

Physics-constrained non-Gaussian probabilistic learning on manifolds

Christian Soize¹  | Roger Ghanem² 

¹Laboratoire Modélisation et Simulation Multi Echelle, MSME UMR 8208 CNRS, Université Paris-Est Marne-la-Vallée, Marne-la-Vallée, France

²Viterbi School of Engineering, University of Southern California, Los Angeles, California

Correspondence

Christian Soize, Laboratoire Modélisation et Simulation Multi Echelle, MSME UMR 8208 CNRS, Université Paris-Est Marne-la-Vallée, 5 bd Descartes, 77454 Marne-la-Vallée, France.
Email: christian.soize@u-pem.fr

Funding information

Defense Advanced Research Projects Agency; PIRATE project

Summary

An extension of the probabilistic learning on manifolds (PLoM), recently introduced by the authors, has been presented: In addition to the initial data set given for performing the probabilistic learning, constraints are given, which correspond to statistics of experiments or of physical models. We consider a non-Gaussian random vector whose unknown probability distribution has to satisfy constraints. The method consists in constructing a generator using the PLoM and the classical Kullback-Leibler minimum cross-entropy principle. The resulting optimization problem is reformulated using Lagrange multipliers associated with the constraints. The optimal solution of the Lagrange multipliers is computed using an efficient iterative algorithm. At each iteration, the Markov chain Monte Carlo algorithm developed for the PLoM is used, consisting in solving an Itô stochastic differential equation that is projected on a diffusion-maps basis. The method and the algorithm are efficient and allow the construction of probabilistic models for high-dimensional problems from small initial data sets and for which an arbitrary number of constraints are specified. The first application is sufficiently simple in order to be easily reproduced. The second one is relative to a stochastic elliptic boundary value problem in high dimension.

KEYWORDS

data driven, Kullback-Leibler, machine learning, probabilistic learning, statistical constraints, uncertainty quantification

1 | INTRODUCTION

The consideration of constraints in learning algorithms remains a very important and active research topic. Bayesian updating^{1–5} provides a rational framework for integrating data into predictive models and has been successfully adapted to situations where likelihoods are not readily available either because of expense or because of the nature of available information.^{6,7} In many instances, however, relevant information is available in the form of sample statistics rather than raw data or samplewise constraints. In these settings, an alternative to the Bayesian framework has evolved over the past decades in the form of constrained maximum entropy⁸ and constrained minimum cross-entropy.^{9–11} The latter, also known as the Kullback-Liebler divergence, forms the basis of the method developed in this paper.

The approach using the Kullback-Liebler divergence has been used extensively over the last three decades for imposing constraints in the framework of learning with statistical models. Many developments and applications have been published such as, for instance, the following interesting papers devoted to the training of neural network classifiers,¹² learning topological,¹³ ensemble learning related to multilayer networks,¹⁴ constraints on parameters during learning,¹⁵ recognition of human activities,¹⁶ text categorization,¹⁷ exploration of data analysis,¹⁸ classification problems

with constrained data,¹⁹ statistical learning algorithms with linguistic constraints,²⁰ reinforcement learning in finite Markov decision processes,²¹ visual recognition,^{22–24} optimal sequential allocation,²⁵ identifiability of parameter learning machines,²⁶ speech enhancement,²⁷ the training of generative neural samplers,²⁸ pessimistic uncertainty quantification problem,²⁹ inference given summary statistics,^{30,31} optimization framework concerning distance metric learning³² unsupervised machine learning techniques to learn latent parameters,³³ graph-based semisupervised learning,³⁴ speech enhancement,³⁵ and finally, to the broad learning systems.³⁶

In spite of these developments, little progress has been made for non-Gaussian probabilistic learning under statistical constraints, for the case of small data sets. This paper is devoted to this framework: learning from a high-dimensional small data set, without invoking the Gaussian assumption. The first ingredient of the proposed approach is the probabilistic learning on manifolds (PLoM) presented in the work of Soize and Ghanem,³⁷ for which complementary developments^{38–40} and applications with validation^{41–43} are presented in other works. This PLoM allows for constructing a *generated data set* D_M^g made up of M additional independent realizations $\{\mathbf{x}_{\text{ar}}^\ell, \ell = 1, \dots, M\}$ of a non-Gaussian random vector \mathbf{X} , defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$, with values in \mathbb{R}^n , for which the probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ is unknown, using only an *initial data set* D_N (training set), which is a small data set, made up of N independent realizations $\{\mathbf{x}^j, j = 1, \dots, N\}$ of \mathbf{X} with $\mathbf{x}^j = \mathbf{X}(\theta_j)$ for $\theta_j \in \Theta$ and such that $M \gg N$.

In this paper, we present an extension of this PLoM for which, not only the initial data set D_N is given, but in addition, constraints are specified, in the form of statistics synthesized from experimental data, from theoretical considerations, or from numerical simulations. The “physics-constrained” terminology used in the title will be explained in Section 4. We thus construct a non-Gaussian random vector \mathbf{X}^c whose probability distribution $P_{\mathbf{X}^c}(d\mathbf{x})$ must satisfy the specified constraints. The method consists of constructing a generator of $P_{\mathbf{X}^c}(d\mathbf{x})$ using the PLoM, for which $P_{\mathbf{X}^c}(d\mathbf{x})$ is closest to $P_{\mathbf{X}}(d\mathbf{x})$, while satisfying the constraints. For that, the classical Kullback-Leibler minimum cross-entropy principle is used, which is the second ingredient of the proposed method. Consequently, an optimization problem is formulated using Lagrange multipliers associated with the constraints. The Lagrange multipliers are evaluated as the solution to a convex optimization problem. An efficient iterative algorithm based on the Newton method is proposed. At each Newton iteration, an Markov chain Monte Carlo (MCMC) algorithm has to be used for estimating the gradient and the Hessian of the objective function. The PLoM, which is detailed in the work of Soize and Ghanem,³⁷ is used for the generation of realizations. It consists in solving an Itô stochastic differential equation (ISDE) that is projected on a diffusion-maps basis. This ISDE corresponds to a nonlinear stochastic dissipative Hamiltonian dynamical system. The method and the algorithm presented are efficient and allow high-dimensional problems to be solved from small initial data sets and for which an arbitrary number of constraints are specified.

This paper is organized as follows. In Section 2, the problem that has to be solved is defined. Section 3 deals with the methodology proposed to take into account the constraints that correspond to statistics coming from experiments or defined by a physical/computational model. Examples of constraints are presented in Section 4. Section 5 is devoted to the computational statistical method for generating realizations of \mathbf{X}^c using the PLoM and involving the iterative Newton algorithm that requires the use of the MCMC generator based on the ISDE projected on a diffusion-maps basis. Finally, two applications are presented. The first one, presented in Section 6, is completely specified and can thus be easily reproduced by the reader. The second one, presented in Section 7, corresponds to a physical system, which is represented by a stochastic computational model that corresponds to the finite element discretization of a stochastic elliptic boundary value problem.

In this paper, any Euclidean space \mathcal{E} (such as \mathbb{R}^n) is equipped with its Borel field $\mathcal{B}_{\mathcal{E}}$, which means that $(\mathcal{E}, \mathcal{B}_{\mathcal{E}})$ is a measurable space on which a probability measure can be defined. The mathematical expectation is denoted by E .

2 | SETTING THE PROBLEM TO BE SOLVED

A typical problem for the use of the approach proposed in this paper is the following. Let $(\mathbf{w}, \mathbf{u}) \mapsto \mathbf{f}(\mathbf{w}, \mathbf{u})$ be any measurable mapping on $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$ with values in \mathbb{R}^{n_q} representing a mathematical/computational model. Let \mathbf{W} and \mathbf{U} be two independent (non-Gaussian) random variables defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$ with values in \mathbb{R}^{n_w} and \mathbb{R}^{n_u} , for which the probability distributions $P_{\mathbf{W}}(d\mathbf{w}) = p_{\mathbf{W}}(\mathbf{w}) d\mathbf{w}$ and $P_{\mathbf{U}}(d\mathbf{u}) = p_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}$ are defined by the probability density functions $p_{\mathbf{W}}$ and $p_{\mathbf{U}}$ with respect to the Lebesgue measures $d\mathbf{w}$ and $d\mathbf{u}$ on \mathbb{R}^{n_w} and \mathbb{R}^{n_u} . The role played by \mathbf{W} and \mathbf{U} are the following. The boundary value problem used for constructing the computational model depends on uncertain parameters (epistemic and/or aleatory uncertainties). These random parameters are, for instance, those used for modeling geometry uncertainties, uncertain boundary conditions, constitutive equations of random media, and uncertain physical

properties, and can also be a finite spatial discretization of random fields (such as the one used for modeling the elasticity field of a heterogeneous material). Random vector \mathbf{W} is made up of a part of these random parameters, which are used for controlling the system, while random vector \mathbb{U} is made up of the other part of these random parameters, which are not used for controlling the system. Let \mathbf{Q} be the vector of the quantities of interest (QoI) that is a random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in \mathbb{R}^{n_q} such that

$$\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbb{U}). \quad (1)$$

Random QoI, \mathbf{Q} , represents the vector of all the observations performed in the system and is expressed as a function of the solution of the computational model. Let us assume that N calculations have been performed with the mathematical/computational model whose solution is represented by Equation (1), allowing N independent realizations $\{\mathbf{q}^j, j = 1, \dots, N\}$ of \mathbf{Q} to be computed such that

$$\mathbf{q}^j = \mathbf{f}(\mathbf{w}^j, \mathbb{u}^j), \quad (2)$$

in which $\{\mathbf{w}^j, j = 1, \dots, N\}$ and $\{\mathbb{u}^j, j = 1, \dots, N\}$ are N independent realizations of (\mathbf{W}, \mathbb{U}) , which have been generated using an adapted generator for $p_{\mathbf{W}}$ and $p_{\mathbb{U}}$. We then consider the random variable \mathbf{X} with values in \mathbb{R}^n , such that

$$\mathbf{X} = (\mathbf{Q}, \mathbf{W}), \quad n = n_q + n_w. \quad (3)$$

The probabilistic learning will be performed for this random vector \mathbf{X} that includes the control parameter \mathbf{W} and the QoI \mathbf{Q} , but that does not include random parameter \mathbb{U} . The initial data set related to random vector \mathbf{X} is then made up of the N independent realizations

$$\{\mathbf{x}^j, j = 1, \dots, N\}, \quad \mathbf{x}^j = (\mathbf{q}^j, \mathbf{w}^j) \in \mathbb{R}^n. \quad (4)$$

It is assumed that the measurable mapping \mathbf{f} is such that the conditional probability distribution $P_{\mathbf{Q}|\mathbf{W}}(d\mathbf{q}|\mathbf{w})$ admits a conditional probability density function $p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w})$ with respect to the Lebesgue measure $d\mathbf{q}$ on \mathbb{R}^{n_q} for $P_{\mathbf{W}}(d\mathbf{w})$, almost all \mathbf{w} in \mathbb{R}^{n_w} (note that the existence of a pdf for \mathbb{U} , $P_{\mathbb{U}}(d\mathbb{u}) = p_{\mathbb{U}}(\mathbb{u})d\mathbb{u}$, is necessary but is not sufficient). Under this hypothesis, it can be deduced that the probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ of \mathbf{X} admits a density $\mathbf{x} \mapsto p_{\mathbf{X}}(\mathbf{x})$ with respect to the Lebesgue measure $d\mathbf{x}$ on \mathbb{R}^n . The proof is the following. Let $P_{\mathbf{Q}|\mathbf{W}}(d\mathbf{q}|\mathbf{w})$ be the conditional probability distribution of the conditional random variable $\mathbf{Q}|\mathbf{W}$ given $\mathbf{W} = \mathbf{w}$ for $\mathbf{w} \in \mathbb{R}^{n_w}$, which is such that $\mathbf{Q}|\mathbf{w} = \mathbf{f}(\mathbf{w}, \mathbb{U})$. Because $P_{\mathbf{W}}(d\mathbf{w}) = p_{\mathbf{W}}(\mathbf{w})d\mathbf{w}$, under the above hypothesis (existence of conditional pdf $\mathbf{q} \mapsto p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w})$), the probability distribution of (\mathbf{Q}, \mathbf{W}) can be written as $P_{\mathbf{Q}, \mathbf{W}}(d\mathbf{q}, d\mathbf{w}) = p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w})p_{\mathbf{W}}(\mathbf{w})d\mathbf{q}d\mathbf{w}$, which shows that the probability distribution of (\mathbf{Q}, \mathbf{W}) admits a pdf with respect to $d\mathbf{q}d\mathbf{w}$ and consequently shows that the probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ of $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ admits a density $p_{\mathbf{X}}(\mathbf{x})$ with respect to the Lebesgue measure $d\mathbf{x}$ on \mathbb{R}^n .

The PLoM allows for generating additional realizations $\{(\mathbf{q}_{\text{ar}}^\ell, \mathbf{w}_{\text{ar}}^\ell), \ell = 1, \dots, M\}$ for $M \gg N$ without using the computational model, but using only the training data set, which we denote by D_N . These additional realizations allow, for instance, a cost function $J(\mathbf{w}) = E\{\mathcal{J}(\mathbf{Q}, \mathbf{W})|\mathbf{W} = \mathbf{w}\}$ to be evaluated, in which $(\mathbf{q}, \mathbf{w}) \mapsto \mathcal{J}(\mathbf{q}, \mathbf{w})$ is a given measurable real-valued mapping on $\mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ as well as constraints related to a nonconvex optimization problem^{39,41,42} and this, without calling the mathematical/computational model.

Sometimes, additional information in the form of statistics may become available, synthesized from partial measurements, published data, or numerical simulations. The goal is then to generate additional realizations by taking into account the constraints defined by these statistics. For instance, statistics can correspond to a specified mean value m_k^{exp} and standard deviation s_k^{exp} of several components Q_k of \mathbf{Q} (see Section 4). Statistics can also correspond to the mean value of a given stochastic equation involving \mathbf{Q} and \mathbf{W} , which models a physical system or a part of it (see Appendix A).

Remark concerning the scaling of the initial data set. In practice, the initial data set can be made up of heterogeneous numerical values and must be scaled for performing computational statistics. In this remark, it is then assumed that the initial data set \mathbb{D}_N is made up of N independent realizations $\{\mathbf{x}^j, j = 1, \dots, N\}$ of the \mathbb{R}^n -valued random variable \mathbf{X} . Let $\mathbf{x}^{\max} = \max_j \{\mathbf{x}^j\}$ and $\mathbf{x}^{\min} = \min_j \{\mathbf{x}^j\}$ be vectors in \mathbb{R}^n . Let $[\alpha_x]$ be the invertible diagonal $(n \times n)$ real matrix such that $[\alpha_x]_{kk} = (\mathbf{x}_k^{\max} - \mathbf{x}_k^{\min})$ if $\mathbf{x}_k^{\max} - \mathbf{x}_k^{\min} \neq 0$, and $[\alpha_x]_{kk} = 1$ if $\mathbf{x}_k^{\max} - \mathbf{x}_k^{\min} = 0$. The scaling of random vector \mathbf{X} is the \mathbb{R}^n -valued random variable \mathbf{X} (that has previously been introduced) such that

$$\mathbf{X} = [\alpha_x]\mathbf{X} + \mathbf{x}^{\min}, \quad \mathbf{X} = [\alpha_x]^{-1}(\mathbf{X} - \mathbf{x}^{\min}), \quad (5)$$

which shows that the N independent realizations $\{\mathbf{x}^j, j = 1, \dots, N\}_j$ of \mathbf{X} are such that

$$\mathbf{x}^j = [\alpha_x]^{-1}(\mathbf{x}^j - \mathbf{x}^{\min}). \quad (6)$$

The scaled initial data set D_N is then defined by

$$D_N = \{\mathbf{x}^j, j = 1, \dots, N\}, \quad (\mathbf{q}^j, \mathbf{w}^j) = \mathbf{x}^j. \quad (7)$$

3 | METHODOLOGY PROPOSED TO TAKE INTO ACCOUNT CONSTRAINTS DEFINED BY EXPERIMENTS OR BY PHYSICAL MODELS

To take into account statistical constraints in the algorithm of the PLoM, the proposed methodology consists of minimizing the probabilistic distance (or the divergence) between the probability distribution $P_{\mathbf{X}}(d\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$ of the \mathbb{R}^n -valued random variable \mathbf{X} estimated in the framework of the PLoM and the probability distribution $P_{\mathbf{X}^c}(d\mathbf{x}) = p_{\mathbf{X}^c}(\mathbf{x}) d\mathbf{x}$ of the \mathbb{R}^n -valued random variable \mathbf{X}^c that, in addition, satisfies the constraints. We thus have to find among all the probability density functions (pdf) $p_{\mathbf{X}^c}$ that satisfy the constraints the one that is closest to the pdf $p_{\mathbf{X}}$. For that, we propose to use the classical Kullback-Leibler minimum cross-entropy principle,⁹⁻¹¹ which is written as

$$p_{\mathbf{X}^c} = \arg \min_{\hat{p} \in C_{ad,\mathbf{X}^c}} \int_{\mathbb{R}^n} \hat{p}(\mathbf{x}) \log \frac{\hat{p}(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})} d\mathbf{x}, \quad (8)$$

in which the admissible set C_{ad,\mathbf{X}^c} is defined by

$$C_{ad,\mathbf{X}^c} = \left\{ \mathbf{x} \mapsto \hat{p}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^+, \int_{\mathbb{R}^n} \hat{p}(\mathbf{x}) d\mathbf{x} = 1, \int_{\mathbb{R}^n} \mathbf{g}^c(\mathbf{x}) \hat{p}(\mathbf{x}) d\mathbf{x} = \beta^c \right\}. \quad (9)$$

In Equation (9), β^c is a given vector in \mathbb{R}^{m_c} with $m_c \geq 1$ and $\mathbf{x} \mapsto \mathbf{g}^c(\mathbf{x})$ is a measurable mapping from \mathbb{R}^n into \mathbb{R}^{m_c} . This means that, in addition to the normalization condition, pdf $p_{\mathbf{X}^c}$ must be such that the following constraint is satisfied:

$$E\{\mathbf{g}^c(\mathbf{X}^c)\} = \int_{\mathbb{R}^n} \mathbf{g}^c(\mathbf{x}) p_{\mathbf{X}^c}(\mathbf{x}) d\mathbf{x} = \beta^c. \quad (10)$$

Remarks

- (1) It should be noted that, in the framework of the PLoM, the estimation $p_{\mathbf{X}}$ of the pdf of \mathbf{X} , which is performed using the Gaussian kernel-density estimation method and initial data set D_N , is such that $p_{\mathbf{X}}(\mathbf{x}) > 0$ for all \mathbf{x} in \mathbb{R}^n .
- (2) The functional that has to be minimized on C_{ad,\mathbf{X}^c} is the Kullback-Leibler divergence (relative entropy),

$$D(\hat{p}; p_{\mathbf{X}}) = \int_{\mathbb{R}^n} \hat{p}(\mathbf{x}) \log \frac{\hat{p}(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})} d\mathbf{x}, \quad (11)$$

that is, nonnegative, additive, and nonsymmetric ($D(\hat{p}; p_{\mathbf{X}}) \neq D(p_{\mathbf{X}}; \hat{p})$). The symmetric divergence

$$D^s(\hat{p}; p_{\mathbf{X}}) = \int_{\mathbb{R}^n} (\hat{p}(\mathbf{x}) - p_{\mathbf{X}}(\mathbf{x})) \log \frac{\hat{p}(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})} d\mathbf{x}$$

could also be introduced. As explained in the work of Kullback and Leibler,¹⁰ the symmetric divergence has exactly the same mathematical properties as the relative entropy, such as nonnegativeness and additivity properties. The introduction of the symmetry property could have only an interest if pdf $p_{\mathbf{X}}$ was constructed as an *a priori* pdf, which is not the case because $p_{\mathbf{X}}$ is constructed using the nonparametric statistics. In addition, the choice of the relative divergence instead of

the symmetric divergence has allowed for developing simpler and more efficient algorithms for solving the optimization problem defined by Equations (8) and (9). Consequently, we will pursue the formulation defined by Equations (8) and (9).

