, equation (102) now writes after discretization

$$J_I = \frac{1}{2\sigma_f^2} \sum_{i=1}^{N} \left(g\left(\underline{x}_i + \left(\underline{\underline{\Pi}} \circ \mathrm{d}\underline{\psi} \circ \underline{u} \circ \underline{\psi}(\underline{x}_i)\right)\right) - f(\underline{x}_i)\right)^2. \tag{104}$$

The adjoint variational problem is derived in a same manner as for FEMU, the main difference is the expression of the source term:

$$\frac{\mathrm{d}J_{\mathrm{IDIC}}}{\mathrm{d}\boldsymbol{u}} = \frac{\mathrm{d}J_I}{\mathrm{d}\boldsymbol{u}} + \frac{\mathrm{d}J_F}{\mathrm{d}\boldsymbol{u}}. \tag{105}$$

The derivative of J_F remains unchanged, and the derivative of J_I is expanded using the chain rule

$$\frac{\mathrm{d}J_I}{\mathrm{d}\boldsymbol{u}} = \frac{1}{\sigma_f^2}(\tilde{\boldsymbol{g}} - \boldsymbol{f}) \frac{\mathrm{d}\tilde{\boldsymbol{g}}}{\mathrm{d}\boldsymbol{U}} \frac{\mathrm{d}\boldsymbol{U}}{\mathrm{d}\widetilde{\boldsymbol{u}}} \frac{\mathrm{d}\widetilde{\boldsymbol{u}}}{\mathrm{d}\boldsymbol{u}}. \tag{106}$$

In this last expression, each term of the chain rule expansion is explained below:

- $\frac{\mathrm{d}\widetilde{\boldsymbol{u}}}{\mathrm{d}\boldsymbol{u}}$ is a block-diagonal sparse matrix of size $(3(N_{\underline{V}} + 1))^2$ that is the discrete representation of $\mathrm{d}\underline{\psi}$ in basis $\mathcal{B}$. If we note $\underline{\psi}(\underline{x}) = \underline{\underline{A}} \cdot \underline{x} + \underline{b}$, then

$$\frac{\mathrm{d}\widetilde{\boldsymbol{u}}}{\mathrm{d}\boldsymbol{u}} = \begin{pmatrix} \underline{\underline{A}} & & 0 \\ & \ddots & \\ 0 & & \underline{\underline{A}} \end{pmatrix}, \tag{107}$$

- $\frac{\mathrm{d}\boldsymbol{U}}{\mathrm{d}\widetilde{\boldsymbol{u}}}$ is a sparse matrix of size $(2N) \times (3N_{\underline{V}})$ that stores the finite element basis values at pixel coordinates. In terms of interpolation, if $\boldsymbol{U}$ is seen as a vector valued image, then that image can be interpreted as a structured mesh made of quadrilateral cells, on which can be constructed a P^1 finite element basis $\mathcal{B}_{\mathrm{pix}}$. That being said, $\frac{\mathrm{d}\boldsymbol{U}}{\mathrm{d}\widetilde{\boldsymbol{u}}}$ is simply the discrete interpolator from $\widetilde{\mathcal{B}}$ to $\mathcal{B}_{\mathrm{pix}}$. The matrix entries are given by

$$\frac{\mathrm{d}\boldsymbol{U}}{\mathrm{d}\widetilde{\boldsymbol{u}}} = \left((\underline{\varphi}_k(\underline{x}_l))_i \right)_{\substack{i=0,1 \\ 0 \leq k \leq N_{\underline{V}} \\ 0 \leq l < N}} \tag{108}$$

  where $\underline{x}_l$ denotes pixel coordinates,

- $\frac{\mathrm{d}\tilde{\boldsymbol{g}}}{\mathrm{d}\widetilde{\boldsymbol{u}}}$ is a block-diagonal sparse matrix of size $N \times 2N$ whose entries are made of the back-transformed image gradient $\nabla\tilde{\boldsymbol{g}}$. There are multiple approaches to compute the discrete gradient of an image, but in order to remain consistent with the bicubic gray level interpolation, we use a centred finite difference scheme:

