



HAL
open science

Probabilistic learning inference constrained by an uncertain model and a target: A general method with application to elasticity homogenization without scale separation

Christian Soize

► To cite this version:

Christian Soize. Probabilistic learning inference constrained by an uncertain model and a target: A general method with application to elasticity homogenization without scale separation. F. Willot; J. Dirrenberger; S. Forest; D. Jeulin; A. Cherkaev. Continuum Models and Discrete Systems, Mathematics and Statistics 457, Springer Nature Switzerland; Springer, pp.1-14, 2024, Springer Proceedings in Mathematics & Statistics, 978-3-031-58664-4. 10.1007/978-3-031-58665-1_1 . hal-04683074

HAL Id: hal-04683074

<https://univ-eiffel.hal.science/hal-04683074v1>

Submitted on 1 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Probabilistic learning inference constrained by an uncertain model and a target: A general method with application to elasticity homogenization without scale separation

Christian Soize

Abstract We present a probabilistic learning inference that assimilates data (target set) into a parameterized large stochastic computational model resulting from discretizing a stochastic boundary value problem (BVP). A target is imposed on a vector-valued random quantity of interest (QoI), observed as the stochastic solution of the BVP. The probabilistic inference estimates the posterior probability model, which is constrained both by the second-order moment of the random residue of the BVP stochastic equations and the target set composed of statistical moments of the QoI. We assume that evaluating a single realization of the BVP is computationally expensive, so the training dataset comprises only a few points differing from big data approaches. The presented application contributes to three-dimensional stochastic homogenization of heterogeneous linear elastic media, specifically when the mesoscale and macroscale are not separated.

1 Introduction

(i) Objective. This paper addresses an important question concerning probabilistic learning algorithms that allow data (target dataset) to be assimilated into predictive models. These models have a small training dataset and the target dataset consists of statistical moments of quantities of interest (QoI). It assumes that these statistical moments, such as mean values and second-order moments, were estimated using realizations (samples) that are no longer available. This situation often occurs when the data is lost, deteriorated, not commented on, no longer exists, or is inaccessible. However, some statistical moment values may have been published or provided in technical reports. Additional physics-based constraints are simultaneously considered, consisting of controlling the learning process with the model to

Christian Soize
Université Gustave Eiffel, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée,
France e-mail: christian.soize@univ-eiffel.fr

minimize the mean-square norm of the random normalized residue of the stochastic partial differential equation of the boundary value problem (BVP). These two types of constraints are both defined by an implicit function denoted as \mathbf{h}^c , for which an algebraic representation is unavailable. They ensure that the algorithm aims to bring the statistical moments closer to their targets while maintaining a small residue of the BVP stochastic equations. The framework presented uses a computational model with a computationally expensive single evaluation. Therefore, the training set comprises a small number of samples (or points), differing from big data approaches. The presented general method for probabilistic learning inference is applied to three-dimensional stochastic homogenization, specifically in scenarios where the mesoscale and macroscale are not separated. It is worth noting that stochastic homogenization has been extensively studied in numerous works (e.g., [35, 1, 13, 5]), with particular attention given to analyzing the size of the representative volume element (RVE) (see [21, 19, 24, 4]), and the morphological models of random structures (see [11, 12]). The presented application focuses on the stochastic homogenization of a random linear elastic medium at the mesoscale, defined on the domain $\Omega \subset \mathbb{R}^3$ of the microstructure. It is assumed that this domain is not a RVE, which means that there is no scale separation between the mesoscale and the macroscale. Since there is no scale separation, the macroscale is in fact another mesoscale at a larger scale than the initial mesoscale, which is described by random effective/apparent quantities. For the construction of the posterior probability measure using the proposed probabilistic learning inference, the constraints are defined by both a target set consisting of given statistical moments of the random effective/apparent elasticity tensor and the second-order moment of the random normalized residue of the BVP stochastic equations.

(ii) *A few words on the main methodologies allowing for addressing the problem.*

The main statistical methods for addressing this type of problems are the following. The *Bayesian method* estimates the posterior probability measure of the control parameter using prior probability and a target dataset of realizations for the quantity of interest (QoI). However, this article introduces a different hypothesis that does not necessitate knowledge of realizations, which are assumed to be unavailable. It is important to note that Bayesian inference can be challenging to use, especially in high dimensions [20]. The *maximum-likelihood method* is employed to estimate the hyperparameters of a probability measure using a target dataset of realizations. Similar to Bayesian inference, the maximum-likelihood method is an effective statistical approach for estimating hyperparameters, but it relies on realizations as the target. The *parameterized functional representations* of random quantities, whether in finite or infinite dimensions, such as spectral methods like polynomial chaos expansions ([7, 27, 34]), facilitate the integration of data into predictive models. Bayesian inference, maximum likelihood, and least-square approaches are commonly used to identify the parameters of these representations. The *deep machine learning method*, which relies on Artificial Neural Networks (ANNs), has gained considerable recognition in diverse scientific and engineering disciplines. However, it is important to note that ANNs usually rely on large datasets for accurate parameter identification

[22] and may not be straightforward to use within a comprehensive probabilistic framework, such as the one proposed in this paper.

(iii) *Proposed methodology.* In this work, we present a probabilistic learning method under implicit constraints. The Kullback-Leibler minimum principle [16, 14, 3] is employed to estimate the posterior probability measure that satisfies the constraints. This principle uses the prior probability measure and constraints related to statistical moments, given as targets. The Kullback-Leibler minimum principle has been used in previous works [29, 31, 32, 26]. The posterior probability measure is represented by an algebraic expression involving the prior probability measure and a Lagrange multiplier vector, λ , associated with the function \mathbf{h}^c . The optimal value of the Lagrange multipliers, λ^{sol} , is obtained as the limit of a sequence $\{\lambda_i\}_i$, allowing for the construction of a sequence of probability measures, where the limit corresponds to the desired posterior probability measure. Due to the high-dimensional framework and a small training dataset, the Probabilistic Learning on Manifolds (PLoM) [28, 30, 33] is used to generate the constrained learned dataset for each value of λ_i . The generation is based on a Markov Chain Monte Carlo (MCMC) method employing a nonlinear Itô stochastic differential equation (ISDE) associated with a nonlinear stochastic dissipative Hamiltonian dynamical system. The inclusion of a dissipative term allows for the rapid elimination of the transient part and facilitates the attainment of the stationary response associated with the invariant measure. Evaluating the drift vector of the ISDE involves computing the gradient of function \mathbf{h}^c multiple times. As an algebraic expression of this gradient is not available, direct numerical calculations considering the high dimensionality are not feasible. In [26], a method is proposed to construct an explicit statistical surrogate model, \mathbf{h}^N , of implicit function \mathbf{h}^c . The surrogate model depends on the number N of points generated in the constrained learned dataset, and its gradient has an explicit algebraic representation.