(3) The objective of this paper is therefore to use the PLoM for the random variable \mathbf{X}^c with values in \mathbb{R}^n for which its probability distribution $P_{\mathbf{X}^c}(d\mathbf{x}) = p_{\mathbf{X}^c}(\mathbf{x})d\mathbf{x}$ is defined by the pdf $p_{\mathbf{X}^c}$ that is unknown but that can possibly be concentrated in the neighborhood of a manifold in \mathbb{R}^n .

(i) Data of the problem are only those defined by

(i-1) initial data set $D_N = \{\mathbf{x}^j, j = 1, \dots, N\}$ of independent realizations of the \mathbb{R}^n -valued random variable \mathbf{X} for which its probability distribution $P_{\mathbf{X}}(d\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$ (defined by a density) is unknown. We consider the small data set case, which means that N is small.

(i-2) a given vector-valued constraint equation on \mathbb{R}^{m_c} (see Equation (10)),

$$\int_{\mathbb{R}^n} \mathbf{g}^c(\mathbf{x}) \hat{p}_{\mathbf{X}^c}(\mathbf{x}) d\mathbf{x} = \boldsymbol{\beta}^c.$$

(ii) The problem, which has to be solved, consists in constructing M realizations $\{\mathbf{x}^{c,\ell}, \ell = 1, \dots, M\}$ of random vector \mathbf{X}^c with $M \gg N$, for which pdf $p_{\mathbf{X}^c}$ satisfies the constraint while being the closest pdf to $p_{\mathbf{X}}$.

4 | EXAMPLE OF CONSTRAINTS

Using the previous notations, random vector \mathbf{X}^c is written as

$$\mathbf{X}^c = (\mathbf{Q}^c, \mathbf{W}^c) \quad \text{with values in } \mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}. \quad (12)$$

In this section, we give two examples of constraints. The first example corresponds to given statistical constraints defined by experimental second-order moments of \mathbf{Q}^c . The second one corresponds to physical constraints defined by the mean value of a stochastic equation that relates \mathbf{Q}^c and \mathbf{W}^c . For illustrations, two other examples are given in Appendix A.

4.1 | Statistical constraints defined by experimental second-order moments of \mathbf{Q}^c

Let $\mathcal{J}_{\text{mom}} = \{k_1, \dots, k_{\kappa_{\text{mom}}}\}$ be $\kappa_{\text{mom}} \leq n_q$ integers such that $\mathcal{J}_{\text{mom}} \subseteq \{1, \dots, n_q\}$. For $\kappa \in \{1, \dots, \kappa_{\text{mom}}\}$, let us assume that the experimental mean value, $\mathbb{m}_\kappa^{\text{exp}}$, and the experimental standard deviation, $\mathbb{s}_\kappa^{\text{exp}}$, of the component $Q_{k_\kappa}^c$ of random vector \mathbf{Q}^c are specified. For instance, these statistics could come from measurements. Note that the measured realizations that have been used for estimating these experimental quantities are not available. On the other hand, $\{(\mathbb{m}_\kappa^{\text{exp}}, \mathbb{s}_\kappa^{\text{exp}}), \kappa = 1, \dots, \kappa_{\text{mom}}\}$ do not correspond to the initial data set D_N . For instance, D_N could come from numerical simulations, which include modeling errors. The experimental second-order moment $\mathbb{r}_\kappa^{\text{exp}} = E\{(X_{k_\kappa}^c)^2\}$ of random variable $X_{k_\kappa}^c$ can be deduced and is written as $\mathbb{r}_\kappa^{\text{exp}} = (\mathbb{s}_\kappa^{\text{exp}})^2 + (\mathbb{m}_\kappa^{\text{exp}})^2$. In this case, $m_c = 2\kappa_{\text{mom}} \leq 2n_q$ and the m_c functions $\mathbf{x} \mapsto \mathbf{g}^c(\mathbf{x}) = (g_1^c(\mathbf{x}), \dots, g_{m_c}^c(\mathbf{x}))$ from \mathbb{R}^n into \mathbb{R}^{m_c} and vector $\boldsymbol{\beta}^c = (\beta_1^c, \dots, \beta_{m_c}^c)$, which define the constraints (see Equation (10)), are such that, for $\kappa = 1, \dots, \kappa_{\text{mom}}$,

$$g_\kappa^c(\mathbf{X}^c) = Q_{k_\kappa}^c, \quad \beta_\kappa^c = \mathbb{m}_\kappa^{\text{exp}}, \quad (13)$$

$$g_{\kappa+\kappa_{\text{mom}}}^c(\mathbf{X}^c) = \left(Q_{k_\kappa}^c\right)^2, \quad \beta_{\kappa+\kappa_{\text{mom}}}^c = \mathbb{r}_\kappa^{\text{exp}}. \quad (14)$$

4.2 | Physical constraints defined by the mean value of a stochastic equation relating \mathbf{Q}^c and \mathbf{W}^c

Let us assume that the initial data set $D_N = \{\mathbf{x}^j, j = 1, \dots, N\}$ corresponds to N experimental realizations (measurements or numerical simulations obtained by the training a stochastic computational model) of random vector $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ whose probability distribution on $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ admits a density with respect to the Lebesgue measure. In this appendix, we consider the case for which the PLoM is used for generating M additional realizations (using only initial data set D_N) under the constraints defined by statistical moments (for instance, the mean value) of a stochastic equation corresponding to a stochastic computational model of a physical system. As previously, let $p_{\mathbf{X}^c}$ be the pdf of the random vector $\mathbf{X}^c = (\mathbf{Q}^c, \mathbf{W}^c)$

with values in $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$, which has to verify, in statistical mean, the equation $\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c) = \boldsymbol{\beta}^{\text{const}}$, that is to say, $E\{\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c)\} = \boldsymbol{\beta}^{\text{const}}$ in which $(\mathbf{q}^c, \mathbf{w}^c) \mapsto \mathbf{f}^{\text{const}}(\mathbf{q}^c, \mathbf{w}^c)$ is a given (deterministic) measurable mapping defined on $\mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ with values in \mathbb{R}^{m_c} and where $\boldsymbol{\beta}^{\text{const}}$ is a given vector in \mathbb{R}^{m_c} . This constraint can then be rewritten as Equation (10) with $\mathbf{g}^c(\mathbf{X}^c) = \mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c)$ and $\boldsymbol{\beta}^c = \boldsymbol{\beta}^{\text{const}}$. For instance, function $\mathbf{f}^{\text{const}}$ could be written, for $m_c = n_q$, as

$$\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c) = \mathbf{Q}^c - [\underline{B}]\mathbf{W}^c, \quad (15)$$

in which $[\underline{B}]$ is a given matrix in \mathbb{M}_{n_q, n_w} .

It should be noted that this formulation can be extended to a stochastic equation that would be written, for instance, as $E\{\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c, \mathbb{U})\} = \boldsymbol{\beta}^{\text{const}}$ in which \mathbb{U} is a random vector independent of \mathbf{W}^c and where $\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c, \mathbb{U}) = \mathbf{Q}^c - [\underline{B}(\mathbb{U})]\mathbf{W}^c$. Clearly, any more complex stochastic equation can be used.

5 | COMPUTATIONAL STATISTICAL METHOD FOR GENERATING REALIZATIONS OF \mathbf{X}^c USING PROBABILISTIC LEARNING ON MANIFOLDS

In this section, we present a computational method for generating M independent realizations $\mathbf{x}^{c,1}, \dots, \mathbf{x}^{c,M}$ of random variable \mathbf{X}^c whose pdf $p_{\mathbf{X}^c}$ is defined by Equations (8) and (9). Because it is assumed that N is small and because the available information is only made up of $\{\mathbf{x}^j, j = 1, \dots, N\}$ and of the constraint equation $E\{\mathbf{g}^c(\mathbf{X}^c)\} = \boldsymbol{\beta}^c$, we propose to use the PLoM that will be developed in Section 5.8. The main steps of the computational statistical method are as follows:

1. principal component analysis (PCA) of \mathbf{X} at order $v \leq n$ using the set $\{\mathbf{x}^j, j = 1, \dots, N\}$, which will introduce the normalized non-Gaussian random vector \mathbf{H} with values in \mathbb{R}^v , centered, with covariance matrix $[I_v]$, for which the N independent realizations are $\{\boldsymbol{\eta}^j, j = 1, \dots, N\}$;
2. Gaussian kernel-density estimation of the pdf $p_{\mathbf{H}}$ of random vector \mathbf{H} ;
3. reformulating the optimization problem defined by Equations (8) and (9) for the construction of pdf $p_{\mathbf{H}^c}$ of \mathbf{H}^c in terms of pdf $p_{\mathbf{H}}$ of \mathbf{H} and of the constraint equation rewritten in terms of $p_{\mathbf{H}^c}$;
4. representation of the solution of the optimization problem with constraints using Lagrange multipliers (λ_0, λ) where λ_0 is associated with the normalization constraint;
5. reformulation introducing a random vector \mathbf{H}_λ and eliminating λ_0 ;
6. introducing an iterative algorithm for computing the optimal value λ^{sol} of the Lagrange multiplier λ ;
7. constructing a nonlinear ISDE as a generator of random variable \mathbf{H}_λ ;
8. computing the additional realizations of \mathbf{H}^c using the PLoM and then deducing additional realizations of \mathbf{X}^c ;
9. estimating statistics of \mathbf{X}^c using the additional realizations.

In the following, each of these steps is expanded into a subsection.

5.1 | PCA of random vector \mathbf{X}

The purpose of the PCA transformation is to improve the statistical condition of \mathbf{X} through decorrelation and normalization. Dimensional reduction is only accepted if it reproduces the original data with a given tolerance. Let $\underline{\mathbf{x}} \in \mathbb{R}^n$ and $[\hat{C}_{\mathbf{X}}] \in \mathbb{M}_n^+$ be the classical empirical estimates of the mean vector and the covariance matrix of \mathbf{X} ,

$$\underline{\mathbf{x}} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}^j, \quad [\hat{C}_{\mathbf{X}}] = \frac{1}{N-1} \sum_{j=1}^N (\mathbf{x}^j - \underline{\mathbf{x}})(\mathbf{x}^j - \underline{\mathbf{x}})^T. \quad (16)$$

Let $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n \geq 0$ be the eigenvalues and let $\boldsymbol{\varphi}^1, \dots, \boldsymbol{\varphi}^n$ be the orthonormal eigenvectors of $[\hat{C}_{\mathbf{X}}]$ such that $[\hat{C}_{\mathbf{X}}] \boldsymbol{\varphi}^\alpha = \mu_\alpha \boldsymbol{\varphi}^\alpha$. For $\varepsilon > 0$ fixed, let v be the integer such that $1 \leq v \leq n$, verifying

$$\text{err}_{\text{PCA}}(v) = 1 - \frac{\sum_{\alpha=1}^v \mu_\alpha}{\text{tr}[\hat{C}_{\mathbf{X}}]} \leq \varepsilon. \quad (17)$$

The PCA of \mathbf{X} allows for representing \mathbf{X} by \mathbf{X}^v such that

$$\mathbf{X}^v = \underline{\hat{\mathbf{x}}} + [\Phi][\mu]^{1/2}\mathbf{H}, \quad E\{\|\mathbf{X} - \mathbf{X}^v\|^2\} \leq \varepsilon E\{\|\mathbf{X}\|^2\}, \quad (18)$$

in which $[\Phi] = [\varphi^1 \dots \varphi^v] \in \mathbb{M}_{n,v}$ such that $[\Phi]^T[\Phi] = [I_v]$ and $[\mu]$ is the diagonal ($v \times v$) matrix such that $[\mu]_{\alpha\beta} = \mu_\alpha \delta_{\alpha\beta}$. The \mathbb{R}^v -valued random variable \mathbf{H} is obtained by projection,

$$\mathbf{H} = [\mu]^{-1/2}[\Phi]^T(\mathbf{X} - \underline{\hat{\mathbf{x}}}), \quad (19)$$

and its N independent realizations $\{\boldsymbol{\eta}^j, j = 1, \dots, N\}$ are such that

$$\boldsymbol{\eta}^j = [\mu]^{-1/2}[\Phi]^T(\mathbf{x}^j - \underline{\hat{\mathbf{x}}}) \in \mathbb{R}^v. \quad (20)$$

The empirical estimates of the mean vector and the covariance matrix of \mathbf{H} verify the following conditions:

$$\underline{\hat{\boldsymbol{\eta}}} = \frac{1}{N} \sum_{j=1}^N \boldsymbol{\eta}^j = \mathbf{0}, \quad [\hat{C}_{\mathbf{H}}] = \frac{1}{N-1} \sum_{j=1}^N (\boldsymbol{\eta}^j - \underline{\hat{\boldsymbol{\eta}}})(\boldsymbol{\eta}^j - \underline{\hat{\boldsymbol{\eta}}})^T = [I_v]. \quad (21)$$

From a numerical point of view, if $N < n$, the estimate $[\hat{C}_{\mathbf{X}}]$ of covariance matrix of \mathbf{X} is not computed and $v, [\mu]$, and $[\Phi]$ are directly computed using a thin SVD of the matrix whose N columns are $(\mathbf{x}^j - \underline{\hat{\mathbf{x}}})$ for $j = 1, \dots, N$.

Hypothesis. Let $L^2(\Theta, \mathbb{R}^n)$ be the space of second-order random variables \mathbf{X} defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in \mathbb{R}^n such that $E\{\|\mathbf{X}\|^2\} < +\infty$. Similarly, let $L^2(\Theta, \mathbb{R}^v)$ be the space of all the second-order random variables \mathbf{H} defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in \mathbb{R}^v such that $E\{\|\mathbf{H}\|^2\} < +\infty$. It should be noted that we do not consider the subset \mathcal{H}_v of $L^2(\Theta, \mathbb{R}^v)$ made up of all the second-order \mathbb{R}^v -valued random variables \mathbf{H} such that $E\{\mathbf{H}\} = \mathbf{0}$ and $[C_{\mathbf{H}}] = [I_v]$ because the constraints will incur a shift and scaling that is a priori unknown. For fixed $v \leq n$ that verifies Equation (17), random vector \mathbf{X}^v defined by Equation (18) spans a subspace \mathcal{X}_v of $L^2(\Theta, \mathbb{R}^n)$ when \mathbf{H} runs through $L^2(\Theta, \mathbb{R}^v)$. It is then assumed that random vector \mathbf{X}^c , for which pdf $p_{\mathbf{X}^c}$ is closest to pdf $p_{\mathbf{X}}$ under the constraint $E\{\mathbf{g}^c(\mathbf{X}^c)\} = \beta^c$ (see Section 3), belongs to \mathcal{X}_v . It should be noted that there always exists v such that $v \leq n$ with $err_{PCA}(v) \leq \varepsilon$ in order that $\mathbf{X}^c \in \mathcal{X}_v$ because for $v = n$ (that is to say, for ε sufficiently small), we have $\mathcal{X}_n = L^2(\Theta, \mathbb{R}^n)$.

5.2 | Gaussian kernel-density estimation of the pdf of random vector \mathbf{H}

The pdf $\boldsymbol{\eta} \mapsto p_{\mathbf{H}}(\boldsymbol{\eta})$ on \mathbb{R}^v of the \mathbb{R}^v -valued random variable \mathbf{H} is constructed by the Gaussian kernel-density estimation method using the N independent realizations $\{\boldsymbol{\eta}^j, j = 1, \dots, N\}$ defined by Equation (20). We use the modification proposed in the work of Soize⁴⁴ of the classical formulation,⁴⁵ which can be written, for all $\boldsymbol{\eta}$ in \mathbb{R}^v , as

$$p_{\mathbf{H}}(\boldsymbol{\eta}) = c_v \rho(\boldsymbol{\eta}), \quad c_v = \frac{1}{\left(\sqrt{2\pi}\hat{s}_v\right)^v}, \quad (22)$$

$$\rho(\boldsymbol{\eta}) = \frac{1}{N} \sum_{j=1}^N \exp \left\{ -\frac{1}{2\hat{s}_v^2} \left\| \frac{\hat{s}_v}{s_v} \boldsymbol{\eta}^j - \boldsymbol{\eta} \right\|^2 \right\}, \quad (23)$$

in which s_v is the usual multidimensional optimal Silverman bandwidth (taking into account that the covariance matrix of \mathbf{H} is $[I_v]$), and where \hat{s}_v has been introduced in order that the second equation in Equation (21) be satisfied,

$$s_v = \left\{ \frac{4}{N(2+v)} \right\}^{1/(v+4)}, \quad \hat{s}_v = \frac{s_v}{\sqrt{s_v^2 + (N-1)/N}}. \quad (24)$$

With this formulation, Equations (21) to (24) yield

$$E\{\mathbf{H}\} = \int_{\mathbb{R}^v} \boldsymbol{\eta} p_{\mathbf{H}}(\boldsymbol{\eta}) d\boldsymbol{\eta} = \frac{\hat{s}_v}{s_v} \hat{\boldsymbol{\eta}} = \mathbf{0}, \quad (25)$$

$$E\{\mathbf{H}\mathbf{H}^T\} = \int_{\mathbb{R}^v} \boldsymbol{\eta}\boldsymbol{\eta}^T p_{\mathbf{H}}(\boldsymbol{\eta}) d\boldsymbol{\eta} = \hat{s}_v^2 [I_v] + \frac{\hat{s}_v^2 (N-1)}{s_v^2 N} [\hat{C}_{\mathbf{H}}] = [I_v]. \quad (26)$$

The pdf of \mathbf{H} , defined by Equations (22) to (24), is rewritten, for all $\boldsymbol{\eta}$ in \mathbb{R}^v , as

$$p_{\mathbf{H}}(\boldsymbol{\eta}) = c_v e^{-\psi(\boldsymbol{\eta})}, \quad \psi(\boldsymbol{\eta}) = -\log \rho(\boldsymbol{\eta}). \quad (27)$$

5.3 | Reformulating the optimization problem as a function of random vector \mathbf{H}

The objective is to reformulate the optimization problem defined by Equations (8) and (9) to evaluate the pdf $p_{\mathbf{H}}$ defined by Equation (20) and (22) to (24). Taking into account Equation (18) and the hypothesis introduced in Section 5.1, the \mathbb{R}^v -valued random variable $\mathbf{X}^{c,v}$ can be written as

$$\mathbf{X}^{c,v} = \underline{\hat{\mathbf{x}}} + [\Phi][\mu]^{1/2} \mathbf{H}^c, \quad (28)$$

in which the \mathbb{R}^v -valued random variable \mathbf{H}^c is defined by

$$\mathbf{H}^c = [\mu]^{-1/2} [\Phi]^T (\mathbf{X}^c - \underline{\hat{\mathbf{x}}}). \quad (29)$$

Note that, in general, \mathbf{H}^c is not centered and its covariance matrix is not $[I_v]$. In addition, for ϵ sufficiently small verifying Equation (17), the mean-square error $E\{\|\mathbf{X}^c - \mathbf{X}^{c,v}\|^2\}$ is sufficiently small (note that such condition always exists because the limiting case corresponds to $v = n$). Due to the introduction of the reduced-order representation $\mathbf{X}^{c,v}$ of \mathbf{X}^c defined by Equation (28), the stochastic dimension of $\mathbf{X}^{c,v}$ is less than or equal to v , because \mathbf{H}^c is a \mathbb{R}^v -valued random variable.

Let $\boldsymbol{\eta} \mapsto \tilde{\mathbf{h}}^c(\boldsymbol{\eta})$ be the measurable mapping from \mathbb{R}^v into \mathbb{R}^{m_c} such that, for all $\boldsymbol{\eta}$ in \mathbb{R}^v ,

$$\tilde{\mathbf{h}}^c(\boldsymbol{\eta}) = \mathbf{g}^c(\underline{\hat{\mathbf{x}}} + [\Phi][\mu]^{1/2} \boldsymbol{\eta}). \quad (30)$$

If it is assumed that the value of v is such that β^c (defined by Equation (10)) belongs to the range space $\tilde{\mathbf{h}}^c(\mathbb{R}^v)$ of mapping $\tilde{\mathbf{h}}^c$, then the constraint defined by Equation (10) is transported on \mathbf{H}^c and writes

$$E\{\tilde{\mathbf{h}}^c(\mathbf{H}^c)\} = \beta^c, \quad \int_{\mathbb{R}^v} \tilde{\mathbf{h}}^c(\boldsymbol{\eta}) p_{\mathbf{H}^c}(\boldsymbol{\eta}) d\boldsymbol{\eta} = \beta^c, \quad (31)$$

in which the pdf $p_{\mathbf{H}^c}$ of \mathbf{H}^c is closest to pdf $p_{\mathbf{H}}$ while satisfying the constraint defined by Equation (31).