$$(\nabla_x \tilde{\boldsymbol{g}})(\underline{x}) = \lim_{\epsilon \to 0} \frac{\tilde{\boldsymbol{g}}(\underline{x} + \epsilon \underline{e}_x) - \tilde{\boldsymbol{g}}(\underline{x} - \epsilon \underline{e}_x)}{2\epsilon}, \tag{109}$$

$$(\nabla_y \tilde{\boldsymbol{g}})(\underline{x}) = \lim_{\epsilon \to 0} \frac{\tilde{\boldsymbol{g}}(\underline{x} + \epsilon \underline{e}_y) - \tilde{\boldsymbol{g}}(\underline{x} - \epsilon \underline{e}_y)}{2\epsilon}, \tag{110}$$

  where $\tilde{\boldsymbol{g}}(\underline{x} \pm \epsilon e_x)$ is defined through bicubic interpolation, $\underline{e}_x = (1, 0)$ and $\underline{e}_y = (0, 1)$. From there,

$$\frac{\mathrm{d}\tilde{\boldsymbol{g}}}{\mathrm{d}\widetilde{\boldsymbol{u}}} = \begin{pmatrix} \nabla\tilde{\boldsymbol{g}}(\underline{x}_0) & & 0 \\ & \ddots & \\ 0 & & \nabla\tilde{\boldsymbol{g}}(\underline{x}_N) \end{pmatrix} \tag{111}$$

  where

$$\nabla\tilde{\boldsymbol{g}}(\underline{x}_l) = \left((\nabla_x \tilde{\boldsymbol{g}})(\underline{x}_l), (\nabla_y \tilde{\boldsymbol{g}})(\underline{x}_l) \right).$$

In (111), we note that the back transformed image gradient is also commonly used for implementing standard DIC, and following the state-of-the-art implementation techniques, it was shown in, e.g., [Leclerc et al., 2009, Neggers et al., 2017, Réthoré, 2010], that $\nabla \tilde{\boldsymbol{g}}$ can be approximated by $\nabla \boldsymbol{f}$ such that it does not need to be recomputed at each iteration. For application of Dirichlet boundary conditions, the simplest approach is to use the correlated displacement $\boldsymbol{u}^{\text{obs}}$ from a DIC algorithm, as commonly done in the literature [Leclerc et al., 2009, Neggers et al., 2017].

B Construction of Observation Noise Matrices

In this section, we reuse notations introduced in appendix A, and discuss the expression of the precision matrix used in the likelihood expression (31). As explained in section 5.2, it largely depends on the algorithm used to perform the correlation. An interesting case is the FE-DIC approach, in which the correlation is performed on a finite element mesh instead of a structured grid. For details about FE-DIC, we refer to [Besnard et al., 2006]. In this case, the displacement precision matrix can be derived analytically. Considering that the reference image f is corrupt by a Gaussian white noise with variance $2\sigma_f^2$, the noise analysis performed in [Besnard et al., 2006, Réthoré, 2010] shows the following expression for the displacement precision matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} = \frac{1}{2\sigma_f^2} \boldsymbol{N}^T (\nabla \boldsymbol{f})^T \nabla \boldsymbol{f} \boldsymbol{N} \tag{112}$$