(iv) *Difficulties and novelties.* In the broader context of the proposed methodology, this statistical inverse problem poses several challenges. Firstly, there is a significant computational cost associated with evaluating a large Stochastic Computational Model (SCM) resulting from the discretization of the stochastic BVP. Consequently, a limitation arises from the availability of a small training dataset. Secondly, the random parameters contribute to a high stochastic dimensionality. Additionally, the constraints are defined by an implicit, nonlinear mapping operating in a high-dimensional space. To overcome each of these challenges, the author provides a brief overview of the recently proposed probabilistic learning methodology [26]. This general method is applied to 3D elasticity homogenization without scale separation.

2 Formulation of the probabilistic learning inference using the Kullback-Leibler divergence

(i) *Stochastic boundary value problem and quantity of interest.* All the random quantities are defined on a probability space $(\Theta, \mathcal{F}, \mathcal{P})$. We consider a stochastic elliptic

BVP on the open bounded domain $\Omega \subset \mathbb{R}^d$ (for instance, $d = 3$), whose partial differential equation (PDE) is written as $\mathcal{N}(\mathbf{Y}, \mathbf{G}, \mathbf{W}) = \mathbf{0}$ a.s. The unknown is the non-Gaussian vector-valued field $\{\mathbf{Y}(\xi), \xi \in \Omega\}$ that satisfies the boundary conditions. The coefficients of the stochastic elliptic operator depend on a non-Gaussian second-order vector-valued random field \mathbf{G} and on a random vector-valued control parameter \mathbf{W} . It is assumed that the weak formulation of this stochastic BVP admits a unique strong stochastic solution $\mathbf{Y} = \mathbf{f}(\mathbf{G}, \mathbf{W})$ that is a second-order random field. The mapping \mathbf{f} is not explicitly known. An approximation is constructed, on the one hand by using the finite element method to discretize the weak formulation, and on the other hand by using the numerical method of Monte Carlo to approximate the strong stochastic solution. The quantity of interest (observation) is a second-order vector-valued random variable $\mathbf{Q} = \mathcal{O}(\mathbf{X})$ with $\mathbf{X} = (\mathbf{Y}, \mathbf{G}, \mathbf{W})$, and where the observation operator, \mathcal{O} , is a given measurable mapping. This parameterization is summarized in Fig. 1.

(ii) *Construction of the small training dataset.* A prior probability model of $\{\mathbf{G}, \mathbf{W}\}$, represented by the joint probability measure $P_{\mathbf{G}, \mathbf{W}}^{\text{prior}}$, is constructed, along with its generator of independent realizations. The subscript "d" designates the quantities related to the training dataset, which is constructed by using the Monte Carlo numerical simulation method. Let $\{\mathbf{g}_d^1, \dots, \mathbf{g}_d^{N_d}\}$ and $\{\mathbf{w}_d^1, \dots, \mathbf{w}_d^{N_d}\}$ independent realizations of $\{\mathbf{G}, \mathbf{W}\}$, generated using the prior probability model of $\{\mathbf{G}, \mathbf{W}\}$. Each realization \mathbf{y}_d^j is computed by solving the weak formulation of the PDE $\mathcal{N}(\mathbf{y}_d^j, \mathbf{g}_d^j, \mathbf{w}_d^j) = \mathbf{0}$ with the boundary conditions. Consequently, N_d independent realizations $\{\mathbf{y}_d^j, j = 1, \dots, N_d\}$ of the random field \mathbf{Y} are computed, where $\mathbf{y}_d^j = \mathbf{f}(\mathbf{g}_d^j, \mathbf{w}_d^j)$. The N_d independent realizations $\{\mathbf{q}_d^j, j = 1, \dots, N_d\}$ of the random observation \mathbf{Q} are deduced as $\mathbf{q}_d^j = \mathcal{O}(\mathbf{x}_d^j)$, where $\mathbf{x}_d^j = \{\mathbf{y}_d^j, \mathbf{g}_d^j, \mathbf{w}_d^j\}$ is the realization of $\mathbf{X} = \{\mathbf{Y}, \mathbf{G}, \mathbf{W}\}$. The training dataset is then composed of a small number

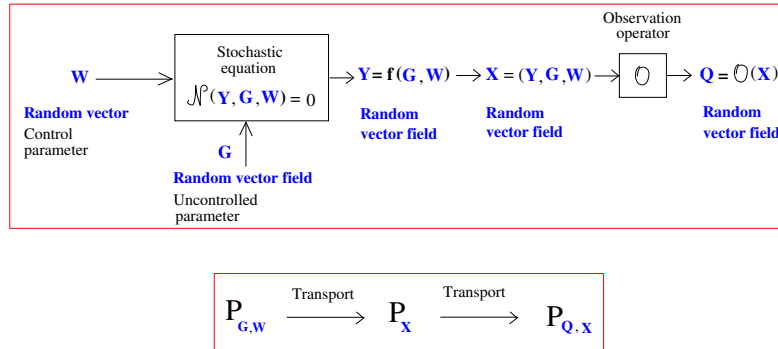


Fig. 1 Diagram illustrating the parameterization of the stochastic boundary value problem and the quantity of interest \mathbf{Q} . The prior probability model is defined by the joint probability measure $P_{\mathbf{G}, \mathbf{W}}$, which is transported into the measure $P_{\mathbf{X}}$ and further transported into the joint probability measure $P_{\mathbf{Q}, \mathbf{X}}$.

N_d of points \mathbf{x}_d^j for $j = 1, \dots, N_d$, which represent N_d independent realizations of \mathbf{X} . It is assumed that the BVP can only be solved a limited number of times. This implies that the training dataset is a small dataset (in contrast to a big dataset). Therefore, the posterior model is constructed using a learning tool to generate the constrained learned realizations of \mathbf{X} without solving the BVP, but solely using the training dataset.