We propose to perform the projection of the \mathbb{R}^{m_c} -valued random variable $\tilde{\mathbf{h}}^c(\mathbf{H})$ in a \mathbb{R}^{v_c} -valued random variable $\mathbf{h}^c(\mathbf{H}) = [\Psi]^T \tilde{\mathbf{h}}^c(\mathbf{H})$ with $v_c \leq v$ in which the matrix $[\Psi] \in \mathbb{M}_{m_c, v_c}$ verifies $[\Psi]^T [\Psi] = [I_{v_c}]$. This projection is carried out in order that the projected constraints be algebraically independent (see Section 5.4-(iv)) and that the components of the random vector $\mathbf{h}^c(\mathbf{H})$ be decorrelated and not degenerated. Note that, in addition, with such a projection, the computation is less costly and more robust for large value of m_c . For the second criterion, matrix $[\Psi]$ and v_c are constructed as follows. Let $[\text{cov}\{\tilde{\mathbf{h}}^c(\mathbf{H})\}]$ be the covariance matrix in $\mathbb{M}_{m_c}^{+0}$ of the \mathbb{R}^{m_c} -valued random variable $\tilde{\mathbf{h}}^c(\mathbf{H})$ in which \mathbf{H} is the \mathbb{R}^v -valued random variable defined by Equation (19) for which the N independent realizations are defined by Equation (20). Let

$$[\text{cov}\{\tilde{\mathbf{h}}^c(\mathbf{H})\}] \boldsymbol{\psi}^\alpha = \lambda_\alpha^c \boldsymbol{\psi}^\alpha$$

be the eigenvalue problem such that

$$\lambda_1^c \geq \dots \geq \lambda_{v_c}^c > 0 = \lambda_{v_c+1}^c = \dots = \lambda_{m_c}^c.$$

Consequently, the rank of $[\text{cov}\{\tilde{\mathbf{h}}^c(\mathbf{H})\}]$ is

$$\text{rank}\{[\text{cov}\{\tilde{\mathbf{h}}^c(\mathbf{H})\}]\} = v_c \leq m_c. \quad (32)$$

Let $[\Psi] = [\boldsymbol{\psi}^1 \dots \boldsymbol{\psi}^{v_c}] \in \mathbb{M}_{m_c, v_c}$ be the matrix whose columns are the orthonormal eigenvectors $\boldsymbol{\psi}^1, \dots, \boldsymbol{\psi}^{v_c}$ associated with $\lambda_1^c \geq \dots \geq \lambda_{v_c}^c > 0$. Therefore, we have $[\Psi]^T [\Psi] = [I_{v_c}]$. Consequently, the covariance matrix $[\text{cov}\{\mathbf{h}^c(\mathbf{H})\}] \in \mathbb{M}_{v_c}^+$ is diagonal and invertible (the components of random vector $\mathbf{h}^c(\mathbf{H})$ are decorrelated and not degenerated).

From a numerical point of view, if $N < m_c$, the estimate $[\widehat{\text{cov}}\{\tilde{\mathbf{h}}^c(\mathbf{H})\}]$ of covariance matrix of $\tilde{\mathbf{h}}^c(\mathbf{H})$ is not computed, and $v_c, \lambda_1^c, \dots, \lambda_{v_c}^c, [\Psi]$ are directly computed using a thin SVD of the matrix whose N columns are $(\tilde{\mathbf{h}}^c(\boldsymbol{\eta}^j) - \underline{\mathbf{h}}^c)$ for $j = 1, \dots, N$, where $\underline{\mathbf{h}}^c$ is the estimate of the mean value of random vector $\tilde{\mathbf{h}}^c(\mathbf{H})$, which is performed using the realizations $\{\tilde{\mathbf{h}}^c(\boldsymbol{\eta}^j), j = 1, \dots, N\}$. The projection of Equation (31) is performed using the basis $[\Psi]$ that spans a subspace of \mathbb{R}^{m_c} of dimension v_c and is written as

$$E\{\mathbf{h}^c(\mathbf{H}^c)\} = \mathbf{b}^c, \quad (33)$$

in which $\boldsymbol{\eta} \mapsto \mathbf{h}^c(\boldsymbol{\eta})$ is the measurable mapping from \mathbb{R}^ν into \mathbb{R}^{v_c} , with

$$v_c \leq \nu, \quad (34)$$

such that, for all $\boldsymbol{\eta}$ in \mathbb{R}^ν ,

$$\mathbf{h}^c(\boldsymbol{\eta}) = [\Psi]^T \tilde{\mathbf{h}}^c(\boldsymbol{\eta}) = [\Psi]^T \mathbf{g}^c(\underline{\mathbf{x}} + [\Phi][\mu]^{1/2} \boldsymbol{\eta}) \in \mathbb{R}^{v_c}, \quad (35)$$

and where \mathbf{b}^c is written as

$$\mathbf{b}^c = [\Psi]^T \boldsymbol{\beta}^c \in \mathbb{R}^{v_c}. \quad (36)$$

Using again the Kullback-Leibler minimum cross-entropy principle (as done in Section 3) for random variable \mathbf{H}^c , the optimization problem defined by Equations (8) and (9) is reformulated as

$$p_{\mathbf{H}^c} = \arg \min_{\hat{p} \in \mathcal{C}_{\text{ad}, \mathbf{H}^c}} \int_{\mathbb{R}^\nu} \hat{p}(\boldsymbol{\eta}) \log \frac{\hat{p}(\boldsymbol{\eta})}{p_{\mathbf{H}}(\boldsymbol{\eta})} d\boldsymbol{\eta}, \quad (37)$$

in which the admissible set $\mathcal{C}_{\text{ad}, \mathbf{H}^c}$ is defined by

$$\mathcal{C}_{\text{ad}, \mathbf{H}^c} = \{\boldsymbol{\eta} \mapsto \hat{p}(\boldsymbol{\eta}) : \mathbb{R}^\nu \rightarrow \mathbb{R}^+, \int_{\mathbb{R}^\nu} \hat{p}(\boldsymbol{\eta}) d\boldsymbol{\eta} = 1, \int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta}) \hat{p}(\boldsymbol{\eta}) d\boldsymbol{\eta} = \mathbf{b}^c\}. \quad (38)$$

Remark concerning the calculation of the rank v_c

The rank v_c defined by Equation (32) is a function of the tolerance $\tau_c > 0$ that is used for numerically testing the condition $\{\lambda_{v_c}^c > 0\} \cap \{\lambda_{v_c+1}^c = 0\}$. This test is rewritten as follows:

$$\text{If } \{\lambda_{v_c}^c > \tau_c \lambda_1^c\} \cap \{\lambda_{v_c+1}^c \leq \tau_c \lambda_1^c\}, \text{ then the rank is } v_c. \quad (39)$$

In the linear algebra libraries devoted to the computation of the rank of a matrix, τ_c is generally chosen as 10^{-12} . It should be noted that the smaller τ_c the larger v_c . Therefore, the value of v_c that is less than or equal to m_c will play an important role on the existence of a unique solution of the optimization problem defined by Equation (37) with Equation (38). This aspect will be discussed in Section 5.11.

5.4 | Existence of a solution and its representation using Lagrange multipliers (λ_0, λ)

The optimization problem defined by Equation (37) with Equation (38) is solved by introducing Lagrange multipliers as done for the maximum entropy principle.^{8,11,46-49}

(i) Lagrange multipliers associated with the constraints

A Lagrange multipliers is introduced for each constraint:

$$\lambda_0 \in \mathbb{R}^+ \text{ for the constraint: } \int_{\mathbb{R}^\nu} \hat{p}(\boldsymbol{\eta}) d\boldsymbol{\eta} = 1,$$

$$\lambda \in \mathcal{C}_{\text{ad}, \lambda} \subset \mathbb{R}^{v_c} \text{ for the constraint: } \int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta}) \hat{p}(\boldsymbol{\eta}) d\boldsymbol{\eta} = \mathbf{b}^c,$$

in which the admissible set $C_{\text{ad},\lambda}$ is defined as the open subset of \mathbb{R}^{v_c} such that

$$C_{\text{ad},\lambda} = \left\{ \boldsymbol{\lambda} \in \mathbb{R}^{v_c}, \int_{\mathbb{R}^v} e^{-\psi(\boldsymbol{\eta}) - \langle \boldsymbol{\lambda}, \mathbf{h}^c(\boldsymbol{\eta}) \rangle} d\boldsymbol{\eta} < +\infty \right\}, \quad (40)$$

where ψ is the potential of \mathbf{H} defined by Equation (27). The form of the admissible set $C_{\text{ad},\lambda}$ follows from the representation using Lagrange multipliers.

(ii) Expression of the Lagrangian

For $\lambda_0 \in \mathbb{R}^+$ and $\lambda \in C_{\text{ad},\lambda}$, the Lagrangian associated with Equations (37) and (38) is written as

$$\begin{aligned} \mathcal{L}\text{ag}(\hat{p}; \lambda_0, \lambda) = & \int_{\mathbb{R}^v} \hat{p}(\boldsymbol{\eta}) \log \frac{\hat{p}(\boldsymbol{\eta})}{p_{\mathbf{H}}(\boldsymbol{\eta})} d\boldsymbol{\eta} - (\lambda_0 - 1) \left(\int_{\mathbb{R}^v} \hat{p}(\boldsymbol{\eta}) d\boldsymbol{\eta} - 1 \right) \\ & - \left\langle \lambda, \int_{\mathbb{R}^v} \mathbf{h}^c(\boldsymbol{\eta}) \hat{p}(\boldsymbol{\eta}) d\boldsymbol{\eta} - \mathbf{b}^c \right\rangle. \end{aligned} \quad (41)$$

(iii) Reformulation of the optimization problem and construction of the solution

If it is assumed that there exists a unique solution $p_{\mathbf{H}^c}$ to the optimization problem defined by Equation (37) with Equation (38), then, using simple arguments from the calculus of variations, it can be seen that there exists $(\lambda_0^{\text{sol}}, \lambda^{\text{sol}}) \in \mathbb{R}^+ \times C_{\text{ad},\lambda}$, such that the functional $(\hat{p}, \lambda_0, \lambda) \mapsto \mathcal{L}\text{ag}(\hat{p}; \lambda_0, \lambda)$ is stationary at $p_{\mathbf{H}^c}$ for $\lambda_0 = \lambda_0^{\text{sol}}$ and $\lambda = \lambda^{\text{sol}}$, and $p_{\mathbf{H}^c}$ can be written, for all $\boldsymbol{\eta}$ in \mathbb{R}^v , as

$$p_{\mathbf{H}^c}(\boldsymbol{\eta}) = c_v \exp \left\{ -\psi(\boldsymbol{\eta}) - \lambda_0^{\text{sol}} - \langle \lambda^{\text{sol}}, \mathbf{h}^c(\boldsymbol{\eta}) \rangle \right\}, \quad (42)$$

in which c_v and $\psi(\boldsymbol{\eta})$ are defined by Equations (22) and (27).

In addition, always under the hypothesis that there exists a unique solution to the optimization problem defined by Equation (37) with Equation (38), and under mild additional hypotheses, it is proven in Appendix B of the work of Guilleminot and Soize^{50(pp126-128)} that the unique solution is given by Equation (42) in which $(\lambda_0^{\text{sol}}, \lambda^{\text{sol}})$ is the unique solution in $\mathbb{R}^+ \times C_{\text{ad},\lambda}$ of the following nonlinear algebraic equations,

$$\int_{\mathbb{R}^v} c_v \exp \left\{ -\psi(\boldsymbol{\eta}) - \lambda_0 - \langle \lambda, \mathbf{h}^c(\boldsymbol{\eta}) \rangle \right\} d\boldsymbol{\eta} = 1, \quad (43)$$

$$\int_{\mathbb{R}^v} c_v \mathbf{h}^c(\boldsymbol{\eta}) \exp \left\{ -\psi(\boldsymbol{\eta}) - \lambda_0 - \langle \lambda, \mathbf{h}^c(\boldsymbol{\eta}) \rangle \right\} d\boldsymbol{\eta} = \mathbf{b}^c. \quad (44)$$

(iv) Comments about the existence of a unique solution

The hypothesis of the existence of a unique solution for the optimization problem defined by Equations (37) with (38) does not necessarily hold for arbitrary combinations of initial data set D_N and constraint equations. As explained in Section 5.3, a first necessary condition to guarantee the existence of a unique solution of the optimization problem is that the constraints be algebraically independent, which can be checked using the following criterion: There exists a bounded subset B of \mathbb{R}^{v_c} with $\int_B d\boldsymbol{\eta} > 0$ such that, for any nonzero vector \mathbf{v} in \mathbb{R}^{1+v_c} ,

$$\int_B \langle \mathbf{v}, \hat{\mathbf{h}}^c(\boldsymbol{\eta}) \rangle^2 d\boldsymbol{\eta} > 0 \quad \text{with} \quad \hat{\mathbf{h}}^c(\boldsymbol{\eta}) = (1, \mathbf{h}^c(\boldsymbol{\eta})) \in \mathbb{R}^{1+v_c}. \quad (45)$$

It should be noted that, in general, this criterion requires a numerical evaluation. If the given constraints are not consistent with D_N , the admissible set $C_{\text{ad},\lambda}$ may be the empty set. This necessary condition is therefore not sufficient and there is no mathematical criterion that can easily be implemented, which gives a necessary and sufficient condition to guarantee the existence of a unique solution as a function of D_N , \mathbf{h}^c , and \mathbf{b}^c . This problem relative to existence and uniqueness will be revisited from a numerical point of view in Section 5.11.

5.5 | Reformulation introducing a random vector \mathbf{H}_λ and eliminating λ_0

In general, only a numerical approximation to the solution of Equations (43) and (44) can be achieved. However, a major difficulty is introduced by the normalization constant $c_v e^{-\lambda_0}$ that induces considerable difficulties when v is large. In addition, integrals in high dimension have to be computed for calculating the Lagrange multipliers. In this section, a procedure is introduced to address this difficulty. This is done by first eliminating λ_0 , followed by introducing a convex objective function that lends itself to efficient iterative schemes, and finally, by using the PLoM as a MCMC generator of realizations allowing the integrals to be computed.

5.5.1 | Introduction of the \mathbb{R}^v -valued random variable \mathbf{H}_λ and eliminating λ_0

For λ given in $C_{ad,\lambda}$, let \mathbf{H}_λ be the \mathbb{R}^v -valued random variable whose pdf with respect to $d\boldsymbol{\eta}$ on \mathbb{R}^v is written, for all $\boldsymbol{\eta} \in \mathbb{R}^v$, as

$$p_{\mathbf{H}_\lambda}(\boldsymbol{\eta}) = c_0(\lambda) \exp\{-\psi(\boldsymbol{\eta}) - \langle \lambda, \mathbf{h}^c(\boldsymbol{\eta}) \rangle\}, \quad (46)$$

in which the function $\lambda \mapsto c_0(\lambda)$ from $C_{ad,\lambda} \subset \mathbb{R}^{v_c}$ into $]0, +\infty[$ is defined by

$$c_0(\lambda) = \left(\int_{\mathbb{R}^v} \exp\{-\psi(\boldsymbol{\eta}) - \langle \lambda, \mathbf{h}^c(\boldsymbol{\eta}) \rangle\} d\boldsymbol{\eta} \right)^{-1}. \quad (47)$$

It can be easily verified that $c_0(\lambda) = c_v \exp\{-\lambda_0\}$ and, consequently, for all $\boldsymbol{\eta}$ in \mathbb{R}^v ,

$$p_{\mathbf{h}^c}(\boldsymbol{\eta}) = \{p_{\mathbf{h}_\lambda}(\boldsymbol{\eta})\}_{\lambda=\lambda^{sol}}, \quad p_{\mathbf{h}}(\boldsymbol{\eta}) = \{p_{\mathbf{h}_\lambda}(\boldsymbol{\eta})\}_{\lambda=0}. \quad (48)$$

Because $c_0(\lambda)$ is defined as the constant of normalization of pdf $p_{\mathbf{h}_\lambda}$, the constraint defined by Equation (43), which is rewritten as $\int_{\mathbb{R}^v} p_{\mathbf{h}_\lambda}(\boldsymbol{\eta}) d\boldsymbol{\eta} = 1$ is automatically satisfied for all λ in $C_{ad,\lambda}$. Using Equations (46) and (47), it can be seen that the constraint defined by Equation (44) can be rewritten, for all λ in $C_{ad,\lambda}$, as

$$E\{\mathbf{h}^c(\mathbf{h}_\lambda)\} = \mathbf{b}^c, \quad (49)$$

because $E\{\mathbf{h}^c(\mathbf{h}_\lambda)\} = \int_{\mathbb{R}^v} \mathbf{h}^c(\boldsymbol{\eta}) p_{\mathbf{h}_\lambda}(\boldsymbol{\eta}) d\boldsymbol{\eta}$. From Section 5.4-(iv), it can then be deduced that $\lambda^{sol} \in C_{ad,\lambda}$ is the unique solution in λ of Equation (49).

5.5.2 | Construction of an optimization problem with a convex objective function for calculating λ^{sol}

Similarly to the discrete-case approach proposed in the work of Agmon et al,⁵¹ let $\lambda \mapsto \Gamma(\lambda)$ be the objective function from the open subset $C_{ad,\lambda}$ of \mathbb{R}^{v_c} into \mathbb{R} such that

$$\Gamma(\lambda) = \langle \lambda, \mathbf{b}^c \rangle - \log c_0(\lambda). \quad (50)$$

It can easily be deduced that the gradient of Γ with respect to λ is written as

$$\nabla \Gamma(\lambda) = \mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H}_\lambda)\}, \quad (51)$$

and that the Hessian matrix, $[\Gamma''(\lambda)]$, of Γ at λ is written as

$$[\Gamma''(\lambda)] = [\text{cov}\{\mathbf{h}^c(\mathbf{H}_\lambda)\}], \quad (52)$$

in which the covariance matrix $[\text{cov}\{\mathbf{h}^c(\mathbf{H}_\lambda)\}]$ of random vector $\mathbf{h}^c(\mathbf{H}_\lambda)$ is such that

$$[\text{cov}\{\mathbf{h}^c(\mathbf{H}_\lambda)\}] = E\{\mathbf{h}^c(\mathbf{H}_\lambda) \mathbf{h}^c(\mathbf{H}_\lambda)^T\} - E\{\mathbf{h}^c(\mathbf{H}_\lambda)\} E\{\mathbf{h}^c(\mathbf{H}_\lambda)\}^T. \quad (53)$$

By construction (see Section 5.3), we have $[\text{cov}\{\mathbf{h}^c(\mathbf{H})\}] = [\Psi]^T [\text{cov}\{\tilde{\mathbf{h}}^c(\mathbf{H})\}] [\Psi]$. We will assume that, for all $\lambda \in C_{\text{ad},\lambda}$, the covariance matrix $[\text{cov}\{\mathbf{h}^c(\mathbf{H}_\lambda)\}]$ is positive definite (this is a reasonable hypothesis due to the construction of $[\Psi]$ and due to the hypothesis of algebraical independence of the constraints (see Equation (45)). Consequently, $\lambda \mapsto \Gamma(\lambda)$ is a strictly convex function in $C_{\text{ad},\lambda}$. Because $\lambda^{\text{sol}} \in C_{\text{ad},\lambda}$ is the unique solution of Equation (49), we have $E\{\mathbf{h}^c(\mathbf{H}_{\lambda^{\text{sol}}})\} = \mathbf{b}^c$. Therefore, Equation (51) yields $\nabla\Gamma(\lambda^{\text{sol}}) = \mathbf{0}$. Consequently, λ^{sol} is the unique solution of the following optimization problem:

$$\lambda^{\text{sol}} = \arg \min_{\lambda \in C_{\text{ad},\lambda} \subset \mathbb{R}^{n_u}} \Gamma(\lambda). \quad (54)$$

Because Γ is convex, this optimization problem would be convex if $C_{\text{ad},\lambda}$ was convex and then λ^{sol} would be the unique global minimum that could be computed with the efficient algorithms devoted to convex optimization problems. However, it cannot be proven that $C_{\text{ad},\lambda}$ is convex for arbitrary constraints \mathbf{h}^c .