where $\boldsymbol{N}$ is a sparse matrix of size $(2N) \times N_{\underline{V}}$ containing the evaluation of FEM basis functions at pixel values. Note that considering the notations of appendix A, $\boldsymbol{N} = \frac{\mathrm{d}\boldsymbol{U}}{\mathrm{d}\widetilde{\boldsymbol{u}}}$. For a specific choice of finite elements, the marginal variance of the displacement noise can be further expanded (see [Besnard et al., 2006] for the demonstration with quadrilateral elements). In the case of multi-camera setups, such as in stereo-DIC, the same noise sensitivity analysis can be performed. When using the FE-stereo-DIC correlation algorithm [Pierré et al., 2016], the derivation of the noise precision can be found in equation 12 of [Hild and Roux, 2020]. The case of I-DIC is simpler in a multi camera setup as the gray level noise remains spatially uncorrelated, and the precision matrix can be derived easily from the gray level uncertainty of each camera.
In some contexts, a closed form expression may not be available, for instance when the DIC algorithm is unknown (e.g. when using commercial software that hides implementation details). An empirical approach may be used to construct the precision matrix numerically. First, a set of n correlated displacements $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$ is obtained by correlating the reference image f with $f + \xi$ where $\xi \sim \mathcal{N}(0, 2\sigma_f^2)$. The covariance matrix can then be estimated, with

$$\boldsymbol{\Sigma}_{\boldsymbol{u}} \approx \sum_{k=1}^{n} (\boldsymbol{u}_k - \boldsymbol{u}^0)(\boldsymbol{u}_k - \boldsymbol{u}^0)^T, \tag{113}$$

$$\boldsymbol{u}^0 = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{u}_k. \tag{114}$$

The resulting covariance matrix is dense and of rank $r \leq n$. If $n < N_{\underline{V}}$, $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ is not invertible and thus cannot be used to compute the precision matrix. A common technique is to employ a ℓ^1-penalized maximum likelihood to obtain an estimator

$$\widehat{\boldsymbol{Q}} = \underset{\boldsymbol{Q}}{\operatorname{argmin}} \left(\operatorname{tr}(\boldsymbol{\Sigma}_{\boldsymbol{u}} \boldsymbol{Q}) - \log \det \boldsymbol{Q} + \gamma \sum_{i \neq j} S(|\boldsymbol{Q}_{ij}|) \right), \tag{115}$$

where γ is a regularization parameter that favours sparse solutions and S a shrinkage function. For details, we refer to [Fan et al., 2015].

References

Hugo Leclerc, Jean-Noël Périé, Stephane Roux, and François Hild. Integrated digital image correlation for the identification of mechanical properties. volume 5496, pages 161–171, 05 2009. ISBN 978-3-642-01810-7. doi: 10.1007/978-3-642-01811-4_15.

Julien Réthoré. A fully integrated noise robust strategy for the identification of constitutive laws from digital images. *International Journal for Numerical Methods in Engineering*, 84:631–660, 2010. doi: 10.1002/nme.2908. URL https://hal.archives-ouvertes.fr/hal-00938286.

Stéphane Avril, Marc Bonnet, Anne Sophie Bretelle, Michel Grediac, François Hild, Patrick Ienny, Félix Latourte, Didier Lemosse, Stéphane Pagano, Emmanuel Pagnacco, and Fabrice Pierron. Overview of identification methods of mechanical parameters based on full-field measurements. *Experimental Mechanics*, 48(4):381–402, August 2008. doi: 10.1007/s11340-008-9148-y. URL https://hal.archives-ouvertes.fr/hal-00274639.

J. Neggers, F. Mathieu, F. Hild, S. Roux, and N. Swiergiel. Improving full-field identification using progressive model enrichments. *International Journal of Solids and Structures*, 118-119:213–223, 2017. ISSN 0020-7683. doi: https://doi.org/10.1016/j.ijsolstr.2017.03.013. URL https://www.sciencedirect.com/science/article/pii/S0020768317301191.

Luca M. Martulli, Leen Muyshondt, Martin Kerschbaum, Soraia Pimenta, Stepan V. Lomov, and Yentl Swolfs. Carbon fibre sheet moulding compounds with high in-mould flow: Linking morphology to tensile and compressive properties. *Composites Part A: Applied Science and Manufacturing*, 126:105600, 2019. ISSN 1359-835X. doi: https://doi.org/10.1016/j.compositesa.2019.105600. URL https://www.sciencedirect.com/science/article/pii/S1359835X19303495.