(iii) *Finite reduced-order representation and training dataset $\mathcal{D}_{N_d}(\boldsymbol{\eta})$.* Let us assume that $\mathbf{X} = \{\mathbf{Y}, \mathbf{G}, \mathbf{W}\}$ is a second-order random quantity (random fields and random variable) defined on $(\Theta, \mathcal{F}, \mathcal{P})$, and that its covariance operator is a Hilbert-Schmidt [6], symmetric, positive operator in a Hilbert space \mathbb{X} equipped with the inner product $\langle \mathbf{x}, \mathbf{x}' \rangle_{\mathbb{X}}$. Therefore, \mathbf{X} can be approximated by the truncated Karhunen-Loève expansion $\mathbf{X}^{(\nu)}$ [15, 17] of \mathbf{X} ,

$$\mathbf{X}^{(\nu)} = \underline{\mathbf{x}} + \sum_{\alpha=1}^{\nu} \sqrt{\kappa_{\alpha}} \boldsymbol{\varphi}^{\alpha} H_{\alpha}, \quad (1)$$

where the eigenvalues of the covariance operator are $\kappa_1 \geq \dots \geq \kappa_{\nu} \geq \dots = 0$ with $\sum_{\alpha=1}^{+\infty} \kappa_{\alpha}^2 < +\infty$. The family of the eigenfunctions $\{\boldsymbol{\varphi}^{\alpha}\}_{\alpha}$ is a Hilbert basis of \mathbb{X} , where $\underline{\mathbf{x}} = E\{\mathbf{X}\}$, and where $\mathbf{H} = (H_1, \dots, H_{\nu})$ is a second-order, centered, \mathbb{R}^{ν} -valued random variable whose covariance matrix is the identity matrix $[I_{\nu}]$ in \mathbb{M}_{ν} . For $\alpha \in \{1, \dots, \nu\}$, the component H_{α} is written as $H_{\alpha} = \kappa_{\alpha}^{-1/2} \langle \mathbf{X} - \underline{\mathbf{x}}, \boldsymbol{\varphi}^{\alpha} \rangle_{\mathbb{X}}$. The training set \mathcal{D}_d related to \mathbf{H} consists of the N_d independent realizations $\{\boldsymbol{\eta}_d^j, j = 1, \dots, N_d\}$ such that $\boldsymbol{\eta}_d^j = \kappa_{\alpha}^{-1/2} \langle \mathbf{x}_d^j - \underline{\mathbf{x}}, \boldsymbol{\varphi}^{\alpha} \rangle_{\mathbb{X}}$. We assume that the kernel of the covariance operator is unknown. Therefore, we can only obtain an approximation of the covariance operator using an empirical estimator built with the N_d points $\{\mathbf{x}_d^j, j = 1, \dots, N_d\}$. Under these conditions the largest value of ν will be $N_d - 1$ and the discretization of Eq. (1) will simply correspond to a normalization of the points that constitute the training set \mathcal{D}_d . If ν can be chosen as $N_d - 1$, then $\mathbf{X}^{(\nu)}$ is a reduced-order representation of \mathbf{X} .

(iv) *Probabilistic learning inference under implicit constraints.* We introduce the superscript "c" to designate the solution with the constraints, which corresponds to the posterior model. Let $\mathbf{X}^c = (\mathbf{Y}^c, \mathbf{G}^c, \mathbf{W}^c)$ and $\mathbf{Q}^c = \mathcal{O}(\mathbf{X}^c)$ be the posterior probability model of the prior probability model of $\mathbf{X} = (\mathbf{Y}, \mathbf{G}, \mathbf{W})$ with $\mathbf{Y} = \mathbf{f}(\mathbf{G}, \mathbf{W})$ and $\mathbf{Q} = \mathcal{O}(\mathbf{X})$, respectively. Let $R^c = r(\mathbf{X}^c)$ be the positive-valued random variable representing the norm of the random residue of the PDE. The probabilistic learning inference belongs to the class of the statistical inverse problems. The prior probability model of $\{\mathbf{G}, \mathbf{W}\}$ is given and we are interested in estimating the posterior probability model in order that the second-order moment $E\{(R^c)^2\}$ of R^c be as small as possible and that some statistical moments of the posterior observations \mathbf{Q}^c are close to some given targets. Using Eq. (1), the implicit function $\boldsymbol{\eta} \mapsto \mathbf{h}^c(\boldsymbol{\eta})$ from \mathbb{R}^{ν} into \mathbb{R}^{n_c} , which globally defines the constraints, yields the constraints equation,

$$E\{\mathbf{h}^c(\mathbf{H}^c)\} = \mathbf{b}^c, \quad (2)$$

in which $\mathbf{b}^c \in \mathbb{R}^{n_c}$ is the target and \mathbf{h}^c is an implicit mapping.

(v) *Formulation using the Kullback-Leibler divergence minimum principle.* We use the Kullback-Leibler divergence minimum principle [16, 14, 3] for estimating the posterior probability measure $P_{\mathbf{H}^c}(d\boldsymbol{\eta}) = p_{\mathbf{H}^c}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ on \mathbb{R}^ν of the \mathbb{R}^ν -valued random variable $\mathbf{H}^c = (H_1^c, \dots, H_\nu^c)$. This estimation of $P_{\mathbf{H}^c}$ is performed using the prior probability measure $P_{\mathbf{H}}(d\boldsymbol{\eta}) = p_{\mathbf{H}}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ on \mathbb{R}^ν in which $p_{\mathbf{H}}$ is estimated with a modified (see (vi) below) Gaussian KDE method [2] using the points $\{\boldsymbol{\eta}_d^1, \dots, \boldsymbol{\eta}_d^{N_d}\}$ of the training dataset \mathcal{D}_d , and using the constraint defined by Eq. (2). Therefore, the probability density function $p_{\mathbf{H}^c}$ on \mathbb{R}^ν , which satisfies the constraint defined by Eq. (1) and which is closest to $p_{\mathbf{H}}$, is the solution of the optimization problem

$$p_{\mathbf{H}^c} = \arg \min_{p \in \mathcal{C}_{\text{ad},p}} \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta}) \log \left(\frac{p(\boldsymbol{\eta})}{p_{\mathbf{H}}(\boldsymbol{\eta})} \right) d\boldsymbol{\eta}, \quad (3)$$

in which the admissible set $\mathcal{C}_{\text{ad},p}$ is defined by

$$\mathcal{C}_{\text{ad},p} = \left\{ \boldsymbol{\eta} \mapsto p(\boldsymbol{\eta}) : \mathbb{R}^\nu \rightarrow \mathbb{R}^+, \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta}) d\boldsymbol{\eta} = 1, \int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta}) p(\boldsymbol{\eta}) d\boldsymbol{\eta} = \mathbf{b}^c \right\}. \quad (4)$$