5.6 | Iterative algorithm for computing the Lagrange multipliers λ^{sol}

Taking into account the convexity of Γ , we propose to solve the minimization problem defined by Equation (54) with a Newton iterative method applied to function $\lambda \mapsto \nabla\Gamma(\lambda)$ ^{49,52} starting from an initial point λ_1 in $C_{\text{ad},\lambda}$. This algorithm allows for finding λ^{sol} such that $\nabla\Gamma(\lambda^{\text{sol}}) = \mathbf{0}$ and is written, for $i \geq 1$, as

$$\lambda^{i+1} = \lambda^i - [\Gamma''(\lambda^i)]^{-1} \nabla\Gamma(\lambda^i), \quad (55)$$

in which $\nabla\Gamma(\lambda^i)$ and $[\Gamma''(\lambda^i)]$ are given by Equations (51) and (52), that is to say,

$$\nabla\Gamma(\lambda^i) = \mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H}_{\lambda^i})\}, \quad [\Gamma''(\lambda^i)] = [\text{cov}\{\mathbf{h}^c(\mathbf{H}_{\lambda^i})\}]. \quad (56)$$

The initial point, which corresponds to $i = 1$, is chosen such that

$$\lambda^1 = -[\Gamma''(\lambda^0)]^{-1} \nabla\Gamma(\lambda^0), = \quad (57)$$

with $\lambda^0 = \mathbf{0}$, and because $\mathbf{H}_0 = \mathbf{H}$ (see Equation (48)), we have

$$\nabla\Gamma(\lambda^0) = \mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H})\}, \quad [\Gamma''(\lambda^0)] = [\text{cov}\{\mathbf{h}^c(\mathbf{H})\}]. \quad (58)$$

Consequently, the necessary condition concerning the initial point in the Newton method is satisfied: λ^1 is not close to origin $\mathbf{0}$ because $\mathbf{b}^c \neq E\{\mathbf{h}^c(\mathbf{H})\}$ and we have $\lambda^1 \in C_{\text{ad},\lambda}$. The right-hand sides of the two equations in Equation (58) can be estimated using the empirical estimators of the mean vector and covariance matrix of $\mathbf{h}^c(\mathbf{H})$ using the initial data set $\{\boldsymbol{\eta}^j, j = 1, \dots, N\}$ defined by Equation (20). At each iteration i , the convergence of the iteration algorithm is controlled by calculating the error

$$\text{err}(i) = \frac{\|\mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H}_{\lambda^i})\}\|}{\|\mathbf{b}^c\|} = \frac{\|\nabla\Gamma(\lambda^i)\|}{\|\mathbf{b}^c\|}. \quad (59)$$

During the iteration, for each λ^i given in $C_{\text{ad},\lambda} \subset \mathbb{R}^{n_u}$, the following mathematical expectations:

$$E\{\mathbf{h}^c(\mathbf{H}_{\lambda^i})\}, \quad E\{\mathbf{h}^c(\mathbf{H}_{\lambda^i})\mathbf{h}^c(\mathbf{H}_{\lambda^i})^T\}, \quad (60)$$

have to be computed using an MCMC algorithm (see Section 5.7). It should be noted that admissible set $C_{\text{ad},\lambda}$ cannot, in the general case, explicitly be described. An additional discussion concerning the convergence of the iteration algorithm in the framework of the existence of a solution is given in Section 5.11.

5.7 | Nonlinear ISDE for the generator of random variable \mathbf{H}_λ

For λ fixed in $C_{\text{ad},\lambda}$, we have to construct a generator for \mathbf{H}_λ in order to compute an estimation of the quantities listed in Equation (60). We propose to use the approach presented in the works of Soize,^{49,53} which is based on a nonlinear ISDE associated with a nonlinear stochastic dissipative Hamiltonian dynamical system. This choice of MCMC generator will

allow for implementing, in Section 5.8, the PLoM proposed in the work of Soize and Ghanem³⁷ for generating a large number of realizations of random variable \mathbf{H}_λ and then of \mathbf{H}^c (at convergence with respect to i).

5.7.1 | Introduction of the potential function associated with pdf $p_{\mathbf{H}_\lambda}$

For λ fixed in $C_{ad,\lambda}$, Equation (46) is rewritten, for all $\boldsymbol{\eta}$ in \mathbb{R}^v , as

$$p_{\mathbf{H}_\lambda}(\boldsymbol{\eta}) = c_0(\lambda) \rho_\lambda(\boldsymbol{\eta}), \quad (61)$$

in which function $\boldsymbol{\eta} \mapsto \rho_\lambda(\boldsymbol{\eta})$ from \mathbb{R}^v into \mathbb{R}^+ is defined by

$$\rho_\lambda(\boldsymbol{\eta}) = \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\}, \quad \mathcal{V}_\lambda(\boldsymbol{\eta}) = \psi(\boldsymbol{\eta}) + \langle \lambda, \mathbf{h}^c(\boldsymbol{\eta}) \rangle. \quad (62)$$

5.7.2 | Nonlinear ISDE for random vector \mathbf{H}_λ

Following the work of Soize and Ghanem,³⁷ we have to consider the nonlinear stochastic dissipative Hamiltonian dynamical system whose potential is \mathcal{V}_λ defined by Equation (62), which is represented by the nonlinear ISDE expressed, for $t > 0$, as

$$d[\mathbf{U}_\lambda(t)] = [\mathbf{V}_\lambda(t)] dt, \quad (63)$$

$$d[\mathbf{V}_\lambda(t)] = [L_\lambda([\mathbf{U}_\lambda(t)])] dt - \frac{1}{2} f_0 [\mathbf{V}_\lambda(t)] dt + \sqrt{f_0} d[\mathbf{W}_\lambda^{\text{wien}}(t)], \quad (64)$$

with the initial condition at $t = 0$,

$$[\mathbf{U}_\lambda(0)] = [\eta^{\text{init}}], \quad [\mathbf{V}_\lambda(0)] = [v^{\text{init}}] \quad \text{a.s.}, \quad (65)$$

in which $\{([\mathbf{U}_\lambda(t)], [\mathbf{V}_\lambda(t)]), t \in \mathbb{R}^+\}$ is a stochastic process with values in $\mathbb{M}_{v,N} \times \mathbb{M}_{v,N}$, which is made up of the independent stochastic processes $\{(\mathbf{U}_\lambda^\ell(t), \mathbf{V}_\lambda^\ell(t)), t \in \mathbb{R}^+\}$ such that $[\mathbf{U}_\lambda(t)] = [\mathbf{U}_\lambda^1(t) \dots \mathbf{U}_\lambda^N(t)]$ and $[\mathbf{V}_\lambda(t)] = [\mathbf{V}_\lambda^1(t) \dots \mathbf{V}_\lambda^N(t)]$, and where:

- (i) $f_0 > 0$ is a free parameter allowing the dissipation to be controlled in the stochastic dynamical system. This parameter is chosen such that $f_0 < 4$. The value, 4, corresponds to the critical damping rate of the linearized ISDE associated with Equations (63) and (64).
- (ii) $\{[\mathbf{W}_\lambda^{\text{wien}}(t)], t \in \mathbb{R}^+\}$ is the stochastic process, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by \mathbb{R}^+ , with values in $\mathbb{M}_{v,N}$, for which the columns of $[\mathbf{W}_\lambda^{\text{wien}}(t)]$ are N independent copies of the \mathbb{R}^v -valued normalized Wiener process whose matrix-valued autocorrelation function is $\min(t, t') [I_v]$. The dependence on λ is such that, if λ and λ' are two distinct values in $C_{ad,\lambda}$, then the two stochastic processes $\{[\mathbf{W}_\lambda^{\text{wien}}(t)], t \in \mathbb{R}^+\}$ and $\{[\mathbf{W}_{\lambda'}^{\text{wien}}(t)], t \in \mathbb{R}^+\}$ are independent.
- (iii) $[u] \mapsto [L_\lambda([u])]$ is a nonlinear mapping from $\mathbb{M}_{v,N}$ into $\mathbb{M}_{v,N}$, expressed in the following form as negative of the gradient of the potential \mathcal{V}_λ

$$[L_\lambda([u])]_{\alpha\ell} = -\frac{\partial}{\partial u_\alpha^\ell} \mathcal{V}_\lambda(\mathbf{u}^\ell), \quad \alpha = 1, \dots, v, \quad \ell = 1, \dots, N, \quad (66)$$

in which $[u] = [\mathbf{u}^1 \dots \mathbf{u}^N] \in \mathbb{M}_{v,N}$ with $\mathbf{u}^\ell = (u_1^\ell, \dots, u_v^\ell) \in \mathbb{R}^v$. For ℓ fixed in $\{1, \dots, N\}$, the Hamiltonian of the associated conservative homogeneous dynamical system related to stochastic process $\{(\mathbf{U}^\ell(t), \mathbf{V}^\ell(t)), t \in \mathbb{R}^+\}$ is thus written as $\mathbb{H}_\lambda(\mathbf{u}^\ell, \mathbf{v}^\ell) = \frac{1}{2} \|\mathbf{v}^\ell\|^2 + \mathcal{V}_\lambda(\mathbf{u}^\ell)$.

- (iv) $[\eta^{\text{init}}]$ and $[v^{\text{init}}]$ are given matrices in $\mathbb{M}_{v,N}$ that will be defined in Section 5.7.5 as a function of iteration i of the algorithm given in Section 5.6.

5.7.3 | Existence and uniqueness of a stationary solution associated with the invariant measure

Using Theorems 4 to 7 in the work of Soize,^{54(pp211-216)} it can be proven, that for $\lambda \in C_{ad,\lambda}$, Equations (63) to (65) admits the unique invariant measure,

$$\bigotimes_{\ell=1}^N \{p_{\mathbf{H}_\lambda}(\mathbf{u}^\ell) p_G(\mathbf{v}^\ell) d\mathbf{u}^\ell d\mathbf{v}^\ell\}, \quad (67)$$

in which $p_{\mathbf{H}_\lambda}$ is defined by Equations (61) and (62), and where the pdf p_G is written as $p_G(\mathbf{v}^\ell) = (2\pi)^{-v/2} \exp\{-1/2 \|\mathbf{v}^\ell\|^2\}$. In addition, these theorems allow for proving that Equations (63) to (65) has a unique solution $\{([\mathbf{U}_\lambda(t)], [\mathbf{V}_\lambda(t)]), t \in \mathbb{R}^+\}$,

which is a second-order diffusion stochastic process that is asymptotic for $t \rightarrow +\infty$ to the stationary stochastic process $\{([\mathbf{U}_\lambda^{\text{st}}(t_{\text{st}})], [\mathbf{V}_\lambda^{\text{st}}(t_{\text{st}})]), t_{\text{st}} \in \mathbb{R}^+\}$ for the right-shift semigroup on \mathbb{R}^+ . For all fixed t_{st} , the joint probability distribution of the random matrices $[\mathbf{U}_\lambda^{\text{st}}(t_{\text{st}})]$ and $[\mathbf{V}_\lambda^{\text{st}}(t_{\text{st}})]$ is the invariant measure defined by Equation (67). The probability distribution of random matrix $[\mathbf{U}_\lambda^{\text{st}}(t_{\text{st}})]$ is

$$\bigotimes_{\ell=1}^N p_{\mathbf{H}_\lambda}(\mathbf{u}^\ell) d\mathbf{u}^\ell, \quad (68)$$

that is to say, is the probability distribution $p_{[\mathbf{H}_\lambda]}([u]) d[u]$ with $d[u] = \bigotimes_{\ell=1}^N d\mathbf{u}^\ell$ of the random matrix $[\mathbf{H}_\lambda]$ with values in $\mathbb{M}_{v,N}$. The columns $\mathbf{H}_\lambda^1, \dots, \mathbf{H}_\lambda^N$ of $[\mathbf{H}_\lambda]$ are N independent copies of random vector \mathbf{H}_λ with values in \mathbb{R}^v whose pdf is $p_{\mathbf{H}_\lambda}$ defined by Equations (61) and (62). It can then be deduced that for any fixed t_{st} ,

$$[\mathbf{H}_\lambda] = \left[\mathbf{U}_\lambda^{\text{st}}(t_{\text{st}}) \right] = \lim_{t \rightarrow +\infty} [\mathbf{U}_\lambda(t)]. \quad (69)$$

Equation (69) means that, for $f_0 > 0$, there exists $t_\lambda^{\min} > 0$, such that $\{[\mathbf{U}_\lambda(t)], t > t_\lambda^{\min}\}$ is a stationary stochastic process that is a copy of stochastic process $\{[\mathbf{U}_\lambda^{\text{st}}(t_{\text{st}})], t_{\text{st}} \in \mathbb{R}^+\}$. As previously explained, f_0 allows for controlling the transient response generated by the initial conditions for quickly reaching the stationary regime. Finally, we introduce the Gaussian random matrices $[\mathbf{V}_\lambda]$ with values in $\mathbb{M}_{v,N}$ such that

$$[\mathbf{V}_\lambda] = \left[\mathbf{V}_\lambda^{\text{st}}(t_{\text{st}}) \right] = \lim_{t \rightarrow +\infty} [\mathbf{V}_\lambda(t)]. \quad (70)$$

5.7.4 | Expression of the mapping $[L_\lambda]$

In this subsection, a general expression for the mapping $[L_\lambda]$ is developed. In addition, specializations to particular common examples are derived.

Using Equations (62) and (66), it can be proven that, for $\alpha = 1, \dots, v$, for $\ell = 1, \dots, N$, and for $[u] = [\mathbf{u}^1 \dots \mathbf{u}^N]$ in $\mathbb{M}_{v,N}$, we have

$$[L_\lambda([u])]_{\alpha\ell} = \left\{ \frac{1}{\rho(\mathbf{u}^\ell) \mathbf{u}^\ell} \nabla \rho(\mathbf{u}^\ell) - [\mu]^{1/2} [\Phi]^T [D_{\mathbf{x}} \mathbf{g}^c(\hat{\mathbf{x}} + [\Phi][\mu]^{1/2} \mathbf{u}^\ell)] [\Psi] \lambda \right\}_\alpha, \quad (71)$$

in which ρ is defined by Equation (23), where the gradient of ρ is such that

$$\nabla_{\mathbf{u}^\ell} \rho(\mathbf{u}^\ell) = \frac{1}{\hat{s}_v^2} \frac{1}{N} \sum_{j=1}^N \left(\frac{\hat{s}_v}{s_v} \boldsymbol{\eta}^j - \mathbf{u}^\ell \right) \exp \left\{ -\frac{1}{2\hat{s}_v^2} \left\| \frac{\hat{s}_v}{s_v} \boldsymbol{\eta}^j - \mathbf{u}^\ell \right\|^2 \right\}, \quad (72)$$

and where $[D_{\mathbf{x}} \mathbf{g}^c(\mathbf{x})] \in \mathbb{M}_{n,m_c}$ is the derivative at point $\mathbf{x} \in \mathbb{R}^n$ of the mapping $\mathbf{x} \mapsto \mathbf{g}^c(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_{m_c}(\mathbf{x}))$ from \mathbb{R}^n into \mathbb{R}^{m_c} such that

$$[D_{\mathbf{x}} \mathbf{g}^c(\mathbf{x})]_{ki} = \frac{\partial g_i^c(\mathbf{x})}{\partial x_k}, \quad k = 1, \dots, n, \quad i = 1, \dots, m_c.$$

(1) Case of the example presented in Section 4.1

Let us consider the example presented in Section 4.1. We have $\mathcal{J}_{\text{mom}} = \{k_1, \dots, k_{\kappa_{\text{mom}}}\}$, $\kappa_{\text{mom}} \leq n_q$, and $m_c = 2\kappa_{\text{mom}} \leq 2n_q$. Therefore,

$$\boldsymbol{\beta}^c = \left(\mathbb{m}_{k_1}^{\text{exp}}, \dots, \mathbb{m}_{k_{\kappa_{\text{mom}}}}^{\text{exp}}, \mathbb{r}_{k_1}^{\text{exp}}, \dots, \mathbb{r}_{k_{\kappa_{\text{mom}}}}^{\text{exp}} \right) \in \mathbb{R}^{m_c}, \quad (73)$$

and, for $k = 1, \dots, n$ and $\mathbf{v} = [\Psi] \lambda$ in \mathbb{R}^{m_c} , we have

$$\{[D_{\mathbf{x}} \mathbf{g}^c(\mathbf{x})] \mathbf{v}\}_k = \sum_{\kappa=1}^{\kappa_{\text{mom}}} \delta_{kk_\kappa} (v_\kappa + 2x_{k_\kappa} v_{\kappa+\kappa_{\text{mom}}}).$$

Let us consider the block expression of $[\Phi] \in \mathbb{M}_{n,v}$ such that $[\Phi]^T = [[\Phi_q]^T [\Phi_w]^T]$ in which $[\Phi_q] \in \mathbb{M}_{n_q,v}$. For $\kappa = 1, \dots, \kappa_{\text{mom}}$, we define $q_{k_\kappa}(\mathbf{u}^\ell)$ such that

$$q_{k_\kappa}(\mathbf{u}^\ell) = \hat{q}_{k_\kappa} + \{[\Phi_q][\mu]^{1/2} \mathbf{u}^\ell\}_{k_\kappa}. \quad (74)$$

Consequently, for $\alpha = 1, \dots, v$, for $\ell = 1, \dots, N$, and for $\mathbf{v} = [\Psi] \lambda$, Equation (71) yields

$$[L_\lambda([u])]_{\alpha\ell} = \left\{ \frac{1}{\rho(\mathbf{u}^\ell)_{\mathbf{u}^\ell}} \nabla \rho(\mathbf{u}^\ell) \right\}_\alpha + [L_\lambda^{c,1}([u])]_{\alpha\ell}. \quad (75)$$

$$[L_\lambda^{c,1}([u])]_{\alpha\ell} = - \sum_{\kappa=1}^{\kappa_{\text{mom}}} \mu_\alpha^{1/2} [\Phi_q]_{k_\kappa \alpha} (v_\kappa + 2 q_{k_\kappa}(\mathbf{u}^\ell) v_{\kappa+\kappa_{\text{mom}}}). \quad (76)$$

For facilitating the numerical implementation and enhancing computational efficiency, the right-hand side of Equation (76) can be rewritten in a matrix form as follows. Let $\underline{\mathbf{q}} = (\underline{q}_1, \dots, \underline{q}_{\kappa_{\text{mom}}})$ be the vector in $\mathbb{R}^{\kappa_{\text{mom}}}$ and let $[\tilde{\Phi}_q]$ be the matrix in $\mathbb{M}_{\kappa_{\text{mom}}, v}$ such that, for $\kappa = 1, \dots, \kappa_{\text{mom}}$ and for all $\alpha = 1, \dots, v$, we have

$$\underline{q}_\kappa = \hat{q}_{k_\kappa}, \quad [\tilde{\Phi}_q]_{\kappa\alpha} = [\Phi_q]_{k_\kappa \alpha}. \quad (77)$$

Let $\tilde{\mathbf{q}}(\mathbf{u}^\ell) = (\tilde{q}_1(\mathbf{u}^\ell), \dots, \tilde{q}_{\kappa_{\text{mom}}}(\mathbf{u}^\ell))$ be the vector in $\mathbb{R}^{\kappa_{\text{mom}}}$ such that $\tilde{q}_\kappa(\mathbf{u}^\ell) = q_{k_\kappa}(\mathbf{u}^\ell)$, which can be written as

$$\tilde{\mathbf{q}}(\mathbf{u}^\ell) = \underline{\mathbf{q}} + [\tilde{\Phi}_q][\mu]^{1/2} \mathbf{u}^\ell. \quad (78)$$

Let $[\tilde{q}(\mathbf{u}^\ell)]$ be the diagonal matrix in $\mathbb{M}_{\kappa_{\text{mom}}}$ such that $[\tilde{q}(\mathbf{u}^\ell)]_{\kappa\kappa'} = \delta_{\kappa\kappa'} \tilde{q}_\kappa(\mathbf{u}^\ell)$ in which $\tilde{q}_\kappa(\mathbf{u}^\ell)$ is component κ of vector $\tilde{\mathbf{q}}(\mathbf{u}^\ell)$ defined by Equation (78). Let us consider the block writing of $[\Psi] \in \mathbb{M}_{m_c, v_c}$ such that $[\Psi]^T = [[\Psi_m]]^T [\Psi_R]^T$ in which $[\Psi_m]$ and $[\Psi_R]$ are matrices that belong to $\mathbb{M}_{\kappa_{\text{mom}}, v_c}$. Using the above notations, Equation (76) can be rewritten as

$$[L_\lambda^{c,1}([u])]_{\alpha\ell} = \left\{ -[\mu]^{1/2} [\tilde{\Phi}_q]^T ([\Psi_m] + 2[\tilde{q}(\mathbf{u}^\ell)][\Psi_R]) \lambda \right\}_\alpha. \quad (79)$$

(2) Case of the example presented in Section 4.2

Let us now consider Equation (15) of the example presented in Section 4.2 with $m_c = n_q$. We then have $\mathbf{g}^c(\mathbf{x}) = \mathbf{q} - [\underline{B}] \mathbf{w}$ with $\mathbf{x} = (\mathbf{q}, \mathbf{w}) \in \mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$, which can be rewritten as $\mathbf{g}^c(\mathbf{x}) = [G^c] \mathbf{x}$ with $[G^c] = [[I_{n_q}] | -[\underline{B}]] \in \mathbb{M}_{n_q, n}$ with $n = n_q + n_w$. Consequently, we have $[D_x \mathbf{g}^c(\mathbf{x})] = [G^c]^T \in \mathbb{M}_{n, n_q}$ and for $\alpha = 1, \dots, v$, for $\ell = 1, \dots, N$, and for $\mathbf{v} = [\Psi] \lambda$, Equation (71) yields

$$[L_\lambda([u])]_{\alpha\ell} = \left\{ \frac{1}{\rho(\mathbf{u}^\ell)_{\mathbf{u}^\ell}} \nabla \rho(\mathbf{u}^\ell) \right\}_\alpha + [L_\lambda^{c,2}([u])]_{\alpha\ell}. \quad (80)$$

$$[L_\lambda^{c,2}([u])]_{\alpha\ell} = \left\{ -[\mu]^{1/2} [\Phi]^T [G^c]^T [\Psi] \lambda \right\}_\alpha. \quad (81)$$

It should be noted that $[L_\lambda^{c,2}([u])]$ depends only on λ and is independent of $[u]$.