Marco Alves, Yizhuo Li, and Soraia Pimenta. Spatial variability and characteristic length-scales of strain fields in tow-based discontinuous composites: Characterisation and modelling. *Composites Part B: Engineering*, 262:110789, 2023. ISSN 1359-8368. doi: https://doi.org/10.1016/j.compositesb.2023.110789. URL https://www.sciencedirect.com/science/article/pii/S1359836823002925.

Sebastian Geyer, Iason Papaioannou, and Daniel Straub. Spatial modeling of concrete strength based on data. *Structural Safety*, 103:102345, 2023. ISSN 0167-4730. doi: https://doi.org/10.1016/j.strusafe.2023.102345. URL https://www.sciencedirect.com/science/article/pii/S0167473023000322.

Yoshimi Watanabe, Hisashi Sato, and Yasuyoshi Fukui. Wear properties of intermetallic compound reinforced functionally graded materials fabricated by centrifugal solid-particle and in-situ methods. *Journal of Solid Mechanics and Materials Engineering*, 2(7):842–853, 2008. doi: 10.1299/jmmp.2.842.

M. T. Nair. On morozov's discrepancy principle for nonlinear ill-posed equations. *Bulletin of the Australian Mathematical Society*, 79(2):337–342, 2009. doi: 10.1017/S0004972708001342.

Per Christian Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Review*, 34(4):561–580, 1992. doi: 10.1137/1034115. URL https://doi.org/10.1137/1034115.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011. doi: https://doi.org/10.1111/j.1467-9868.2011.00777.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x.

Michael Sutton, Jean-José Orteu, and Hubert Schreier. *Image Correlation for Shape, Motion and Deformation Measurements. Basic Concepts, Theory and Applications*. 01 2009. ISBN 978-0-387-78746-6. doi: 10.1007/978-0-387-78747-3.

Andre P. Ruybalid, Johan P. M. Hoefnagels, Olaf van der Sluis, and Marc G. D. Geers. Comparison of the identification performance of conventional fem updating and integrated dic. *International Journal for Numerical Methods in Engineering*, 106(4):298–320.

D. Thomas Seidl, Assad A. Oberai, and Paul E. Barbone. The coupled adjoint-state equation in forward and inverse linear elasticity: Incompressible plane stress. *Computer Methods in Applied Mechanics and Engineering*, 357:112588, 2019. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2019.112588. URL https://www.sciencedirect.com/science/article/pii/S0045782519304591.

Marc Bonnet and A. Constantinescu. Inverse problems in elasticity. *Inverse Problems*, 21:R1–R50, 2005. doi: 10.1088/0266-5611/21/2/R01. URL https://hal.archives-ouvertes.fr/hal-00111264.

Evan Ricketts, Peter Cleall, Anthony Jefferson, Pierre Kerfriden, and Paul Lyons. Representation of three-dimensional unsaturated flow in heterogeneous soil through tractable gaussian random fields. *Géotechnique*, pages 1–34, 08 2023. doi: 10.1680/jgeot.22.00316.

Lu Hai and Meng-Ze Lyu. Modeling tensile failure of concrete considering multivariate correlated random fields of material parameters. *Probabilistic Engineering Mechanics*, 74:103529, 2023. ISSN 0266-8920. doi: https://doi.org/10.1016/j.probengmech.2023.103529. URL https://www.sciencedirect.com/science/article/pii/S0266892023001182.

Srinivas Sriramula and Marios K. Chryssanthopoulos. Quantification of uncertainty modelling in stochastic analysis of frp composites. *Composites Part A: Applied Science and Manufacturing*, 40 (11):1673–1684, 2009. ISSN 1359-835X. doi: https://doi.org/10.1016/j.compositesa.2009.08.020. URL https://www.sciencedirect.com/science/article/pii/S1359835X09002577.

Masoumeh Dashti and Andrew M. Stuart. The bayesian approach to inverse problems, 2015.