(vi) *Lagrange multipliers associated with the constraints and solution of the functional optimization problem.* As explained in Section 1-(iii), the constraints within the admissible set $\mathcal{C}_{\text{ad},p}$ are taken into account by introducing the Lagrange multipliers $\lambda_0 - 1$ with $\lambda_0 \in \mathbb{R}^+$ (associated with the normalization condition of the pdf) and $\boldsymbol{\lambda} \in \mathcal{C}_{\text{ad},\boldsymbol{\lambda}} \subset \mathbb{R}^{n_c}$ (associated with the imposed constraints defined by the implicit function \mathbf{h}^c). It has been proven [26] that $\boldsymbol{\lambda}^{\text{sol}}$ and the pdf $p_{\mathbf{H}^c}$ of \mathbf{H}^c can be constructed as follows. For $1 \leq n_c \leq \nu$, let us assume that $\mathbf{h}^c \in C^1(\mathbb{R}^\nu, \mathbb{R}^{n_c})$ and there exist constants $\alpha > 0$, $\beta > 0$, $c_\alpha > 0$, and $c_\beta > 0$, independent of $\boldsymbol{\eta}$, such that for $\|\boldsymbol{\eta}\| \rightarrow +\infty$, $\|\mathbf{h}^c(\boldsymbol{\eta})\| \leq c_\alpha \|\boldsymbol{\eta}\|^\alpha$ and $\|[\nabla_{\boldsymbol{\eta}} \mathbf{h}^c(\boldsymbol{\eta})]\|_F \leq c_\beta \|\boldsymbol{\eta}\|^\beta$ in which $[\nabla_{\boldsymbol{\eta}} \mathbf{h}^c(\boldsymbol{\eta})] \in \mathbb{M}_{\nu, n_c}$ with $[\nabla_{\boldsymbol{\eta}} \mathbf{h}^c(\boldsymbol{\eta})]_{\alpha k} = \partial h_k^c(\boldsymbol{\eta}) / \partial \eta_\alpha$, and where $\|\cdot\|_F$ is the Frobenius norm. Then, the admissible set $\mathcal{C}_{\text{ad},\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ is a convex open subset of \mathbb{R}^{n_c} . For all $\boldsymbol{\lambda}$ in $\mathcal{C}_{\text{ad},\boldsymbol{\lambda}}$, the pdf $p_{\mathbf{H}_\lambda}$ of \mathbf{H}_λ can be written as

$$p_{\mathbf{H}_\lambda}(\boldsymbol{\eta}; \boldsymbol{\lambda}) = c_0(\boldsymbol{\lambda}) \zeta(\boldsymbol{\eta}) \exp\{-\langle \boldsymbol{\lambda}, \mathbf{h}^c(\boldsymbol{\eta}) \rangle\}, \quad (5)$$

with $0 < c_0(\boldsymbol{\lambda}) < +\infty$ and where $\boldsymbol{\eta} \mapsto \zeta(\boldsymbol{\eta}) : \mathbb{R}^\nu \rightarrow \mathbb{R}^+$ is written as $\zeta(\boldsymbol{\eta}) = N_d^{-1} \sum_{j=1}^{N_d} \exp\{-\frac{1}{2\hat{s}^2} \|\frac{\hat{s}}{s} \boldsymbol{\eta}_d^j - \boldsymbol{\eta}\|^2\}$ where $\hat{s} = s (s^2 + (N_d - 1)/N_d)^{-1/2}$ with $s = (4/(N_d(2 + \nu)))^{1/(\nu+4)}$. The \mathbb{R}^{n_c} -valued random variable $\mathbf{h}^c(\mathbf{H}_\lambda)$ is such that $E\{\|\mathbf{h}^c(\mathbf{H}_\lambda)\|^2\} < +\infty$. Let $\boldsymbol{\lambda} \mapsto \Gamma(\boldsymbol{\lambda}) : \mathcal{C}_{\text{ad},\boldsymbol{\lambda}} \rightarrow \mathbb{R}$ be defined by $\Gamma(\boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \mathbf{b}^c \rangle - \log c_0(\boldsymbol{\lambda})$ in which $\mathbf{b}^c \in \mathbb{R}^{n_c}$ is defined in Eq. (2). For all $\boldsymbol{\lambda}$ in $\mathcal{C}_{\text{ad},\boldsymbol{\lambda}}$, we have

$$\nabla_{\boldsymbol{\lambda}} \Gamma(\boldsymbol{\lambda}) = \mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H}_\lambda)\} \in \mathbb{R}^{n_c}, \quad [\Gamma''(\boldsymbol{\lambda})] = [\text{cov}\{\mathbf{h}^c(\mathbf{H}_\lambda)\}] \in \mathbb{M}_{n_c}^+, \quad (6)$$

where $[\Gamma''(\boldsymbol{\lambda})]$ is the positive-definite covariance matrix of $\mathbf{h}^c(\mathbf{H}_\lambda)$, which is such that $[\Gamma''(\boldsymbol{\lambda})]_{kk'} = \partial^2 \Gamma(\boldsymbol{\lambda}) / \partial \lambda_k \partial \lambda_{k'}$. Function Γ is a strictly convex function on

$\mathcal{C}_{\text{ad},\lambda}$. There is a unique solution λ^{sol} in $\mathcal{C}_{\text{ad},\lambda}$ of the convex optimization problem,

$$\lambda^{\text{sol}} = \arg \min_{\lambda \in \mathcal{C}_{\text{ad},\lambda}} \Gamma(\lambda). \quad (7)$$

If the following equation in λ ,

$$\nabla_{\lambda} \Gamma(\lambda) = \mathbf{0}_{n_c}, \quad (8)$$

has a solution $\tilde{\lambda}$ in $\mathcal{C}_{\text{ad},\lambda}$, then this solution is unique and we have $\lambda^{\text{sol}} = \tilde{\lambda}$. The pdf $p_{\mathbf{H}^c}$ of \mathbf{H}^c , which satisfies the constraint $E\{\mathbf{h}^c(\mathbf{H}^c)\} = \mathbf{b}^c$ is written as

$$p_{\mathbf{H}^c}(\boldsymbol{\eta}) = p_{\mathbf{H}_{\lambda^{\text{sol}}}}(\boldsymbol{\eta}; \lambda^{\text{sol}}) \quad , \quad \forall \boldsymbol{\eta} \in \mathbb{R}^{\nu}. \quad (9)$$

Constant $c_0(\lambda)$ cannot be computed when ν is large. As a result, the standard conjugate gradient algorithm cannot be used to solve the convex optimization problem defined by Eq. (7). Instead, λ^{sol} is obtained by solving Eq. (8) using a Newton iterative method, which necessitates estimating the mathematical expectations in Eq. (6). To accomplish this, a learned dataset $\mathcal{D}_{\text{learn}}(\boldsymbol{\eta}_{\lambda}) = \boldsymbol{\eta}_{\lambda}^1, \dots, \boldsymbol{\eta}_{\lambda}^N$ is computed using the PLoM algorithm (refer to Section 1-(iii)) for any fixed value of λ in $\mathcal{C}_{\text{ad},\lambda}$ and for which the invariant measure admits the marginal probability measure $P_{\mathbf{H}_{\lambda}}(d\boldsymbol{\eta}; \lambda) = p_{\mathbf{H}_{\lambda}}(\boldsymbol{\eta}; \lambda) d\boldsymbol{\eta}$, defined by Eq. (5). This learned dataset comprises a large number $N \gg N_d$ of realizations of \mathbf{H}_{λ} . Since the implicit function \mathbf{h}^c and its gradient must be evaluated a large number of times by the MCMC algorithm used by PLoM, an algebraic surrogate model of \mathbf{h}^c must be constructed.