(3) Case of the example presented in Appendix A.1

Finally, we consider the example presented in Appendix A.1 for which $\mathcal{J}_{\text{mom}} = \{k_1, \dots, k_{\kappa_{\text{mom}}}\}$, $\kappa_{\text{mom}} \leq n_q$, and $m_c = 2\kappa_{\text{mom}} + n_q \leq 3n_q$. Therefore,

$$\boldsymbol{\beta}^c = \left(\mathbb{m}_{k_1}^{\text{exp}}, \dots, \mathbb{m}_{k_{\kappa_{\text{mom}}}}^{\text{exp}}, \mathbb{r}_{k_1}^{\text{exp}}, \dots, \mathbb{r}_{k_{\kappa_{\text{mom}}}}^{\text{exp}}, \beta_1^{\text{const}}, \dots, \beta_{n_q}^{\text{const}} \right) \in \mathbb{R}^{m_c}, \quad (82)$$

and using Equations (75), (76), (80), and (81), we have, for $\alpha = 1, \dots, v$, for $\ell = 1, \dots, N$, and for $\lambda \in \mathbb{R}^{v_c}$,

$$[L_\lambda([u])]_{\alpha\ell} = \left\{ \frac{1}{\rho(\mathbf{u}^\ell)_{\mathbf{u}^\ell}} \nabla \rho(\mathbf{u}^\ell) \right\}_\alpha + [L_\lambda^c([u])]_{\alpha\ell}, \quad (83)$$

$$[L_\lambda^c([u])] = [L_\lambda^{c,1}([u])] + [L_\lambda^{c,2}([u])]. \quad (84)$$

5.7.5 | Expression of the initial conditions for the MCMC algorithm

In Equation (65), the initial conditions $[\eta^{\text{init}}]$ and $[\nu^{\text{init}}]$ are constructed in $\mathbb{M}_{v, N}$ as a function of iteration ι of the algorithm presented in Section 5.6 in order to calculate the sequence $\{\lambda^\iota\}_{\iota \geq 2}$ from the initial value λ^1 as follows:

- λ^1 is computed (see Equation (55)) by $\lambda^1 = -[\Gamma''(\lambda^0)]^{-1} \nabla \Gamma(\lambda^0)$ with $\lambda^0 = \mathbf{0}$, using Equation (58) as explained in Section 5.6.

- λ^2 is calculated using Equation (55), in which $[\Gamma''(\lambda^1)]$ and $\nabla\Gamma(\lambda^1)$ are computed with Equations (51) and (52) in which $E\{\mathbf{h}^c(\mathbf{H}_{\lambda^1})\}$ and $[\text{cov}\{\mathbf{h}^c(\mathbf{H}_{\lambda^1})\}]$ are estimated with the MCMC algorithm for which the initial conditions at $t = 0$ are chosen as $[\eta^{\text{init}}] = [\eta^{\text{init},1}]$ and $[v^{\text{init}}] = [v^{\text{init},1}]$ with

$$[\eta^{\text{init},1}] = [\boldsymbol{\eta}^1 \dots \boldsymbol{\eta}^N] \in \mathbb{M}_{v,N}, \quad (85)$$

and where the N columns of $[v^{\text{init},1}]$ are N independent realizations of a Gaussian, centered, \mathbb{R}^v -valued random variable for which its covariance matrix is $[I_v]$. The MCMC generator (see Section 5.8) allows for generating n_{MC} realizations $[\eta_{\lambda^1}^1], \dots, [\eta_{\lambda^1}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{H}_{\lambda^1}]$ and $[v_{\lambda^1}^1], \dots, [v_{\lambda^1}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{V}_{\lambda^1}]$ (see Equations (69) and (70)).

- Knowing λ^i at iteration $i \geq 1$, the vector λ^{i+1} is calculated using Equation (55) in which $[\Gamma''(\lambda^i)]$ and $\nabla\Gamma(\lambda^i)$ are computed with Equations (51) and (52) using the MCMC algorithm for which the initial conditions at $t = 0$ are chosen as $[\eta^{\text{init}}] = [\eta^{\text{init},i}]$ and $[v^{\text{init}}] = [v^{\text{init},i}]$ with

$$[\eta^{\text{init},i}] = \left[\eta_{\lambda^{i-1}}^{n_{\text{MC}}} \right], \quad [v^{\text{init},i}] = \left[v_{\lambda^{i-1}}^{n_{\text{MC}}} \right]. \quad (86)$$

5.8 | Computing the additional realizations of \mathbf{H}_{λ^i} using the probabilistic learning on manifolds

For given $\lambda = \lambda^i$ in \mathbb{R}^{v_c} in which λ^i comes from the iteration algorithm detailed in Section 5.6, and for given initial conditions $[\eta^{\text{init},i}]$ and $[v^{\text{init},i}]$ defined by Equation (86), n_{MC} realizations $[\eta_{\lambda^i}^1], \dots, [\eta_{\lambda^i}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{H}_{\lambda^i}]$ and $[v_{\lambda^i}^1], \dots, [v_{\lambda^i}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{V}_{\lambda^i}]$ have to be computed using the MCMC generator defined by Equations (63) to (65) in which $\lambda = \lambda^i$. However, as it has previously been explained, N is assumed to be small and consequently, the PLoM presented in the work of Soize and Ghanem³⁷ is used. This algorithm consists in projecting the nonlinear ISDE defined by Equations (63) to (65) on the diffusion-maps basis. This approach allows for preserving the concentration of the invariant measure $p_{\mathbf{H}_{\lambda^i}}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ and to avoid the scattering of the realizations computed with the MCMC generator. In order to simplify the reading of this paper, we briefly summarized the algorithm that is proposed in the work of Soize and Ghanem³⁷ in adapting the notations to the problem under consideration.

5.8.1 | Construction of the diffusion-maps basis

The diffusion-maps basis is independent of λ^i , only depends on the initial data set $\{\boldsymbol{\eta}^1, \dots, \boldsymbol{\eta}^N\}$, and is represented by the matrix

$$[g] = [\mathbf{g}^1 \dots \mathbf{g}^m] \in \mathbb{M}_{N,m} \quad \text{with} \quad 1 < m \leq N. \quad (87)$$

For $m = N$, it is an algebraic basis of vector space \mathbb{R}^N . This basis is constructed using the diffusion maps proposed in the work of Coifman et al.⁵⁵ Let $[\mathbb{K}]$ be the symmetric $(N \times N)$ real matrix such that $[\mathbb{K}]_{jj'} = \exp(-\frac{1}{4\varepsilon_{\text{diff}}} \|\boldsymbol{\eta}^j - \boldsymbol{\eta}^{j'}\|)$ that depends on a real smoothing parameter $\varepsilon_{\text{diff}} > 0$. Let $[\mathbb{P}]$ be the transition matrix in \mathbb{M}_N of a Markov chain such that $[\mathbb{P}] = [\mathbb{b}]^{-1} [\mathbb{K}]$, in which $[\mathbb{b}]$ is the positive-definite diagonal real matrix such that $[\mathbb{b}]_{ij} = \delta_{ij} \sum_{j'=1}^N [\mathbb{K}]_{jj'}$. For m fixed in $\{1, \dots, N\}$, let $\mathbf{g}^1, \dots, \mathbf{g}^m$ be the right eigenvectors in \mathbb{R}^N of matrix $[\mathbb{P}]$ such that $[\mathbb{P}]\mathbf{g}^\alpha = \Lambda_\alpha \mathbf{g}^\alpha$ in which the eigenvalues are sorted such that $1 = \Lambda_1 > \Lambda_2 > \dots > \Lambda_m$. It can easily be proven that the eigenvalues are positive and the largest is Λ_1 for which all the components of the corresponding eigenvector \mathbf{g}^1 are equal. Considering the normalization $[g]^T [\mathbb{b}] [g] = [I_m]$, the right-eigenvalue problem of the nonsymmetric matrix $[\mathbb{P}]$ can then be done solving the eigenvalue problem $[\mathbb{b}]^{-1/2} [\mathbb{K}] [\mathbb{b}]^{-1/2} \xi^\alpha = \Lambda_\alpha \xi^\alpha$ related to a positive-definite symmetric real matrix with the orthonormalization $\langle \xi^\alpha, \xi^\beta \rangle = \delta_{\alpha\beta}$. Therefore, \mathbf{g}^α can be deduced from ξ^α by $\mathbf{g}^\alpha = [\mathbb{b}]^{-1/2} \xi^\alpha$.

The construction introduces two hyperparameters: the dimension $m \leq N$ and the smoothing parameter $\varepsilon_{\text{diff}} > 0$. An algorithm is proposed in the work of Soize et al⁴⁰ for estimating their values. Most of the time, m and $\varepsilon_{\text{diff}}$ can be chosen as follows. Let $\varepsilon_{\text{diff}} \mapsto \hat{m}(\varepsilon_{\text{diff}})$ be the function from $]0, +\infty[$ into the set $\mathbb{N} = \{0, 1, 2, \dots\}$ of all the integers such that

$$\hat{m}(\varepsilon_{\text{diff}}) = \arg \min_{\alpha \mid \alpha \geq 3} \left\{ \frac{\Lambda_\alpha(\varepsilon_{\text{diff}})}{\Lambda_2(\varepsilon_{\text{diff}})} < 0.1 \right\}. \quad (88)$$

If function \hat{m} is a decreasing function of $\varepsilon_{\text{diff}}$ in the broad sense (if not, see the work of Soize et al⁴⁰), then the optimal value $\varepsilon_{\text{diff}}^{\text{opt}}$ of $\varepsilon_{\text{diff}}$ can be chosen as the smallest value of the integer $\hat{m}(\varepsilon_{\text{diff}}^{\text{opt}})$ such that

$$\left\{ \hat{m}\left(\varepsilon_{\text{diff}}^{\text{opt}}\right) < \hat{m}(\varepsilon_{\text{diff}}), \forall \varepsilon_{\text{diff}} \in [0, \varepsilon_{\text{diff}}^{\text{opt}}] \right\} \cap \left\{ \hat{m}\left(\varepsilon_{\text{diff}}^{\text{opt}}\right) = \hat{m}(\varepsilon_{\text{diff}}), \forall \varepsilon_{\text{diff}} \in [\varepsilon_{\text{diff}}^{\text{opt}}, 1.5\varepsilon_{\text{diff}}^{\text{opt}}] \right\}. \quad (89)$$

The corresponding optimal value m^{opt} of m is then given by $m^{\text{opt}} = \hat{m}(\varepsilon_{\text{diff}}^{\text{opt}})$. Random vector \mathbf{H} is centered, but due to the constraints, \mathbf{X}^c can be not centered. Consequently, vector \mathbf{g}^1 associated with Λ_1 has to be kept in the basis.

5.8.2 | Reduced-order representation of random matrix $[\mathbf{H}_{\lambda'}]$

The diffusion-maps basis represented by matrix $[g] \in \mathbb{M}_{N,m}$ spans a subspace of \mathbb{R}^N that characterizes the local geometry structure of data set $\{\boldsymbol{\eta}^j, j = 1, \dots, N\}$. The reduced-order representation is obtained by projecting each column of the $\mathbb{M}_{N,v}$ -valued random matrix $[\mathbf{H}_{\lambda'}]^T$ on the subspace of \mathbb{R}^N , spanned by $\{\mathbf{g}^1, \dots, \mathbf{g}^m\}$. Let $[\mathbf{Z}_{\lambda'}]$ be the random matrix with values in $\mathbb{M}_{v,m}$ such that

$$[\mathbf{H}_{\lambda'}] = [\mathbf{Z}_{\lambda'}][g]^T. \quad (90)$$

Because the matrix $[g]^T[g] \in \mathbb{M}_m^+$ is invertible, the least-square approximation of $[\mathbf{Z}_{\lambda'}]$ can be written as

$$[\mathbf{Z}_{\lambda'}] = [\mathbf{H}_{\lambda'}][a], \quad (91)$$

in which the matrix $[a]$ is defined by

$$[a] = [g]([g]^T[g])^{-1} \in \mathbb{M}_{N,m}. \quad (92)$$

It should be noted that the least-square approximation defined by Equation (91) is only used for calculating the initial conditions for the reduced-order nonlinear ISDE.

5.8.3 | Generation of realizations of random matrix $[\mathbf{H}_{\lambda'}]$

The objective is to compute n_{MC} independent realizations $[\eta_{\lambda'}^1], \dots, [\eta_{\lambda'}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{H}_{\lambda'}]$ in which n_{MC} is assumed to be independent of λ' . The reduced-order nonlinear ISDE is constructed as the projection of the nonlinear ISDE defined by Equations (63) to (65) on the diffusion-maps basis represented by matrix $[g] \in \mathbb{M}_{N,m}$. We therefore define the $\mathbb{M}_{v,m} \times \mathbb{M}_{v,m}$ -valued stochastic process $\{([\mathcal{Z}_{\lambda'}(t)], [\mathcal{Y}_{\lambda'}(t)]), t \geq 0\}$ such that, for all $t \geq 0$,

$$[\mathbf{U}_{\lambda'}(t)] = [\mathcal{Z}_{\lambda'}(t)][g]^T, \quad [\mathbf{V}_{\lambda'}(t)] = [\mathcal{Y}_{\lambda'}(t)][g]^T. \quad (93)$$

The projection of Equations (63) to (65) yields

$$d[\mathcal{Z}_{\lambda'}(t)] = [\mathcal{Y}_{\lambda'}(t)]dt, \quad (94)$$

$$d[\mathcal{Y}_{\lambda'}(t)] = [\mathcal{L}_{\lambda'}([\mathcal{Z}_{\lambda'}(t)])]dt - \frac{1}{2}f_0[\mathcal{Y}_{\lambda'}(t)]dt + \sqrt{f_0}d\left[\mathcal{W}_{\lambda'}^{\text{wien}}(t)\right], \quad (95)$$

with the projected initial conditions (see Equations (86) and (91)) at $t = 0$,

$$[\mathcal{Z}_{\lambda'}(0)] = \left[\eta_{\lambda'}^{n_{\text{MC}}}\right][a], \quad [\mathcal{Y}_{\lambda'}(0)] = \left[v_{\lambda'}^{n_{\text{MC}}}\right][a], \text{ a.s.,} \quad (96)$$

in which the random matrix $[\mathcal{L}_{\lambda'}([\mathcal{Z}_{\lambda'}(t)])]$ with values in $\mathbb{M}_{v,m}$ is written as

$$[\mathcal{L}_{\lambda'}([\mathcal{Z}_{\lambda'}(t)])] = [L_{\lambda'}([\mathcal{Z}_{\lambda'}(t)][g]^T)][a], \quad (97)$$

$$\left[\mathcal{W}_{\lambda'}^{\text{wien}}(t)\right] = \left[\mathbf{W}_{\lambda'}^{\text{wien}}(t)\right][a]. \quad (98)$$

Using the results presented in Section 5.7.3, it can be deduced that random matrix $[\mathbf{Z}_{\lambda'}] = \lim_{t \rightarrow +\infty} [\mathcal{Z}_{\lambda'}(t)]$ in probability distribution. We can then compute n_{MC} independent realizations $[z_{\lambda'}^1], \dots, [z_{\lambda'}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{Z}_{\lambda'}]$, using the

stationary regime of stochastic process $\{[\mathcal{Z}_{\lambda'}(t)], t > 0\}$ that is the solution of Equations (94) to (98). The realizations of $[\mathbf{H}_{\lambda'}]$ are then obtained using Equation (90),

$$\left[\eta_{\lambda'}^{\ell} \right] = \left[z_{\lambda'}^{\ell} \right] [g]^T \in \mathbb{M}_{v,N}, \quad \ell = 1, \dots, n_{MC}. \quad (99)$$

The Störmer-Verlet algorithm is used for solving the reduced-order nonlinear ISDE defined by Equations (94) to (98) as proposed in the work of Soize and Ghanem.^{37,44} This algorithm is summarized in Appendix B.

5.8.4 | Deducing the realization of random vector $\mathbf{H}_{\lambda'}$

Let be $M = n_{MC} \times N$. Once the realizations $\{\eta_{\lambda'}^{\ell}, \ell = 1, \dots, n_{MC}\}$ of the $\mathbb{M}_{v,N}$ -valued random variable $[\mathbf{H}_{\lambda'}]$ have been computed using Equation (99), the $M \gg N$ realizations $\{\eta_{\lambda'}^{\ell'}, \ell' = 1, \dots, M\}$ of the \mathbb{R}^v -valued random variable $\mathbf{H}_{\lambda'}$ are deduced by reshaping the matrices $\{\eta_{\lambda'}^{\ell}, \ell = 1, \dots, n_{MC}\}$.

5.9 | Computing the additional realizations of random vector \mathbf{X}^c

As soon as the error function defined by Equation (59) is less than a given small tolerance (that is to say, considering the iteration number i such that $\text{err}(i) \leq \text{tol}$), the iteration algorithm detailed in Section 5.6 is converged and the solution λ^{sol} defined by Equation (54) is such that $\lambda^{\text{sol}} \simeq \lambda'$. The realizations $\{\eta_{\lambda'}^{\ell'}, \ell' = 1, \dots, M\}$ of the random vector \mathbf{H}^c constructed in Section 5.8.4, which correspond to the last iteration λ' , are rewritten as $\{\eta^{c,\ell'}, \ell' = 1, \dots, M\}$ (that is to say, $\eta^{c,\ell'} = \eta_{\lambda'}^{\ell'}$). Using Equation (28), the M realizations $\{\mathbf{x}^{c,\ell'}, \ell' = 1, \dots, M\}$ of random vector \mathbf{X}^c are such that

$$\mathbf{x}^{c,\ell'} = \underline{\mathbf{x}} + [\Phi][\mu]^{1/2} \eta^{c,\ell'}, \quad \ell' = 1, \dots, M. \quad (100)$$

5.10 | Convergence analysis of the probabilistic learning with respect dimension N of initial data set

Let us assumed that ε , which has been introduced in Equation (18), is fixed independently of the value of N . Except for the case for which ε is chosen very small, in general, $v = v(N)$ depends on N for N belonging to the ordered subset $\mathcal{J}_N = \{N_{\min}, \dots, N_{\max}\} \subset \mathbb{N}$ of integers, with $0 < N_{\min} < N_{\max}$. The value of N_{\max} corresponds to the maximum value that is available for constructing the initial data set $D_{N_{\max}}$. Consequently, the random variable $\mathbf{X}^{c,v(N)}$ depends on $N \in \mathcal{J}_N$ and will simply be denoted as $\mathbf{X}^{c,N}$. A question related to the proposed PLoM is the convergence of the probability distribution $P_{\mathbf{X}^{c,N}}$ when N goes through the ordered set \mathcal{J}_N . This means that, for a given tolerance, when N goes from N_{\min} to N_{\max} , the probability distribution $P_{\mathbf{X}^{c,N}}$ must converge. If it is not the case, this means that the learning is not convergent and then N_{\max} has to be increased. Therefore, additional calculations have to be carried out with the computational models for constituting a new bigger training set.