Lassi Roininen, Janne M. J. Huttunen, and Sari Lasanen. Whittle-matérn priors for bayesian statistical inversion with applications in electrical impedance tomography, 2014. ISSN 1930-8337. URL http://aimsciences.org/article/doi/10.3934/ipi.2014.8.561.

Lassi Roininen, Mark Girolami, Sari Lasanen, and Markku Markkanen. Hyperpriors for matérn fields with applications in bayesian inversion, 2019. ISSN 1930-8337. URL http://aimsciences.org/article/doi/10.3934/ipi.2019001.

Christopher Bishop. *Pattern Recognition and Machine Learning*, volume 16, pages 140–155. 01 2006. doi: 10.1117/1.2819119.

Finn Lindgren, David Bolin, and Håvard Rue. The spde approach for gaussian and non-gaussian fields: 10 years and still running. *Spatial Statistics*, 50:100599, 2022. ISSN 2211-6753. doi: https://doi.org/10.1016/j.spasta.2022.100599. URL https://www.sciencedirect.com/science/article/pii/S2211675322000057. Special Issue: The Impact of Spatial Statistics.

David Bolin, Kristin Kirchner, and Mihá ly Kovács. Numerical solution of fractional elliptic stochastic PDEs with spatial white noise. *IMA Journal of Numerical Analysis*, 40(2):1051–1073, dec 2018. doi: 10.1093/imanum/dry091. URL https://doi.org/10.1093%2Fimanum%2Fdry091.

Lassi Roininen, Sari Lasanen, Mikko Orispää, and Simo Särkkä. Sparse approximations of fractional matérn fields. *Scandinavian Journal of Statistics*, 45(1):194–216, 2018. doi: https://doi.org/10.1111/sjos.12297. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12297.

P. Diggle and P. Ribiero. Model–based geostatistics, 2007.

Maitreyee Bose, James S. Hodges, and Sudipto Banerjee. Toward a diagnostic toolkit for linear models with gaussian-process distributed random effects, 2018.

Victor De Oliveira and Zifei Han. On information about covariance parameters in gaussian matérn random fields. *Journal of Agricultural, Biological and Environmental Statistics*, 27(4):690–712, aug 2022. doi: 10.1007/s13253-022-00510-5. URL https://doi.org/10.1007%2Fs13253-022-00510-5.

Hao Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261, 02 2004. doi: 10.1198/016214504000000241.

Wenpin Tang, Lu Zhang, and Sudipto Banerjee. On identifiability and consistency of the nugget in gaussian spatial process models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(5):1044–1070, oct 2021. doi: 10.1111/rssb.12472. URL https://doi.org/10.1111%2Frssb.12472.

Vidhi Lalchand, Wessel P. Bruinsma, David R. Burt, and Carl E. Rasmussen. Sparse gaussian process hyperparameters: Optimize or integrate?, 2022.

Richard D Brown, Johnathan M Bardsley, and Tiangang Cui. Semivariogram methods for modeling whittle–maté rn priors in bayesian inverse problems. *Inverse Problems*, 36(5):055006, may 2020. doi: 10.1088/1361-6420/ab762e. URL https://doi.org/10.1088%2F1361-6420%2Fab762e.

Noel Cressie. Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, 1985. doi: 10.1007/BF01032109. URL https://doi.org/10.1007/BF01032109.

Mårten Gulliksson. Analyzing the nonlinear l-curve. 01 1998.

Heinz W. Engl and Wilhelm Grever. Using the l–curve for determining optimal regularization parameters. *Numerische Mathematik*, 69, 1994. doi: https://doi.org/10.1007/s002110050078.

Alessandro Cultrera and Luca Callegaro. A simple algorithm to find the l-curve corner in the regularisation of ill-posed inverse problems. *IOP SciNotes*, 1(2):025004, aug 2020. doi: 10.1088/2633-1357/abad0d. URL https://dx.doi.org/10.1088/2633-1357/abad0d.

Nicholas G. Polson and Vadim Sokolov. Bayesian regularization: From tikhonov to horseshoe, 2019.