(vii) *Statistical surrogate model of the implicit function \mathbf{h}^c and convergence analysis.* Let $\mathbf{A}_{\lambda} = \mathbf{h}^c(\mathbf{H}_{\lambda})$ be the \mathbb{R}^{n_c} -valued random variable whose N independent realizations $\mathbf{a}_{\lambda}^1, \dots, \mathbf{a}_{\lambda}^N$ are such that $\mathbf{a}_{\lambda}^{\ell} = \mathbf{h}^c(\boldsymbol{\eta}_{\lambda}^{\ell}) \in \mathbb{R}^{n_c}$ for $\ell = 1, \dots, N$. The surrogate model $\boldsymbol{\eta} \mapsto \mathbf{h}^N(\boldsymbol{\eta}; \lambda) : \mathbb{R}^{\nu} \rightarrow \mathbb{R}^{n_c}$ of \mathbf{h}^c is defined [26], for all $\boldsymbol{\eta}$ in \mathbb{R}^{ν} , by

$$\mathbf{h}^N(\boldsymbol{\eta}; \lambda) = \sum_{\ell=1}^N \mathbf{a}_{\lambda}^{\ell} \frac{\beta_{\boldsymbol{\eta}}^N(\boldsymbol{\eta}_{\lambda}^{\ell})}{\sum_{\ell'=1}^N \beta_{\boldsymbol{\eta}}^N(\boldsymbol{\eta}_{\lambda}^{\ell'})}, \quad (10)$$

in which for all $\boldsymbol{\eta}$ and $\tilde{\boldsymbol{\eta}}$ in \mathbb{R}^{ν} ,

$$\beta_{\boldsymbol{\eta}}^N(\tilde{\boldsymbol{\eta}}) = \exp\left\{-\frac{1}{2s_{\text{SB}}^2} \|\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}\|_H^2\right\} \quad , \quad \|\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}\|_H^2 = \langle [\sigma_{\mathbf{H}_{\lambda}}]^{-2}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}), \tilde{\boldsymbol{\eta}} - \boldsymbol{\eta} \rangle, \quad (11)$$

in which $[\sigma_{\mathbf{H}_{\lambda}}]$ is the diagonal positive-definite matrix in $\mathbb{M}_{\nu}^{\dagger}$ such that $[\sigma_{\mathbf{H}_{\lambda}}]_{\alpha\alpha}$ is the standard deviation of the real-valued random variable $H_{\lambda,\alpha}$, estimated using $\mathcal{D}_{\text{learn}}(\boldsymbol{\eta}_{\lambda})$, and where s_{SB} is the Silverman bandwidth $s_{\text{SB}} = (4/\{N(2 + n_c + \nu)\})^{1/(n_c + \nu + 4)}$. Propositions 2 to 4 of [26] demonstrate the convergence of the sequence of MCMC generator of PLoM using this statistical surrogate model.

3 Application to 3D stochastic homogenization of heterogeneous material with random spectrum and without scale separation

The details of this application and additional results can be found in [26]. The nonseparation of the mesoscale with the macroscale means that the macroscale is another mesoscale at larger scale with random effective/apparent elastic properties. The posterior probability distribution is constructed using the general methodology presented in Section 2, where the constraints are defined as follows:

- a target set made up of given "experimental" statistical moments of the random effective/apparent elasticity tensor \mathbb{C}^{eff} .
- the second-order moment of the random normalized residue of the random equations of the stochastic computational model.

These constraints ensure that the probabilistic learning algorithm aims to bring the statistical moments closer to their target values while maintaining a small residue of the random equations of the stochastic computational model.

In this application, the primary objective is to identify, in cases where the mesoscopic and macroscopic scales are not separable, the hyperparameters of a prior probabilistic model for the random apparent elasticity tensor field \mathbb{C} at the mesoscopic scale. This identification is achieved through the solution of a statistical inverse problem using "experimental data", which are available at the macroscopic scale. These experimental data concern the statistical moments of the effective/apparent elasticity tensor \mathbb{C}^{eff} at the macroscopic scale. The only constraints imposed on the construction of the mesoscopic prior probabilistic model for \mathbb{C} are twofold: first, it must consider all mathematical and physical properties of \mathbb{C} at the mesoscopic scale, including symmetries and positivity; second, the parameterization of \mathbb{C} must allow the generation of a sufficiently large (random) family for \mathbb{C}^{eff} through stochastic homogenization, encompassing the experimental effective/apparent elasticity tensor \mathbb{C}^{exp} at the macroscopic scale. These conditions enable the creation of a predictive statistical model based on experimental data. Such a probabilistic model, based on experimentation, is particularly valuable. Imagine conducting experimental tests on a specimen of a heterogeneous material subjected to deterministic imposed forces/displacements, where deformations are measured (e.g., a concrete specimen). If the specimen is smaller than a representative elementary volume (RVE), the measured macroscopic deformations are random, rendering the effective tensor \mathbb{C}^{eff} not deterministic but random (apparent tensor). Once the probabilistic model for \mathbb{C} has been experimentally identified, the updated meso-macro probabilistic model can be used to simulate the deterministic effective tensor through numerical simulations on an RVE.

(i) *Stochastic elliptic boundary value problem at mesoscale.* We consider the stochastic homogenization of a heterogeneous linear elastic microstructure, which is modeled at the mesoscale rather than the microscale, and it occupies the 3D bounded open domain $\Omega =]0, 1[\times]0, 1[\times]0, 0.1[\subset \mathbb{R}^3$ (square thick plate) with boundary $\partial\Omega$. For all m and r in $\{1, 2, 3\}$ the \mathbb{R}^3 -valued displacement random field $\{\mathbf{Y}^{mr}(\boldsymbol{\xi}) \in \Omega\}$

at mesoscale is defined on $(\Theta, \mathcal{T}, \mathcal{P})$, is indexed by Ω , and satisfies the stochastic boundary value problem associated with the stochastic homogenization of a random elastic medium at mesoscale without scale separation, such that for $i = 1, 2, 3$, and almost surely,

$$-\frac{\partial}{\partial \xi_j} (\mathbb{C}_{ijpq}(\xi) \varepsilon_{pq}(\mathbf{Y}^{mr}(\xi))) = \mathbf{0}_3 \quad , \quad \forall \xi \in \Omega, \quad (12)$$

$$\mathbf{Y}^{mr}(\xi) = \mathbf{y}_0^{mr}(\xi), \quad \forall \xi \in \partial\Omega, \quad (13)$$