(i) Because the mean-square convergence implies the convergence in probability distribution, we can study the convergence of the learning in analyzing the function

$$N \mapsto \text{conv}_{L^2,\mathbf{X}^c}(N) = E\{\|\mathbf{X}^{c,N}\|^2\}. \quad (101)$$

In fact, the limit is unknown and a Cauchy sequence could be studied (Hilbert space $L^2(\Theta, \mathbb{R}^n)$ is a complete vector space). In practice, because set \mathcal{J}_N is not sufficiently big for using such a Cauchy criterion, we can check that $N \mapsto \text{conv}_{L^2,\mathbf{X}^c}(N)$ is flat on $\{N_f, \dots, N_{\max}\}$ for $N_{\max} < N_f < N_{\max}$.

(ii) Because the contribution of $\mathbf{Q}^{c,N}$ and $\mathbf{W}^{c,N}$, such that $\mathbf{X}^{c,N} = (\mathbf{Q}^{c,N}, \mathbf{W}^{c,N})$, can be very different in the value of $\text{conv}_{L^2,\mathbf{X}^c}(N)$, the definition of $\text{conv}_{L^2,\mathbf{X}^c}(N)$ given by Equation (101) can be replaced by the following one:

$$N \mapsto \text{conv}_{L^2,\mathbf{X}^c}(N) = \frac{1}{\sqrt{2}} \left\{ \frac{E\{\|\mathbf{Q}^{c,N}\|^2\}}{E\{\|\mathbf{Q}^{N_{\max}}\|^2\}} + \frac{E\{\|\mathbf{W}^{c,N}\|^2\}}{E\{\|\mathbf{W}^{N_{\max}}\|^2\}} \right\}. \quad (102)$$

(iii) Sometimes, for a given \mathcal{J}_N , the criteria defined by Equations (101) or (102) are not sufficiently sensitive when N goes through \mathcal{J}_N . In such a case, and if the constraints are only applied to the QoI, indicators of the convergence of the

learning can be analyzed by studying, for a given subset of components k , the graphs of the functions

$$N \mapsto \mathbb{m}_{\mathbf{Q}_k^c}(N), \quad N \mapsto \mathbb{s}_{\mathbf{Q}_k^c}(N), \quad N \mapsto \{q \mapsto p_{\mathbf{Q}_k^c}(q; N)\}, \quad (103)$$

of the mean value, the standard deviation, and the pdf of \mathbf{Q}_k^c , which are estimated by using the M additional realizations generated with the PLoM.

5.11 | Remarks concerning the convergence of the iteration algorithm in the framework of the existence of a solution

If there exists a unique solution (see the comments formulated in Section 5.4-(iv)), then, for all $i \geq 1$, $[\Gamma''(\lambda^i)]$ is a positive-definite matrix, and reciprocally, let us assume that this hypothesis holds. Because the initial point is correctly chosen as explained in Section 5.6, then the iteration algorithm defined by Equation (55) is convergent and it is not necessary to introduce an underrelaxation of the type $\lambda^{i+1} = \lambda^i - \beta_{\text{relax}} [\Gamma''(\lambda^i)]^{-1} \nabla \Gamma(\lambda^i)$ with $\beta_{\text{relax}} \in]0, 1[$. If, for a given iteration number i , the rank of positive matrix $[\Gamma''(\lambda^i)]$ becomes less than v_c , then solution λ^{sol} cannot be constructed in $C_{\text{ad}, \lambda}$, which means that there is no solution. This difficulty is related to the value of v_c and cannot be compensated using an underrelaxation. The introduction of an underrelaxation does not modify the fact that there will exist $i' > i$ (before reaching the convergence) such that the rank of $[\Gamma''(\lambda^{i'})]$ is less than v_c (all the numerical experiments that have been done have confirmed this analysis). In fact, the existence of a solution is related to the number v_c of projected constraint equations defined by Equation (33), which is the projection by $[\Psi] \in \mathbb{M}_{m_c, v_c}$ of the constraint equations defined by the Equation (10). If τ_c is too small, then v_c is too big compared with the stochastic dimension v_s that have been introduced in Section 5.4-(iv). This means that the number v_c of projected constraint equations is too large with respect to the number v_s of “stochastic degrees of freedom,” and consequently, the projected constraint equations defined by Equation (33) cannot be exactly satisfied. An effective method to solve such a nonconvergence problem is to slightly increase the value τ_c , which makes it possible to obtain a slightly smaller rank v_c and thus to reduce the number of projected constraint equations (all the numerical experiments that have been conducted validate this analysis).

Assumptions for the applicability of the proposed method. The proposed method is general. The given constraints for performing the PLoM are not limited to second-order moments but can be expressed as the mathematical expectation of any measurable mapping of the control parameters and the QoI. For example, marginal probability densities can be imposed as shown in Appendix A. There is obviously a domain of validity, which is not due to the method proposed itself, but which is due to the coherence of the given constraints with respect to the initial data set (that is used for the probabilistic learning). If the introduced constraints are not consistent with the initial data set, then the optimization problem will have no solution. The existence of this solution is therefore conditioned by hypotheses that are discussed above and in Section 5.4-(iv).

6 | APPLICATION (AP1)

In this section, an application is presented and is used for performing the validation of the methodology and the algorithms presented. All the random variables are defined on probability space $(\Theta, \mathcal{T}, \mathcal{P})$. This application that will be referenced as (AP1) is sufficiently simple in order to be easily reproducible by the reader and will also allow for presenting a detailed analysis for several aspects of the proposed method.

6.1 | Stochastic model

The stochastic model used for generating the initial data set $\mathbb{D}_N = \{\mathbb{x}^j = (\mathbb{q}^j, \mathbb{w}^j), j = 1, \dots, N\}$ (see the Remark at the end of Section 2), related to random variable $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$ in which $\mathbb{Q} = (\mathbb{Q}_1, \dots, \mathbb{Q}_{n_q})$ and $\mathbb{W} = (\mathbb{W}_1, \dots, \mathbb{W}_{n_w})$, is written as

$$\mathbb{Q} = [\mathbb{B}(\mathbb{U})](\mathbb{W} + V \mathbb{b}_{\text{AP1}}), \quad (104)$$

in which \mathbb{U} , V , and \mathbb{W} are independent random variables. The maximum value of N is 300 and $n_w = 20$. We have $n_q = 200$. The deterministic vector \mathbb{b}_{AP1} in \mathbb{R}^{n_w} is written as $\mathbb{b}_{\text{AP1}} = 0.2 \mathbf{u} + 0.9$ in which all the components of \mathbf{u} belong to $]0, 1[$ (generated with the Matlab script: `rng('default');` $\mathbf{u} = \text{rand}(n_w, 1)$). The real-valued random variable $V = 0.2 \mathcal{U} + 0.9$ in which \mathcal{U} is a uniform random variable on $[0, 1]$. The random vector $\mathbb{U} = (\mathbb{U}_1, \dots, \mathbb{U}_{n_u})$ with $n_u = 6$

is written, for $\alpha = 1, \dots, n_u$, as $\mathbb{U}_\alpha = 2u_\alpha \mathcal{U}_\alpha + 1 - u_\alpha$ in which $\mathcal{U}_1, \dots, \mathcal{U}_{n_u}$ are n_u independent uniform random variables on $[0, 1]$ and where $u_\alpha = 0.2(\alpha - 1)/(n_u - 1)$. The entries $[\mathbb{B}(\mathbb{U})]_{kj}$ of the $(n_q \times n_w)$ random matrix are defined by $[\mathbb{B}(\mathbb{U})]_{kj} = \sum_{\alpha=1}^{n_u} \omega_\alpha(\mathbb{U}_\alpha) \phi_k^\alpha(\mathbb{U}_\alpha) \phi_{j+n_q/2}^\alpha(\mathbb{U}_\alpha)$ in which $\phi_k^\alpha(\mathbb{U}_\alpha) = \sin\{\alpha k \pi / (n_q + 1)\}$ is independent of \mathbb{U}_α (deterministic) and $\omega_\alpha(\mathbb{U}_\alpha) = 1/(\alpha \mathbb{U}_\alpha)^2$. The random vector \mathbb{W} is written as $\mathbb{W} = \sum_{\beta=1}^3 \sqrt{\varpi_\beta} \boldsymbol{\phi}_{\mathbb{W}}^\beta \mathbb{A}_\beta$, in which $\varpi_\beta = 1/\beta^2$ and $\boldsymbol{\phi}_{\mathbb{W}}^\beta = (\phi_{\mathbb{W},1}^\beta, \dots, \phi_{\mathbb{W},n_w}^\beta)$ with $\phi_{\mathbb{W},j}^\beta = \sin\{\beta \pi j / (1 + n_w)\}$. The non-Gaussian centered random vector $\mathbb{A} = (\mathbb{A}_1, \mathbb{A}_2, \mathbb{A}_3)$ is written as $\mathbb{A} = \sum_{\gamma=1}^{27} \mathbb{Q}^\gamma \psi_{\alpha_1^{(\gamma)}}(\Xi_1) \psi_{\alpha_2^{(\gamma)}}(\Xi_2)$ in which Ξ_1 and Ξ_2 are independent normalized Gaussian random variables. The indices $\alpha_1^{(\gamma)}$ and $\alpha_2^{(\gamma)}$ are such that $0 < \alpha_1^{(\gamma)} + \alpha_2^{(\gamma)} \leq 6$, and $\psi_{\alpha_1^{(\gamma)}}(\Xi_1)$ and $\psi_{\alpha_2^{(\gamma)}}(\Xi_2)$ are the polynomial chaos in Gaussian variables. The matrix $[\mathbb{Q}] = [\mathbb{Q}^1 \dots \mathbb{Q}^{27}]$ is such that $[\mathbb{Q}][\mathbb{Q}]^T = [I_3]$ and is generated using the Matlab script: `rng('default');` `M1 = randn(27,27);` `[M2, ~] = eig(M1*(M1)');` `M2(:, 4:27) = [];` `[\mathbb{Q}] = (M2)'`.

6.2 | Simulated experiments

Simulated experiments are not required for the proposed methodology. Only some statistics corresponding to the constraints must be specified. Nevertheless, in order to generate coherent statistics for the constraints (but also for validating the methodology), we introduce an experimental data set that is generated with an “experimental model.” This experimental data set is used for estimating the statistics that correspond to the constraints. We insist on the fact that, for applying the method proposed, the experimental data set is not used and, consequently, is not required.

The experimental data set $\mathbb{D}_{n_r}^{\text{exp}}$ is generated with $n_r = 200$ independent experimental realizations $\{\mathbb{Q}^{\text{exp},r}, r = 1, \dots, n_r\}$ of $\mathbb{Q}^{\text{exp}} = (\mathbb{Q}_1^{\text{exp}}, \dots, \mathbb{Q}_{n_q}^{\text{exp}})$, which are used for estimating the right-hand side member of Equation (10). The simulated experiments correspond to model uncertainties induced by modeling errors and are constructed using significant perturbations of the stochastic model. The experimental stochastic model is then written as

$$\mathbb{Q}^{\text{exp}} = [\mathbb{B}(\mathbb{U}^{\text{exp}})](\mathbb{W}^{\text{exp}} + V^{\text{exp}}|_{\mathbb{B}\mathbf{AP}1}), \quad (105)$$

in which \mathbb{U}^{exp} , V^{exp} , and \mathbb{W}^{exp} are independent random variables that are also independent of \mathbb{U} , V , and \mathbb{W} . The real-valued random variable $V^{\text{exp}} = 0.2\mathcal{U}^{\text{exp}} + 0.9$ in which \mathcal{U}^{exp} is a uniform random variable on $[0, 1]$ independent of \mathcal{U} . The random vector $\mathbb{U}^{\text{exp}} = (\mathbb{U}_1^{\text{exp}}, \dots, \mathbb{U}_{n_u}^{\text{exp}})$ is written, for $\alpha = 1, \dots, n_u$, as $\mathbb{U}_\alpha^{\text{exp}} = 2u_\alpha^{\text{exp}} \mathcal{U}_\alpha^{\text{exp}} + 1 - u_\alpha^{\text{exp}}$ in which $\mathcal{U}_1^{\text{exp}}, \dots, \mathcal{U}_{n_u}^{\text{exp}}$ are n_u independent uniform random variables on $[0, 1]$ and where $u_\alpha^{\text{exp}} = 0.3(\alpha - 1)/(n_u - 1)$. Note that the coefficient is 0.3 and not 0.2 as in the stochastic model. The mapping $u \mapsto [\mathbb{B}(u)]$ is the same as the one of the stochastic model. The random vector \mathbb{W}^{exp} is written as

$$\mathbb{W}^{\text{exp}} = \zeta_{\text{MEAN}} \times \mathbf{1}_{n_w} + \zeta_{\text{STD}} \times \tilde{\mathbb{W}}^{\text{exp}}, \quad (106)$$

in which $\mathbf{1}_{n_w} \in \mathbb{R}^{n_w}$ is the vector whose components are equal to 1 and where $\tilde{\mathbb{W}}^{\text{exp}}$ is an independent copy of the stochastic model of \mathbb{W} . The parameter ζ_{MEAN} controls the mean value of \mathbb{W}^{exp} while ζ_{STD} controls the amplitude of its covariance matrix. Many values of $(\zeta_{\text{MEAN}}, \zeta_{\text{STD}})$ have been considered in the subset $[-0.2, +0.7] \times [0, 1.4]$ for analyzing the robustness of the algorithm. In order to limit the number of figures, we will limit the presentation of the results to the value $\zeta_{\text{MEAN}} = 0.4$ and $\zeta_{\text{STD}} = 1.4$.

6.3 | Scaling

Using the block writing $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$, the scaling of the \mathbb{R}^{n_q} -valued random variable \mathbb{Q} and the scaling of the \mathbb{R}^{n_w} -valued random variable \mathbb{W} are written, using Equation (5), as

$$\mathbb{Q} = [\alpha_q] \mathbf{Q} + \mathbb{Q}^{\text{min}}, \quad \mathbf{Q} = [\alpha_q]^{-1}(\mathbb{Q} - \mathbb{Q}^{\text{min}}), \quad (107)$$

$$\mathbb{W} = [\alpha_w] \mathbf{W} + \mathbb{W}^{\text{min}}, \quad \mathbf{W} = [\alpha_w]^{-1}(\mathbb{W} - \mathbb{W}^{\text{min}}), \quad (108)$$

in which $\mathbf{x}^{\text{min}} = (\mathbb{Q}^{\text{min}}, \mathbb{W}^{\text{min}})$ and where $[\alpha_q]$ and $[\alpha_w]$ correspond to the block writing of $[\alpha_x]$.

The scaling of Equation (104) is then written as

$$\mathbf{Q} = [\mathbb{B}(\mathbb{U})](\mathbf{W} + V \mathbf{b}_{\mathbf{AP}1} + \mathbf{c}_{\mathbf{AP}1}) + \mathbf{d}_{\mathbf{AP}1}, \quad (109)$$

in which $[B(\mathbb{U})] = [\alpha_q]^{-1} [\mathbb{B}(\mathbb{U})] [\alpha_w]$, $\mathbf{b}_{AP1} = [\alpha_w]^{-1} \mathbf{a}_{AP1}$, $\mathbf{c}_{AP1} = [\alpha_w]^{-1} \mathbf{w}^{\min}$, and $\mathbf{d}_{AP1} = -[\alpha_q]^{-1} \mathbf{q}^{\min}$. The scaling of \mathbb{Q}^{\exp} is such that $\mathbf{Q}^{\exp} = [\alpha_q]^{-1} (\mathbb{Q}^{\exp} - \mathbf{q}^{\min})$, which shows that the n_r independent experimental realizations $\{\mathbf{q}^{\exp,r}, r = 1, \dots, n_r\}$ of \mathbf{Q}^{\exp} are such that

$$\mathbf{q}^{\exp,r} = [\alpha_q]^{-1} (\mathbf{q}^{\exp,r} - \mathbf{q}^{\min}). \quad (110)$$

6.4 | Constraints

The constraints are those that are defined in Section 4, that is to say by Equations (13) and (14) in which $\kappa_{\text{mom}} = n_q$ and where the experimental moments $\{(\mathfrak{m}_k^{\exp}, s_k^{\exp}), k = 1, \dots, \kappa_{\text{mom}}\}$ are estimated using the n_r independent experimental realizations $\{\mathbf{q}^{\exp,r}, r = 1, \dots, n_r\}$ defined by Equation (110). Consequently, the number of constraints is $m_c = 2 n_q = 400$.

6.5 | Values of the parameters of the algorithms

In this paragraph, all the values of the parameters are related to $N = 300$. The PCA of \mathbf{X} (see Section 5.1) is performed with $\varepsilon = 10^{-6}$ in Equation (17) and yields $v = 9$. The values of the two parameters of the Gaussian kernel-density estimation defined by Equation (24) are $s_v = 0.5966$ and $\hat{s}_v = 0.5129$. Concerning the construction of the diffusion-maps basis, the use of Equations (88) and (89) yields $\varepsilon_{\text{diff}}^{\text{opt}} = 48$ and the optimal value of m is $m^{\text{opt}} = 12$. The graph $\alpha \mapsto \Lambda_\alpha$ of the eigenvalues of the transition matrix defined in Section 5.8.1, which corresponds to these optimal values, is shown in Figure 1A. For solving the reduced-order nonlinear ISDE defined in Section 5.8.3, the parameters related to the algorithm, which are introduced in Appendix B, have been estimated using the procedure that is detailed in Section 4.7.2 in the work of Soize and Ghanem,³⁷ and are the following: $f_0 = 1.5$, $n_{MC} = 50$, $\ell_0 = 10$, $M_0 = 10$, $\Delta t = 0.16115$, and then, $M = 15\,000$. The projection of the constraints introduced in Section 5.3 has been performed with $\tau_c = 0.001$ that yields $v_c = 3$ (see Equation (39)). Figure 1B displays the graph of the error function $i \mapsto \text{err}(i)$ that is defined by Equation (59), which shows the convergence of the iteration algorithm for computing the Lagrange multiplier λ^{sol} . It can be seen that the convergence is fast, and the error decreases exponentially as a function of the iteration number i .