Mathias Richter. *Inverse problems: Basics, theory and applications in geophysics*. Springer Nature, 2021.

Kyriakos C. Giannakoglou and Dimitrios I. Papadimitriou. *Adjoint Methods for Shape Optimization*, pages 79–108. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-72153-6. doi: 10.1007/978-3-540-72153-6_4. URL https://doi.org/10.1007/978-3-540-72153-6_4.

Mark Asch, Marc Bocquet, and Maëlle Nodet. *Data assimilation: methods, algorithms, and applications*. SIAM, 2016.

Dong C. Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45:503–528, 1989. doi: 10.1007/BF01589116. URL https://doi.org/10.1007/BF01589116.

Michal Habera, Jack S. Hale, Chris Richardson, Johannes Ring, Marie Rognes, Nate Sime, and Garth N. Wells. FEniCSX: A sustainable future for the FEniCS Project. 2 2020. doi: 10.6084/m9.figshare.11866101.v1. URL https://figshare.com/articles/presentation/FEniCSX_A_sustainable_future_for_the_FEniCS_Project/11866101.

Martin S. Alnæs, Anders Logg, Kristian B. Ølgaard, Marie E. Rognes, and Garth N. Wells. Unified form language: A domain-specific language for weak formulations of partial differential equations. *ACM Trans. Math. Softw.*, 40(2), mar 2014. ISSN 0098-3500. doi: 10.1145/2566630. URL https://doi.org/10.1145/2566630.

Satish Balay, Shrirang Abhyankar, Mark F. Adams, Steven Benson, Jed Brown, Peter Brune, Kris Buschelman, Emil M. Constantinescu, Lisandro Dalcin, Alp Dener, Victor Eijkhout, William D. Gropp, Václav Hapla, Tobin Isaac, Pierre Jolivet, Dmitry Karpeev, Dinesh Kaushik, Matthew G. Knepley, Fande Kong, Scott Kruger, Dave A. May, Lois Curfman McInnes, Richard Tran Mills, Lawrence Mitchell, Todd Munson, Jose E. Roman, Karl Rupp, Patrick Sanan, Jason Sarich, Barry F. Smith, Stefano Zampini, Hong Zhang, Hong Zhang, and Junchao Zhang. PETSc Web page. https://petsc.org/, 2022. URL https://petsc.org/.

S. Müller, L. Schüler, A. Zech, and F. Heße. `GSTools` v1.3: a toolbox for geostatistical modelling in python. *Geoscientific Model Development*, 15(7):3161–3182, 2022. doi: 10.5194/gmd-15-3161-2022. URL https://gmd.copernicus.org/articles/15/3161/2022/.

Sebastian Müller and Lennart Schüler. Geostat-framework/gstools: v1.5.0 'nifty neon', June 2023. URL https://doi.org/10.5281/zenodo.8044720.

Ian Knowles and Robert J. Renka. Methods for numerical differentiation of noisy data. 2014.

Carl Edward Rasmussen. *Gaussian Processes in Machine Learning*, pages 63–71. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-540-28650-9. doi: 10.1007/978-3-540-28650-9_4. URL https://doi.org/10.1007/978-3-540-28650-9_4.

G. Besnard, François Hild, and Stephane Roux. "finite-element" displacement fields analysis from digital images: Application to portevin–le châtelier bands. *Experimental Mechanics*, 46, 12 2006. doi: 10.1007/s11340-006-9824-8.

Jean-Emmanuel Pierré, Jean-Charles Passieux, and Jean-Noël Périé. Finite element stereo digital image correlation: Framework and mechanical regularization. *Experimental Mechanics*, 57, 12 2016. doi: 10.1007/s11340-016-0246-y.

François Hild and Stephane Roux. Displacement uncertainties with multiview correlation schemes. *The Journal of Strain Analysis for Engineering Design*, 55, 06 2020. doi: 10.1177/0309324720927102.

Jianqing Fan, Yuan Liao, and Han Liu. An overview on the estimation of large covariance and precision matrices, 2015.