in which the strain tensor is $\varepsilon_{pq}(\mathbf{y}) = (\partial y_p / \partial \xi_q + \partial y_q / \partial \xi_p) / 2$ for all differentiable function $\xi \mapsto \mathbf{y}(\xi)$ and where the components of $\mathbf{y}_0^{mr}(\xi)$ are defined by $y_{0,j}^{mr}(\xi) = (\delta_{jm} \xi_r + \delta_{jr} \xi_m) / 2$, in which δ_{jm} is the Kronecker symbol. At mesoscale, the linear elastic heterogeneous medium is described by the random apparent elasticity field $\{\mathbb{C}(\xi), \xi \in \mathbb{R}^3\}$, which is a non-Gaussian fourth-order tensor-valued random field $\mathbb{C} = \{\mathbb{C}_{ijpq}\}_{ijpq}$ with i, j, p , and q in $\{1, 2, 3\}$, defined on $(\Theta, \mathcal{T}, \mathcal{P})$. The stochastic homogenization consists, for i, j, m , and r in $\{1, 2, 3\}$, in analyzing at macroscale the component $\mathbb{C}_{ijmr}^{\text{eff}}$ of the random effective/apparent elasticity tensor $\{\mathbb{C}_{ijmr}^{\text{eff}}\}_{ijmr}$, which is defined by

$$\mathbb{C}_{ijmr}^{\text{eff}} = \frac{1}{|\Omega|} \int_{\Omega} \mathbb{C}_{ijpq}(\xi) \varepsilon_{pq}(\mathbf{Y}^{mr}(\xi)) d\xi, \quad (14)$$

in which \mathbf{Y}^{mr} is the \mathbb{R}^3 -valued random field that satisfies Eqs. (12) to (13) and where $|\Omega| = \int_{\Omega} d\xi$. The random effective/apparent elasticity tensor \mathbb{C}^{eff} is symmetric and positive definite almost surely. If there was a scale separation, then the statistical fluctuations of this tensor would be negligible.

(ii) *Prior probability model of the apparent elasticity field \mathbb{C} at mesoscale.* It is essential to note that the prior probability model of the apparent elastic field is not carried out at the microscopic scale of the microstructure but rather at the mesoscopic scale. This means that no properties specific to any symmetry class of the elastic field at the microscale are utilized. We refer the reader to [9, 10] for the implementation of symmetry classes in mesoscopic random models of the random apparent elasticity field and to [12, 18] for microscopic random models of the microstructure when it can be described in terms of constituents. Regarding the construction of the mesoscopic prior probabilistic model for the random apparent elasticity field $\mathbb{C} = \{\mathbb{C}(\xi), \xi \in \Omega\}$, it is essential to note that describing only the correlation structure of \mathbb{C} is insufficient. This is because \mathbb{C} is non-Gaussian due to its positivity. To solve the nonlinear statistical inverse problem, it is necessary to build the probability measure of \mathbb{C} and its associated random generator. Such a construction has been proposed in [23, 24] for cases involving anisotropic statistical fluctuations around the mean model at the mesoscale. Its extension to the case involving a random spectral measure, as applied here, is detailed in [25]. The mesoscopic prior probability model of \mathbb{C} is defined as the restriction to Ω of a non-Gaussian, homogeneous, mean-square continuous, fourth-order tensor-valued random field indexed by \mathbb{R}^3 . The mesoscopic

mean model, also known as the mesoscopic nominal model, is isotropic and characterized by $\underline{C}_{\text{bulk}} = 1.09 \times 10^{11} \text{ N/m}^2$ and $\underline{C}_{\text{shear}} = 6.85 \times 10^{10} \text{ N/m}^2$. The mesoscopic statistical fluctuations around the mean mesoscopic model are anisotropic. The parameterization of \mathbb{C} is expressed as $\mathbb{C}(\boldsymbol{\xi}) = \mathfrak{c}(\mathbf{G}(\boldsymbol{\xi}), \underline{\mathbf{z}})$ for $\boldsymbol{\xi} \in \Omega$, where \mathfrak{c} is a tensor-valued function on $\mathbb{R}^{21} \times \mathbb{R}^3$, and $\mathbf{G} = \{\mathbf{G}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}$ is a non-Gaussian, second-order, homogeneous, mean-square continuous, \mathbb{R}^{21} -valued random field, indexed by \mathbb{R}^3 . Here, $\underline{\mathbf{z}} = (\underline{C}_{\text{bulk}}, \underline{C}_{\text{shear}}, \underline{\delta}_{\mathbb{C}})$. The non-Gaussian probability measure of \mathbb{G} is fully determined through statistical conditioning of a Gaussian \mathbb{R}^{21} -valued random field. The mapping \mathfrak{c} is constructed using the maximum entropy principle under constraints, ensuring that the properties of symmetry and positivity of the non-Gaussian random tensor $\mathbb{C}(\boldsymbol{\xi})$ are satisfied at any point $\boldsymbol{\xi}$. The correlation structure of the non-Gaussian field \mathbb{C} is determined by the mapping \mathfrak{c} from the correlation structure of \mathbb{G} . The random matrix spectral measure of \mathbb{G} is defined in [25] based on a prior algebraic model with several hyperparameters: (1) the level of uncertainties on the spectral measure that is controlled by a dispersion parameter $\delta_s = 0.1$, (2) the level of mesoscopic statistical fluctuations of the random medium that is governed by the dispersion coefficient $\underline{\delta}_{\mathbb{C}} = 0.3$, (3) and finally, three spatial correlations lengths denoted as $\underline{\mathbf{L}}_{\mathbb{C}} = (\underline{L}_{c1}, \underline{L}_{c2}, \underline{L}_{c3})$. It is worth noting that the proposed model allows for the introduction of a total of 3×21 spatial correlation lengths, representing the most general parameterization.

(iii) *Spatial correlation lengths and scale separation.* Cases, SC1, SC2, and SC3 of the correlation lengths are considered for analyzing the level of scale separation:

- SC1: $\underline{\mathbf{L}}_{\mathbb{C}} = (0.1, 0.1, 0.1)$ (partial separation in ξ_1, ξ_2 , not in ξ_3).
- SC2: $\underline{\mathbf{L}}_{\mathbb{C}} = (0.3, 0.3, 0.1)$ (not separated in ξ_1, ξ_2 , and ξ_3).
- SC3: $\underline{\mathbf{L}}_{\mathbb{C}} = (0.5, 0.5, 0.2)$ (strongly not separated in ξ_1, ξ_2 , and ξ_3).

(iv) *Random control parameter for the probabilistic learning inference.* The components of the \mathbb{R}^3 -valued random control parameter \mathbf{W} are $(\log C_{\text{bulk}}, \log C_{\text{shear}}, \log \delta_{\mathbb{C}})$ in which C_{bulk} and C_{shear} are Gamma independent random variables (see [8]) with mean values $\underline{C}_{\text{bulk}}, \underline{C}_{\text{shear}}$ and coefficients of variation $\delta_{\text{bulk}} = 0.5, \delta_{\text{shear}} = 0.25$, and where $\delta_{\mathbb{C}}$ is a uniform random variable on $[0.1, 0.5]$.