6.6 | Convergence of the learning with respect to the value of N

For this application, the use of the criterion defined by Equation (101) or by Equation (102) for analyzing the convergence of the learning with respect to dimension N of the initial data set is not sufficiently informative. For $N \in [125, 300]$, $\text{conv}_{L^2, \mathbf{X}^c}(N)$ fluctuates between 9.86 and 9.92. Consequently, for this application, the convergence of the learning is analyzed by using Section 5.10-(iii), that is to say by studying the sequence in N of graphs of the functions $\{k \mapsto s_{\mathbb{Q}_k^c}(N)\}_N$ and $\{q \mapsto p_{\mathbb{Q}_{110}^c}(q; N)\}_N$ for $N \in \{125, 150, 250, 300\}$. These quantities are related to the unscaled vector-valued random variable \mathbb{Q}^c with values in \mathbb{R}^{n_q} for which $n_q = 200$. For fixed N , $s_{\mathbb{Q}_k^c}(N)$ is the standard deviation of component \mathbb{Q}_k^c and $q \mapsto p_{\mathbb{Q}_{110}^c}(q; N)$ is the pdf of \mathbb{Q}_{110}^c (only component 110 that corresponds to the maximum of response has been selected for limiting the number of figures). Figure 2A shows the graphs of standard-deviation functions $k \mapsto s_{\mathbb{Q}_k^c}(N)$ for which a zoom in k has been carried out in order to better separate the curves for viewing the convergence in N . Figure 2B shows

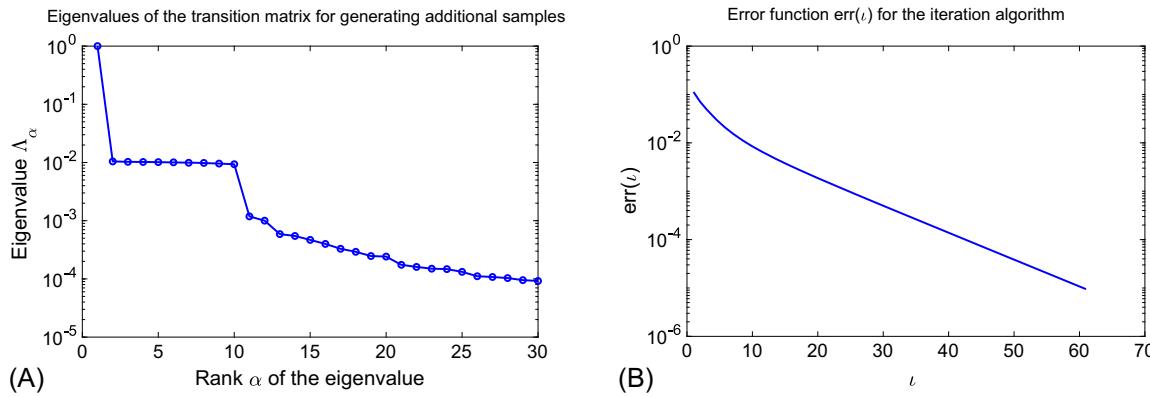


FIGURE 1 Application AP1. A, Graph $\alpha \mapsto \Lambda_\alpha$ of the eigenvalues of the transition matrix defined in Section 5.8.1; B, Graph of the error function $i \mapsto \text{err}(i)$ defined by Equation (59) showing the convergence of the iteration algorithm for computing the Lagrange multiplier λ^{sol} . The vertical axis is in log [Colour figure can be viewed at wileyonlinelibrary.com]

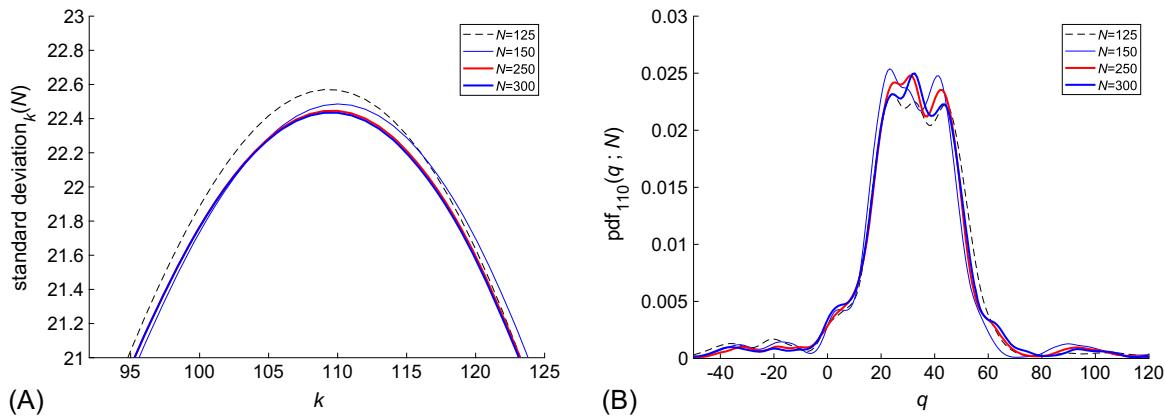


FIGURE 2 Convergence of the probabilistic learning for application AP1 with $M = 15\,000$ and for $N \in \{125, 150, 250, 300\}$. A, Graphs of the standard-deviation function $k \mapsto s_{Q_k^c}(N)$ for $k = 92, \dots, 125$; B, Graphs of probability density function $q \mapsto p_{Q_{110}^c}(q; N)$

the graphs of pdf functions $q \mapsto p_{Q_{110}^c}(q; N)$ for the same values of N . The convergence of the learning is particularly clear for the standard-deviation functions. For $N = 250$ and $N = 300$, the two curves are practically overlapping.

6.7 | Results obtained and validation of the method for $N = 300$ and $M = 15\,000$

For $k = 1, \dots, 200$, let \mathbb{Q}_k be the component k of the random unscaled QoI \mathbb{Q} (the scaled one is \mathbf{Q}) associated with the initial data set \mathbb{D}_N , let $\mathbb{Q}_k^{\text{exp}}$ be the component k of the random unscaled experimental QoI \mathbb{Q}^{exp} (the scaled one is \mathbf{Q}^{exp}), and let \mathbb{Q}_k^c be the component k of the random unscaled QoI \mathbb{Q}^c that verifies the constraints (the scaled one is \mathbf{Q}^c). The quality assessment consists in comparing \mathbb{Q}^c with the random vector $\mathbb{Q}^{c,\text{model}}$ that is transformed from \mathbb{W}^c using the stochastic computational model, that is to say, $\mathbb{Q}^{c,\text{model}} = [\mathbb{B}(\mathbb{U})](\mathbb{W}^c + V_{\text{model}})$. For that, the pdf of component \mathbb{Q}_k^c can be compared with the pdf of component $\mathbb{Q}_k^{c,\text{model}}$.

(i) For $k = 1, \dots, 200$, Figure 3A displays the graphs of the mean functions $k \mapsto m_{\mathbb{Q}_k}$, $k \mapsto m_{\mathbb{Q}_k^{\text{exp}}}$, $k \mapsto m_{\mathbb{Q}_k^c}$, and $k \mapsto m_{\mathbb{Q}_k^{c,\text{model}}}$. Figure 3B shows the graphs of the standard-deviation functions $k \mapsto s_{\mathbb{Q}_k}$, $k \mapsto s_{\mathbb{Q}_k^{\text{exp}}}$, $k \mapsto s_{\mathbb{Q}_k^c}$, and $k \mapsto s_{\mathbb{Q}_k^{c,\text{model}}}$. In each one of these two figures, it can be seen that the two curves $k \mapsto m_{\mathbb{Q}_k^{\text{exp}}}$ and $k \mapsto m_{\mathbb{Q}_k^c}$ are indistinguishable and that the two curves $k \mapsto s_{\mathbb{Q}_k^{\text{exp}}}$ and $k \mapsto s_{\mathbb{Q}_k^c}$ validate the method proposed: The constraints on the mean and the standard deviation are well taken into account. Concerning the quality assessment, it can also be seen that the two curves $k \mapsto m_{\mathbb{Q}_k}$ and $k \mapsto m_{\mathbb{Q}_k^{c,\text{model}}}$ are overlapping and that the two curves $k \mapsto s_{\mathbb{Q}_k^c}$ and $k \mapsto s_{\mathbb{Q}_k^{c,\text{model}}}$ are also overlapping, demonstrating the capability of the method at least for the second-order moments of the QoI.

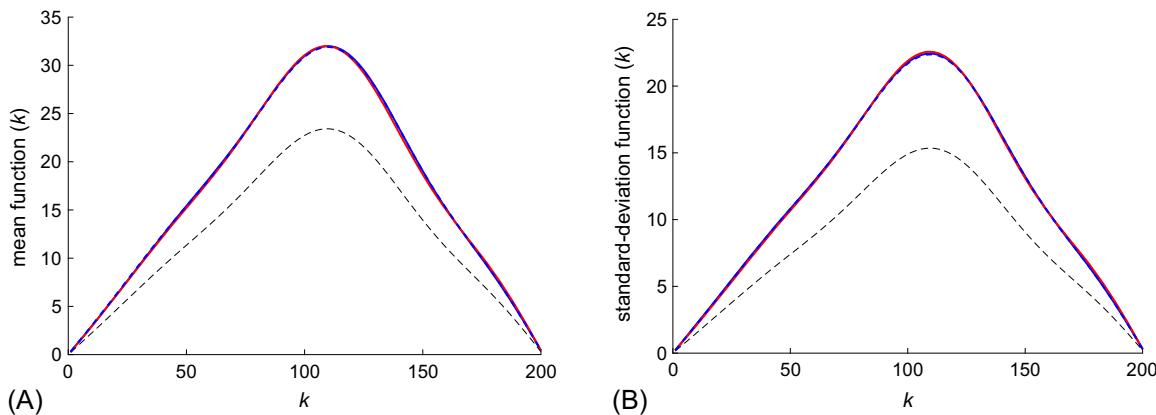


FIGURE 3 Application AP1 with $M = 15\,000$, for $N = 300$, for $n_r = 200$, and for $k = 1, \dots, 200$. A, Graphs of the mean functions $k \mapsto m_{\mathbb{Q}_k}$ (black dashed line), $k \mapsto m_{\mathbb{Q}_k^{\text{exp}}}$ (red line), $k \mapsto m_{\mathbb{Q}_k^c}$ (blue line), and $k \mapsto m_{\mathbb{Q}_k^{c,\text{model}}}$ (blue dashed line); B, Graphs of the standard-deviation functions $k \mapsto s_{\mathbb{Q}_k}$ (black dashed line), $k \mapsto s_{\mathbb{Q}_k^{\text{exp}}}$ (red line), $k \mapsto s_{\mathbb{Q}_k^c}$ (blue line), and $k \mapsto s_{\mathbb{Q}_k^{c,\text{exp}}}$ (blue dashed line). In each figure, the red line, the blue line, and the blue dashed line are overlapping

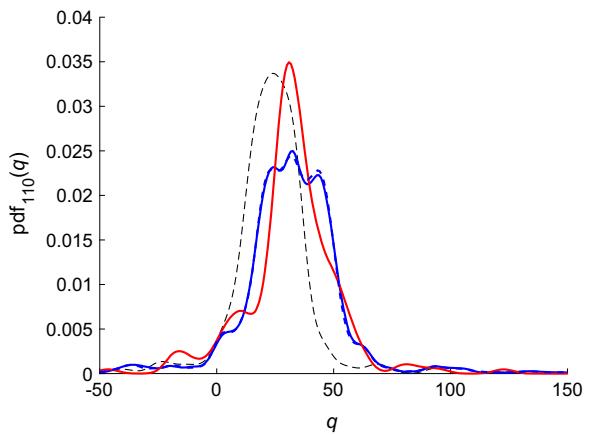


FIGURE 4 Application AP1: For $M = 15\,000$, for $N = 300$, for $n_r = 200$, graphs of the probability density functions $q \mapsto p_{Q_{110}}(q)$ (black dashed line), $q \mapsto p_{Q_{110}^{\text{exp}}}(q)$ (red line), $q \mapsto p_{Q_{110}^c}(q)$ (blue line), and $q \mapsto p_{Q_{110}^{c,\text{model}}}(q)$ (blue dashed line, almost overlapping with the blue line)

(ii) To complete the analysis and the validation, we present the results obtained for the probability density functions. In order to limit the number of figures, we have selected, as previously, component 110 for \mathbb{Q} . Figure 4 displays the graphs of the probability density functions $q \mapsto p_{Q_{110}}(q)$, $q \mapsto p_{Q_{110}^{\text{exp}}}(q)$, $q \mapsto p_{Q_{110}^c}(q)$, and $q \mapsto p_{Q_{110}^{c,\text{model}}}(q)$. The constraints that are imposed concern only the second-order moments of \mathbb{Q} . Consequently, there is no reason for that the pdf of Q_{110}^c coincides with the pdf of Q_{110}^{exp} . Nevertheless, it can be seen that the PLoM under these constraints yields a pdf that is not so bad. Note also that, in practice and in the framework of this application, the pdf of Q_{110}^{exp} cannot be estimated because the n_r realizations of Q^{exp} are not available. Nevertheless, it can be seen that the quality assessment is good enough because pdf $q \mapsto p_{Q_{110}^{c,\text{model}}}(q)$ is very close to pdf $q \mapsto p_{Q_{110}^c}(q)$.

7 | APPLICATION (AP2): PROBABILISTIC LEARNING OF THE RANDOM ELASTICITY FIELD FOR A HETEROGENEOUS ANISOTROPIC MEDIUM

In this section, we present an application in high dimension related to the learning of the random elasticity field for a heterogeneous anisotropic medium. This application is described in details in order that it can be reproduced. The \mathbb{R}^{n_w} -valued random system parameter, \mathbb{W} , corresponds to the spatial discretization of the tensor-valued random elasticity field for which $n_w = 420\,000$. The \mathbb{R}^{n_q} -valued random quantity of interest, \mathbb{Q} , has dimension $n_q = 1683$ corresponding to observed components of the displacement field at 561 points located on the boundary of the domain occupied by the heterogeneous material, and $m_c = 3\,366$ constraints defined in Section 4 are introduced. All the random variables are defined on probability space $(\Theta, \mathcal{T}, \mathcal{P})$.

7.1 | Construction of the initial data set

(i) Stochastic model defined by a stochastic boundary value problem

The stochastic model is used for generating the initial data set and is based on the stochastic boundary value problem defined hereinafter. The medium occupies the domain $\Omega = [0, 1.0] \times [0, 0.2] \times [0, 0.1]$ m^3 with generic point $\zeta = (\zeta_1, \zeta_2, \zeta_3)$. The boundary $\partial\Omega$ of Ω is written as $\partial\Omega = \Gamma_0 \cup \Gamma \cup \Gamma_{\text{obs}} \cup \Gamma_c$ in which

$$\begin{aligned}\Gamma_0 &= \{\zeta_1 = 1.0, 0 \leq \zeta_2 \leq 0.2, 0 \leq \zeta_3 \leq 0.1\}, \\ \Gamma &= \{\zeta_1 = 0.0, 0 \leq \zeta_2 \leq 0.2, 0 \leq \zeta_3 \leq 0.1\}, \\ \Gamma_{\text{obs}} &= \{0 \leq \zeta_1 \leq 1.0, 0 \leq \zeta_2 \leq 0.2, \zeta_3 = 0.1\},\end{aligned}$$

and where $\Gamma_c = \partial\Omega^{\text{meso}} \setminus \{\Gamma_0 \cup \Gamma \cup \Gamma_{\text{obs}}\}$. The outward unit normal to $\partial\Omega$ is denoted by $\mathbf{n}(\zeta)$. The heterogeneous complex material is modeled by a heterogeneous and anisotropic elastic random medium for which the elastic properties are defined by the fourth-order tensor-valued non-Gaussian random field $\{\mathbb{C} = \{\mathbb{C}_{ijkl}(\zeta)\}_{ijkl}, \zeta \in \Omega\}$, which is rewritten, for i, j, k , and h in $\{1, 2, 3\}$, as

$$\mathbb{C}_{ijkl}(\zeta) = [\mathbf{K}(\zeta)]_{IJ} \quad \text{with} \quad I = (i, j) \quad \text{and} \quad J = (k, h), \quad (111)$$

in which indices I and J belong to $\{1, \dots, 6\}$ and where the matrix-valued non-Gaussian random elasticity field $\{[\mathbf{K}(\zeta)], \zeta \in \Omega\}$ will be defined after. The \mathbb{R}^3 -valued displacement field $\boldsymbol{\Xi} = (\Xi_1, \Xi_2, \Xi_3)$ is defined in Ω . A Dirichlet

condition $\boldsymbol{\Xi} = \mathbf{0}$ is given on Γ_0 , whereas a Neumann condition is given on $\Gamma \cup \Gamma_{\text{obs}} \cup \Gamma_c$. The Neumann condition is zero on $\Gamma_{\text{obs}} \cup \Gamma_c$, whereas on Γ , a given random surface force field \mathcal{G}^Γ is applied (note that there is no surface force field applied to Γ_{obs} , which is the part of the boundary for which field $\boldsymbol{\Xi}$ is observed). The stochastic boundary value problem is written as

$$-\operatorname{div} \boldsymbol{\Sigma} = \mathbf{0} \quad \text{in } \Omega, \quad (112)$$

$$\boldsymbol{\Xi} = \mathbf{0} \quad \text{on } \Gamma_0, \quad (113)$$

$$\boldsymbol{\Sigma} \mathbf{n} = \mathcal{G}^\Gamma \quad \text{on } \Gamma, \quad (114)$$

$$\boldsymbol{\Sigma} \mathbf{n} = \mathbf{0} \quad \text{on } \Gamma_{\text{obs}} \cup \Gamma_c. \quad (115)$$

The stress tensor $\boldsymbol{\Sigma} = \{\Sigma_{ij}\}_{ij}$ is related to the strain tensor $\mathbf{E}(\boldsymbol{\Xi}) = \{E_{kh}(\boldsymbol{\Xi})\}_{kh}$ by the constitutive equation,

$$\Sigma_{ij}(\zeta) = \mathbb{C}_{ijkl}(\zeta) E_{kl}(\boldsymbol{\Xi}(\zeta)) \quad (116)$$

in which the strain tensor is such that $E_{kh}(\boldsymbol{\Xi}) = (\partial \Xi_k / \partial \zeta_h + \partial \Xi_h / \partial \zeta_k) / 2$.

(ii) Stochastic model of the applied forces

The surface force field applied on Γ induces the superposition (1) of a mean tension in the ζ_1 -axis induced by a surface force field applied in the $-\zeta_1$ -axis direction with a constant amplitude g^{ζ_1} , (2) of a mean bending around ζ_3 -axis induced by a surface force field applied in the $-\zeta_2$ -axis direction with a constant amplitude g^{ζ_2} , and (3) of a mean bending around ζ_2 -axis induced by a surface force field applied in the $-\zeta_3$ -axis direction with a constant amplitude g^{ζ_3} . Consequently, the \mathbb{R}^3 -valued random surface force field $\{\mathcal{G}^\Gamma(\zeta) = (\mathcal{G}_1^\Gamma(\zeta), \mathcal{G}_2^\Gamma(\zeta), \mathcal{G}_3^\Gamma(\zeta)), \zeta \in \Gamma\}$ applied on surface Γ is defined as follows. For all $\zeta = (\zeta_1, \zeta_2, \zeta_3)$ in Γ ,

$$\mathcal{G}_1^\Gamma(\zeta) = -g^{\zeta_1} \cos(\mathbb{U}_1), \quad (117)$$

$$\mathcal{G}_2^\Gamma(\zeta) = -g^{\zeta_1} \sin(\mathbb{U}_1) - g^{\zeta_2} \cos(\mathbb{U}_2) - g^{\zeta_3} \sin(\mathbb{U}_3), \quad (118)$$

$$\mathcal{G}_3^\Gamma(\zeta) = -g^{\zeta_2} \sin(\mathbb{U}_2) - g^{\zeta_3} \cos(\mathbb{U}_3), \quad (119)$$

where $\mathbb{U} = (\mathbb{U}_1, \mathbb{U}_2, \mathbb{U}_3)$ is a \mathbb{R}^3 -valued random variable such that

$$\mathbb{U}_k = \alpha_0(2\mathcal{U}_k - 1), \quad \alpha_0 = 2\pi\alpha_0^d/360, \quad \alpha_0^d = 1^\circ, \quad (120)$$

in which $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3$ are three independent uniform random variables on $[0, 1]$, which are also independent of random field \mathbb{C} . It can then be deduced that \mathbb{U} is independent of random field \mathbb{C} . The amplitudes are such that $g^{\zeta_1} = 15000 \text{ N/m}^2$, $g^{\zeta_2} = 130 \text{ N/m}^2$, and $g^{\zeta_3} = 50 \text{ N/m}^2$.