(v) *Stochastic computational model.* The finite element method is used for discretizing the weak formulation of the stochastic BVP and yields the stochastic equations on \mathbb{R}^{n_y} where $n_y = 52\,215$. The discretization of the vector-valued random field \mathbf{G} yields a \mathbb{R}^{n_g} -valued random variable where $n_g = 3\,626\,800$. The \mathbb{R}^{n_w} -valued random control parameter, with $n_w = 3$.

(vi) *Training dataset and finite reduced-order representation.* The mesoscopic prior probability model and the stochastic computational model are used to construct $N_d = 50$ realizations $\{\mathbf{x}_d^j, j = 1, \dots, N_d\}$ of the random vector \mathbf{X} with values in \mathbb{R}^{n_x} where $n_x = 6n_y + n_g + n_w = 3\,942\,093$. The dimension ν of the reduced representation $\mathbf{X}^{(\nu)}$ of \mathbf{X} constructed by PCA is $\nu = N_d - 1 = 49$.

(vii) *Random normalized residue.* For λ fixed in $\mathcal{C}_{\text{ad},\lambda}$, the dimensionless positive-valued random variable R_λ^c represents the norm of the vector-valued random residue of the stochastic equations. A realization ℓ is denoted as $\rho_\lambda^\ell = \hat{\rho}_\lambda^\ell / \hat{\rho}_0$, where $\hat{\rho}_0$ is an adapted constant for normalization (refer to [26]). This ensures that the constraint equation for the random residue can be expressed as $E\{(R_\lambda^c)^2\} = b_R^c$, with $b_R^c = 1$.

(viii) *Statistical moments and their targets.* Using the Voigt notation, let $[\mathbb{C}^{\text{eff}}]$ be the second-order \mathbb{M}_6^+ -valued random variable associated with the random tensor \mathbb{C}^{eff} . The first statistical moment of interest is the mean value $[\underline{\mathbb{C}}^{\text{eff}}] = E\{[\mathbb{C}^{\text{eff}}]\} \in \mathbb{M}_6^+$ of random matrix $[\mathbb{C}^{\text{eff}}]$ while its target counterpart is the given matrix $[\underline{\mathbb{C}}^{\text{exp}}] \in \mathbb{M}_6^+$. The corresponding constraint equation, which allows the mean value to be fitted, will then be written as $E\{[\mathbb{C}^{\text{eff}}]\} = [\underline{\mathbb{C}}^{\text{exp}}]$. The second statistical moment of interest is the coefficient of dispersion δ^{eff} of random matrix $[\mathbb{C}^{\text{eff}}]$ and its target counterpart δ^{exp} , which allows the level of statistical fluctuations to be fitted. We then introduce the positive-valued random variable $\Delta_2^{\text{eff}} = \|\mathbb{C}^{\text{eff}} - [\underline{\mathbb{C}}^{\text{eff}}]\|_F^2 / \|\underline{\mathbb{C}}^{\text{eff}}\|_F^2$. Let δ^{eff} be defined by $\delta^{\text{eff}} = (E\{\Delta_2^{\text{eff}}\})^{1/2}$. The constraint equation to control the statistical fluctuations are then written as $\delta^{\text{eff}} = \delta^{\text{exp}}$. It should be noted that, if δ^{eff} goes to zero, then the statistical fluctuations (represented by $(\Delta_2^{\text{eff}})^{1/2}$) of $[\mathbb{C}^{\text{eff}}]$ around $[\underline{\mathbb{C}}^{\text{eff}}]$ goes to zero.

(ix) *Second-order statistics of the random residue and of the effective/apparent elasticity matrix estimated with the constrained learned dataset.* For cases SC1, SC2, and SC3, Table 1 gives the posterior statistics computed with the constrained learned set for $N = 10\,000$ (subscript "c"), the prior statistics computed with the training set (subscript "d"), and the targets (superscript "exp"). It can be read,

- (a) the second-order moment $E\{R_c^2\}$ of the random normalized residue R_c and the corresponding target b_R^c .
- (b) the Frobenius norm $\|\underline{\mathbb{C}}^{\text{eff}}\|_F$ of the mean value of the random effective/apparent elasticity matrix.
- (c) the coefficient of dispersion $\delta_{\text{ML}}^{\text{eff}} = \{\max_{\delta_2} p_{\Delta_2^{\text{eff}}}(\delta_2)\}^{1/2}$ in which $p_{\Delta_2^{\text{eff}}}$ is the pdf of random variable Δ_2^{eff} .

(x) *Discussion about the presented results.* Table 1 demonstrates that the constrained learned dataset significantly modifies the mesoscopic prior probability model. The posterior statistics align well with the targets. These results confirm controlled residue and small deviations from the reference (training) for the optimal solution. The mean value of the random effective/apparent elasticity matrix achieves the target. When there is no scale separation, there is a significant coefficient of dispersion, which increases as the spatial correlation lengths of the random apparent elasticity field at the mesoscale grow. There are statistical fluctuations in the effective/apparent elasticity tensor at the macroscale. Table 1 further demonstrates that the constraints in place ensure that the probabilistic learning algorithm strives to align the statistical moments with their target values, while simultaneously maintaining a negligible residue of the random equations within the stochastic computational model.

Table 1 Posterior statistics with the constrained learned set for $N = 10000$ (subscript "c"), prior statistics with the training dataset (subscript "d"), and targets (superscript "exp").

	SC1	SC2	SC3
$E\{R_c^2\}$	1.2938	1.2687	1.2413
b_R^c	1	1	1
$\ [\underline{C}_d^{\text{eff}}] \ _F \times 10^{11}$	4.2106	4.1925	4.1943
$\ [\underline{C}_c^{\text{eff}}] \ _F \times 10^{11}$	4.6294	4.6923	4.6816
$\ [\underline{C}^{\text{exp}}] \ _F \times 10^{11}$	4.6317	4.6549	4.6706
$\delta_{d,\text{ML}}^{\text{eff}}$	0.2257	0.2469	0.2701
$\delta_{c,\text{ML}}^{\text{eff}}$	0.1329	0.1476	0.1671
δ^{exp}	0.0946	0.1374	0.1825

References

- Allaire, G.: Homogenization and two-scale convergence. *SIAM Journal on Mathematical Analysis* **23**(6), 1482–1518 (1992). DOI 10.1137/0523084
- Bowman, A., Azzalini, A.: *Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-Plus Illustrations*, vol. 18. Oxford University Press, Oxford: Clarendon Press, New York (1997). DOI 10.1007/s001800000033
- Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. Second Edition, John Wiley & Sons, Hoboken (2006)
- Dirrenberger, J., Forest, S., Jeulin, D.: Towards gigantic rve sizes for 3d stochastic fibrous networks. *International Journal of Solids and Structures* **51**(2), 359–376 (2014). DOI 10.1016/j.ijsolstr.2013.10.011
- Forest, S.: Homogenization methods and mechanics of generalized continua-part 2. *Theoretical and applied mechanics* **28-29**, 113–143 (2002)
- Gelfand, I.M., Vilenkin, N.I.: *Les Distributions*. Tome 4. Application de l'Analyse Harmonique. Dunod (1967). DOI 10.1016/0375-9474(67)90547-7
- Ghanem, R., Spanos, P.D.: *Stochastic Finite Elements: a Spectral Approach*. Springer-Verlag, New York (1991)
- Guilleminot, J., Soize, C.: On the statistical dependence for the components of random elasticity tensors exhibiting material symmetry properties. *Journal of Elasticity* **111**(2), 109–130 (2013). DOI 10.1007/s10659-012-9396-z
- Guilleminot, J., Soize, C.: Stochastic model and generator for random fields with symmetry properties: application to the mesoscopic modeling of elastic random media. *Multiscale Modeling & Simulation (A SIAM Interdisciplinary Journal)* **11**(3), 840–870 (2013). DOI 10.1137/120898346
- Guilleminot, J., Soize, C.: Non-Gaussian random fields in multiscale mechanics of heterogeneous materials. In: Altenbach H., Ochsner A. (eds) *Encyclopedia of Continuum Mechanics* pp. 1826–1834 (2020). DOI 10.1007/978-3-662-55771-6_68
- Jeulin, D.: Random texture models for material structures. *Statistics and Computing* **10**(2), 121–132 (2000). DOI 10.1023/A:1008942325749
- Jeulin, D.: *Morphological Models of Random Structures*. Springer Cham (2022). DOI 10.1007/978-3-030-75452-5
- Jeulin, D., Ostojic-Starzewski, M.: *Mechanics of random and multiscale microstructures*. Springer (2001)
- Kapur, J.N., Kesavan, H.K.: *Entropy Optimization Principles with Applications*. Academic Press, San Diego (1992)
- Karhunen, K.: Ueber lineare methoden in der wahrscheinlichkeitsrechnung. *Amer. Acad. Sci., Fennicade, Ser. A, I* **37**, 3–79 (1947)

16. Kullback, S., Leibler, R.A.: On information and sufficiency. *The Annals of Mathematical Statistics* **22**(1), 79–86 (1951). DOI 10.1214/aoms/1177729694
17. Lœve, M.: Fonctions aleatoires du second ordre. In: P. Lévy (ed.) *Processus Stochastiques et Mouvement Brownien*, pp. 366–420. Gauthier-Villars, Paris (1948)
18. Malyarenko, A., Ostoja-Starzewski, M.: *Tensor-Valued Random Fields for Continuum Physics*. Cambridge University Press (2018)
19. Ostoja-Starzewski, M.: Material spatial randomness: From statistical to representative volume element. *Probabilistic Engineering Mechanics* **21**(2), 112–132 (2006). DOI 10.1016/j.probenmech.2005.07.007
20. Owhadi, H., Scovel, C., Sullivan, T.: On the brittleness of Bayesian inference. *SIAM Review* **57**(4), 566–582 (2015). DOI 10.1137/130938633
21. Sab, K., Nedjar, B.: Periodization of random media and representative volume element size for linear composites. *Comptes Rendus Mécanique* **333**(2), 187–195 (2005). DOI 10.1016/j.crme.2004.10.003
22. Schmidhuber, J.: Deep learning in neural networks: An overview. *Neural networks* **61**, 85–117 (2015). DOI 10.1016/j.neunet.2014.09.003
23. Soize, C.: Non Gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators. *Computer Methods in Applied Mechanics and Engineering* **195**(1-3), 26–64 (2006). DOI 10.1016/j.cma.2004.12.014
24. Soize, C.: Tensor-valued random fields for meso-scale stochastic model of anisotropic elastic microstructure and probabilistic analysis of representative volume element size. *Probabilistic Engineering Mechanics* **23**(2-3), 307–323 (2008). DOI 10.1016/j.probenmech.2007.12.019
25. Soize, C.: Stochastic elliptic operators defined by non-Gaussian random fields with uncertain spectrum. *The American Mathematical Society Journal Theory of Probability and Mathematical Statistics* **105**, 113–136 (2021). DOI 10.1090/tpms/1159
26. Soize, C.: Probabilistic learning inference of boundary value problem with uncertainties based on Kullback-Leibler divergence under implicit constraints. *Computer Methods in Applied Mechanics and Engineering* **395**, 115078 (2022). DOI 10.1016/j.cma.2022.115078
27. Soize, C., Ghanem, R.: Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM Journal on Scientific Computing* **26**(2), 395–410 (2004). DOI 10.1137/S1064827503424505
28. Soize, C., Ghanem, R.: Data-driven probability concentration and sampling on manifold. *Journal of Computational Physics* **321**, 242–258 (2016). DOI 10.1016/j.jcp.2016.05.044
29. Soize, C., Ghanem, R.: Physics-constrained non-Gaussian probabilistic learning on manifolds. *International Journal for Numerical Methods in Engineering* **121**(1), 110–145 (2020). DOI 10.1002/nme.6202
30. Soize, C., Ghanem, R.: Probabilistic learning on manifolds. *Foundations of Data Science* **2**(3), 279–307 (2020). DOI 10.3934/fods.2020013
31. Soize, C., Ghanem, R.: Probabilistic learning on manifolds constrained by nonlinear partial differential equations for small datasets. *arXiv:2010.14324 [stat.ML]* pp. 1–38 (2020). URL <http://arxiv.org/abs/2010.14324>
32. Soize, C., Ghanem, R.: Probabilistic learning on manifolds constrained by nonlinear partial differential equations for small datasets. *Computer Methods in Applied Mechanics and Engineering* **380**, 113777 (2021). DOI 10.1016/j.cma.2021.113777
33. Soize, C., Ghanem, R.: Probabilistic learning on manifolds (PLoM) with partition. *International Journal for Numerical Methods in Engineering* **123**(1), 268–290 (2022). DOI 10.1002/nme.6856
34. Tipireddy, R., Ghanem, R.: Basis adaptation in homogeneous chaos spaces. *Journal of Computational Physics* **259**, 304–317 (2014). DOI 10.1016/j.jcp.2013.12.009
35. Torquato, S., Stell, G.: Microstructure of two-phase random media. v. the n-point matrix probability functions for impenetrable spheres. *The Journal of Chemical Physics* **82**(2), 980–987 (1985). DOI 10.1063/1.448475