(iii) Stochastic model of the random elasticity field

The non-Gaussian fourth-order tensor-valued random field $\{\mathbb{C}(\zeta), \zeta \in \Omega\}$ is assumed to be of second order and statistically homogeneous. Its mean function is thus independent of ζ and is an elasticity tensor denoted by $\underline{\mathbb{C}}$, which is defined using its representation $[\underline{K}] \in \mathbb{M}_6^+$ (see Equation (111)) such that $\underline{\mathbb{C}}_{ijkl} = [\underline{K}]_{IJ}$. Elasticity tensor $\underline{\mathbb{C}}$ is associated with a homogeneous isotropic elastic material whose Young modulus is 10^{10} N/m^2 and Poisson coefficient 0.15 (note that the fluctuations are those of a heterogeneous anisotropic elastic material).

The non-Gaussian \mathbb{M}_6^+ -valued random field $\{[\mathbf{K}(\zeta)], \zeta \in \Omega\}$ is constructed using the stochastic model^{49,56} of random elasticity fields for heterogeneous anisotropic elastic media that are isotropic in statistical mean and exhibit anisotropic statistical fluctuations, for which the parameterization consists of spatial-correlation lengths and of a positive-definite lower bound. The random field $\{[\mathbf{K}(\zeta)], \zeta \in \Omega\}$ is written as

$$[\mathbf{K}(\zeta)] = [C_\ell] + [\underline{\mathbb{C}}]^{1/2} [\mathbf{G}_0(\zeta)] [\underline{\mathbb{C}}]^{1/2}, \quad \forall \zeta \in \Omega. \quad (121)$$

The lower-bound matrix is defined by $[C_\ell] = \varepsilon_\ell(1 + \varepsilon_\ell)^{-1} [\underline{K}] \in \mathbb{M}_6^+$ in which ε_ℓ is chosen equal to 10^{-6} . Consequently, $[\underline{\mathbb{C}}] = [\underline{K}] - [C_\ell] = (1 + \varepsilon_\ell)^{-1} [\underline{K}]$ belongs to \mathbb{M}_6^+ . Note that $[\underline{\mathbb{C}}]^{1/2}$ is the square root of matrix $[\underline{\mathbb{C}}]$ in \mathbb{M}_n^+ . The non-Gaussian

random field $\{[\mathbf{G}_0(\zeta)], \zeta \in \mathbb{R}^3\}$, which is indexed by \mathbb{R}^3 with values in \mathbb{M}_6^+ , is homogeneous in \mathbb{R}^3 and is a second-order random field such that, for all ζ in \mathbb{R}^3 ,

$$E\{[\mathbf{G}_0(\zeta)]\} = [I_6], \quad [\mathbf{G}_0(\zeta)] > 0 \quad \text{a.s.} \quad (122)$$

It can then be deduced that, for all ζ in Ω ,

$$E\{[\mathbf{K}(\zeta)]\} = [\underline{\mathbf{K}}] \in \mathbb{M}_6^+, \quad [\mathbf{K}(\zeta)] - [C_\ell] > 0 \quad \text{a.s.} \quad (123)$$

Random field $\{[\mathbf{G}_0(\zeta)], \zeta \in \mathbb{R}^3\}$ depends on three spatial correlation lengths, L_1 , L_2 , and L_3 , and on a dispersion parameter, δ_{G_0} . Its construction and its generator are summarized in Appendix C.

With such a stochastic model of the elasticity field, it is proven in the work of Soize⁵⁶ that the stochastic boundary value problem defined by Equations (112) to (116) admits a unique second-order solution.

(iv) Finite element approximation of the stochastic boundary value problem and construction of random vectors \mathbb{Q} and \mathbb{W}

Domain $\Omega =]0, 1.0[\times]0, 0.2[\times]0, 0.1[$ is meshed with $50 \times 10 \times 5 = 2500$ finite elements using eight-node finite elements. There are 3366 nodes and 10 098 degrees of freedom (dofs). The displacements are locked at all the 66 nodes belonging to surface Γ_0 , and therefore, there are 198 zero Dirichlet conditions. The \mathbb{R}^{n_q} -valued random variable \mathbb{Q} of the QoIs are constituted of the three dofs of the 561 observed nodes on surface Γ_{obs} ; we then have $n_q = 1683 = 3 \times 561$. There are eight integration points in each finite element. Consequently, there are $N_i = 20000$ integration points $\zeta^1, \dots, \zeta^{N_i}$. The \mathbb{R}^{n_w} -valued random variable \mathbb{W} is generated as follows. Let us consider the set $\{[\mathbf{L}(\zeta^1)], \dots, [\mathbf{L}(\zeta^{N_i})]\}$ of the random values at the N_i integration points of the random field $\{[\mathbf{L}(\zeta)], \zeta \in \mathbb{R}^3\}$ with values in the set of all the upper triangular (6×6) matrices with positive-valued diagonal entries. The random matrix $[\mathbf{L}(\zeta)]$ is the factor in the Cholesky factorization $[\mathbf{G}_0(\zeta)] = [\mathbf{L}(\zeta)]^T [\mathbf{L}(\zeta)]$ given by Equation (C1) (from Appendix C) of random matrix $[\mathbf{G}_0(\zeta)]$ used in Equation (121). The components of random vector \mathbb{W} are constituted, for $i = 1, \dots, N_i$ and for $1 \leq j \leq k \leq 6$, of the family of random variables $\log\{[\mathbf{L}(\zeta^i)]_{jj}\}$ and $[\mathbf{L}(\zeta^i)]_{jk}$ for $j < k$. Because there are 21 entries in the upper triangular random matrix $[\mathbf{L}(\zeta^i)]$, we have $n_w = 420000 = 20000 \times 21$.

7.1.1 | (v) Generation of the initial data set

The finite element discretization of the stochastic boundary value problem allows for constructing \mathbb{Q} as a function of \mathbb{W} and \mathbb{U} that is formally written as

$$\mathbb{Q} = \mathbb{f}(\mathbb{W}, \mathbb{U}), \quad (124)$$

in which $(w, u) \mapsto \mathbb{f}(w, u)$ is a measurable mapping on $\mathbb{R}^{n_w} \times \mathbb{R}^3$ with values in \mathbb{R}^{n_q} . Note that the deterministic mapping \mathbb{f} cannot explicitly be described because this mapping is associated with the solution of the finite element approximation of the stochastic boundary value problem. The initial data set $\mathbb{D}_N = \{\mathbb{x}^j = (\mathbb{q}^j, w^j), j = 1, \dots, N\}$, relative to the random variable $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$, is then constructed by using the Monte Carlo simulation method for solving the discretized stochastic equation, in which $\mathbb{q}^j = \mathbb{f}(w^j, u^j)$ where $\{w^j, j = 1, \dots, N\}$ and $\{u^j, j = 1, \dots, N\}$ are N independent realizations of random vectors \mathbb{W} and \mathbb{U} (note that u^j is also an independent realization of w^j). The generation is carried out with $N = 200$, $L_1 = L_2 = L_3 = 0.2$, and $\delta_{G_0} = 0.25$. For illustration, Figure 5 displays the 2D plot in the plane $\zeta_3 = 0.0958 \text{ m}$ of one realization of the random fields $(\zeta_1, \zeta_2) \mapsto \log\{[\mathbf{L}(\zeta_1, \zeta_2, 0.0958)]_{jj}\}$ for $j = 1, 2, 4$, and $(\zeta_1, \zeta_2) \mapsto [\mathbf{L}((\zeta_1, \zeta_2, 0.0958))]_{12}$.

7.2 | Generation of the experimental data set

As explained in Section 6 devoted to Application (AP1), simulated experiments are not required for the proposed methodology. Only some statistics that correspond to the constraints must be specified. Nevertheless, similarly to Application (AP1), in order to generate coherent statistics for the constraints, we introduce an experimental data set generated as a perturbation of the stochastic boundary value problem (previously, we insist on the fact that for applying the method proposed, the experimental data set is not used and consequently, is not required). The experimental data set $\mathbb{D}_{n_r}^{\text{exp}}$ is generated with $n_r = 1000$ independent experimental realizations $\{\mathbb{q}^{\text{exp}, r}, r = 1, \dots, n_r\}$ of $\mathbb{Q}^{\text{exp}} = (\mathbb{Q}_1^{\text{exp}}, \dots, \mathbb{Q}_{n_q}^{\text{exp}})$ such that

$$\mathbb{Q}^{\text{exp}} = \mathbb{f}(\mathbb{W}^{\text{exp}}, \mathbb{U}^{\text{exp}}), \quad (125)$$

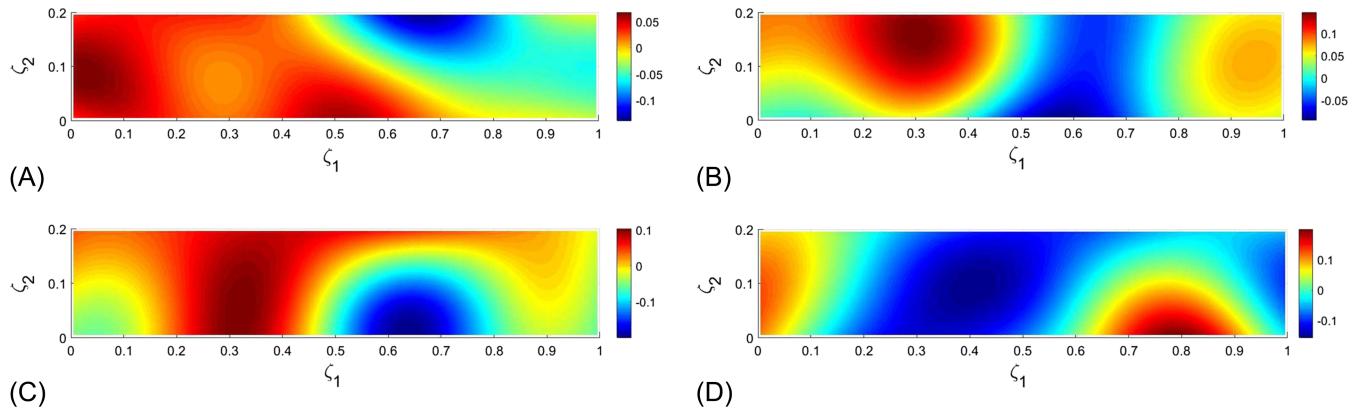


FIGURE 5 Application AP2: 2D plot in the plane $\zeta_3 = 0.0958\text{ m}$ of one realization of the random field $(\zeta_1, \zeta_2) \mapsto \log\{[\mathbf{L}(\zeta_1, \zeta_2, 0.0958)]_{jj}\}$ for $j = 1$ (a), $j = 2$ (b), $j = 4$ (c), and of the random field $(\zeta_1, \zeta_2) \mapsto [\mathbf{L}(\zeta_1, \zeta_2, 0.0958)]_{12}$ (d)

in which the deterministic mapping \mathbb{f} is the same as the one used in Equation (124). The \mathbb{R}^3 -valued random variable \mathbb{U}^{\exp} is an independent copy of the \mathbb{R}^3 -valued random variable \mathbb{U} introduced in Section 7.1-(ii). The random vector \mathbb{W}^{\exp} is constructed as an independent copy of random vector \mathbb{W} defined in Section 7.1-(iv) for which the spatial correlation lengths are $L_1^{\exp} = L_2^{\exp} = L_3^{\exp} = 0.16$ and the dispersion parameter is $\delta_{G_0}^{\exp} = 0.45$.

7.3 | Scaling

Using the block notation $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$, the scaling of the \mathbb{R}^{n_q} -valued random variable \mathbb{Q} and the scaling of the \mathbb{R}^{n_w} -valued random variable \mathbb{W} is carried out as explained in Section 6.3 and is obtained by using Equations (107) and (108) whose values of the parameterized transformation are those of the present application (AP2). The scaling of Equation (124) is then written as

$$\mathbb{Q} = \mathbf{f}(\mathbb{W}, \mathbb{U}), \quad (126)$$

in which $\mathbf{f}(\mathbb{W}, \mathbb{U}) = [\alpha_q]^{-1}(\mathbb{f}([\alpha_w]\mathbb{W} + \mathbb{w}^{\min}, \mathbb{U}) - \mathbb{q}^{\min})$. The scaling of \mathbb{Q}^{\exp} is such that $\mathbb{Q}^{\exp} = [\alpha_q]^{-1}(\mathbb{Q}^{\exp} - \mathbb{q}^{\min})$, which shows that the n_r independent experimental realizations $\{\mathbf{q}^{\exp,r}, r = 1, \dots, n_r\}$ of \mathbb{Q}^{\exp} are such that

$$\mathbf{q}^{\exp,r} = [\alpha_q]^{-1}(\mathbb{q}^{\exp,r} - \mathbb{q}^{\min}). \quad (127)$$

7.4 | Constraints

The constraints are those that are defined by Equations (13) and (14) in which $\kappa_{\text{mom}} = n_q$ and where the experimental moments $\{(\mathbb{m}_\kappa^{\exp}, s_\kappa^{\exp}), \kappa = 1, \dots, \kappa_{\text{mom}}\}$ are estimated using the n_r independent experimental realizations $\{\mathbf{q}^{\exp,r}, r = 1, \dots, n_r\}$ defined by Equation (127). Consequently, the number of constraints is $m_c = 3366$.

7.5 | Values of the parameters of the algorithms

All the values of the parameters given in this paragraph are related to $N = 200$.

- (i) The PCA of \mathbf{X} (see Section 5.1) is performed with $\varepsilon = 10^{-3}$ in Equation (17) and yields $v = 193$. The graph of the error function $v \mapsto \text{err}_{\text{PCA}}(v)$ for $v \in [1, 193]$, defined by Equation (17), is shown in Figure 6A.
- (ii) The parameters of the Gaussian kernel-density estimation defined by Equation (24) are $s_v = 0.9544$ and $\hat{s}_v = 0.6913$.
- (iii) For the diffusion-maps basis, the use of Equations (88) and (89) yields $\varepsilon_{\text{diff}}^{\text{opt}} = 800$ and the optimal value of m is $m^{\text{opt}} = 195$. The graph $\alpha \mapsto \Lambda_\alpha$ of the eigenvalues of the transition matrix defined in Section 5.8.1, which corresponds to these optimal values, is shown in Figure 6B.
- (iv) For solving the reduced-order nonlinear ISDE defined in Section 5.8.3, the parameters related to the algorithm are those introduced in Appendix B, have been estimated using the procedure that is detailed in Section 4.7.2 in the work of Soize and Ghanem,³⁷ and are the following: $f_0 = 1.5$, $n_{\text{MC}} = 250$, $\ell_0 = 10$, $M_0 = 10$, $\Delta t = 0.217192$, and then, $M = 50\,000$.

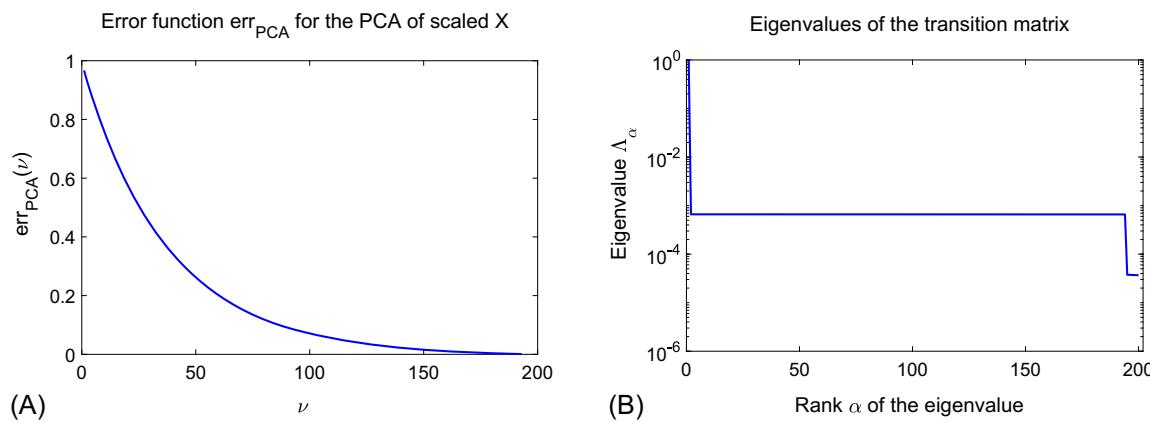


FIGURE 6 Application AP2. A, Graph of principal component analysis–error function $\nu \mapsto \text{err}_{\text{PCA}}(\nu)$ for $\nu \in [1, 193]$, defined by Equation (17). B, Graph $\alpha \mapsto \Lambda_\alpha$ of the eigenvalues of the transition matrix defined in Section 5.8.1 (the vertical axis is in log) [Colour figure can be viewed at wileyonlinelibrary.com]

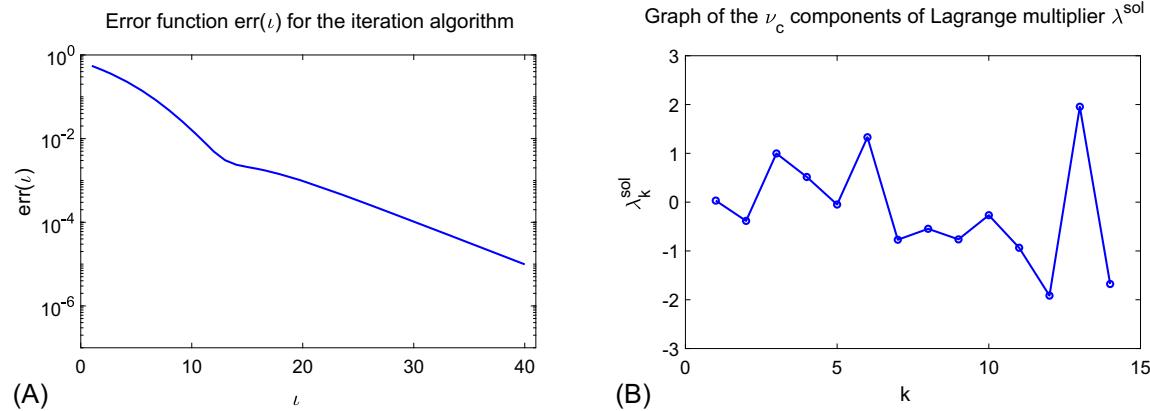


FIGURE 7 Application AP2. A, Graph of the error function $i \mapsto \text{err}(i)$ defined by Equation (59) showing the convergence of the iteration algorithm for computing the Lagrange multiplier λ^{sol} (the vertical axis is in log). B, Graph of the components $k \mapsto \lambda_k^{\text{sol}}$, for $k \in [1, v_c]$ with $v_c = 14$, of the optimal solution $\lambda^{\text{sol}} = (\lambda_1^{\text{sol}}, \dots, \lambda_{v_c}^{\text{sol}})$ of the Lagrange multipliers λ computed by Equation (54) [Colour figure can be viewed at wileyonlinelibrary.com]

(v) The projection of the constraints introduced in Section 5.3 has been performed with $\tau_c = 0.004$ that yields $v_c = 14$ (see Equation (39)). Figure 7A displays the graph of the error function $i \mapsto \text{err}(i)$ that is defined by Equation (59), which shows the convergence of the iteration algorithm for computing the Lagrange multiplier λ^{sol} . It can be seen that the convergence is fast, the error decreases exponentially as a function of the iteration number i . Figure 7B shows the graph of the components $k \mapsto \lambda_k^{\text{sol}}$, for $k \in [1, v_c]$ with $v_c = 14$, of the optimal solution $\lambda^{\text{sol}} = (\lambda_1^{\text{sol}}, \dots, \lambda_{v_c}^{\text{sol}})$ of the Lagrange multipliers λ computed by Equation (54).

7.6 | Convergence of the learning with respect to the value of N

Similarly to Application (AP1), the use of the criterion defined by Equation (101) or by Equation (102) for analyzing the convergence with respect to N (convergence of the learning with respect to dimension of the initial data set) is not sufficiently informative for this application. In addition, taking into account the high dimensions $n_w = 420\,000$ and $n_q = 1683$, it is neither easy nor efficient to perform the analysis as carried out for Application (AP1). In this condition, we therefore propose to analyze the convergence of the mean value and the standard deviation of unscaled displacements Q_k of three selected nodes, for which the displacements are maximum for the mean responses: Q_2 corresponding to the ζ_2 -displacement of the node of coordinates $(0, 0, 0.1)$, Q_{31} corresponding to the ζ_1 -displacement of the node of coordinates $(0, 0.2, 0.1)$, and Q_{33} corresponding to the ζ_3 -displacement of node of coordinates $(0, 0.2, 0.1)$.

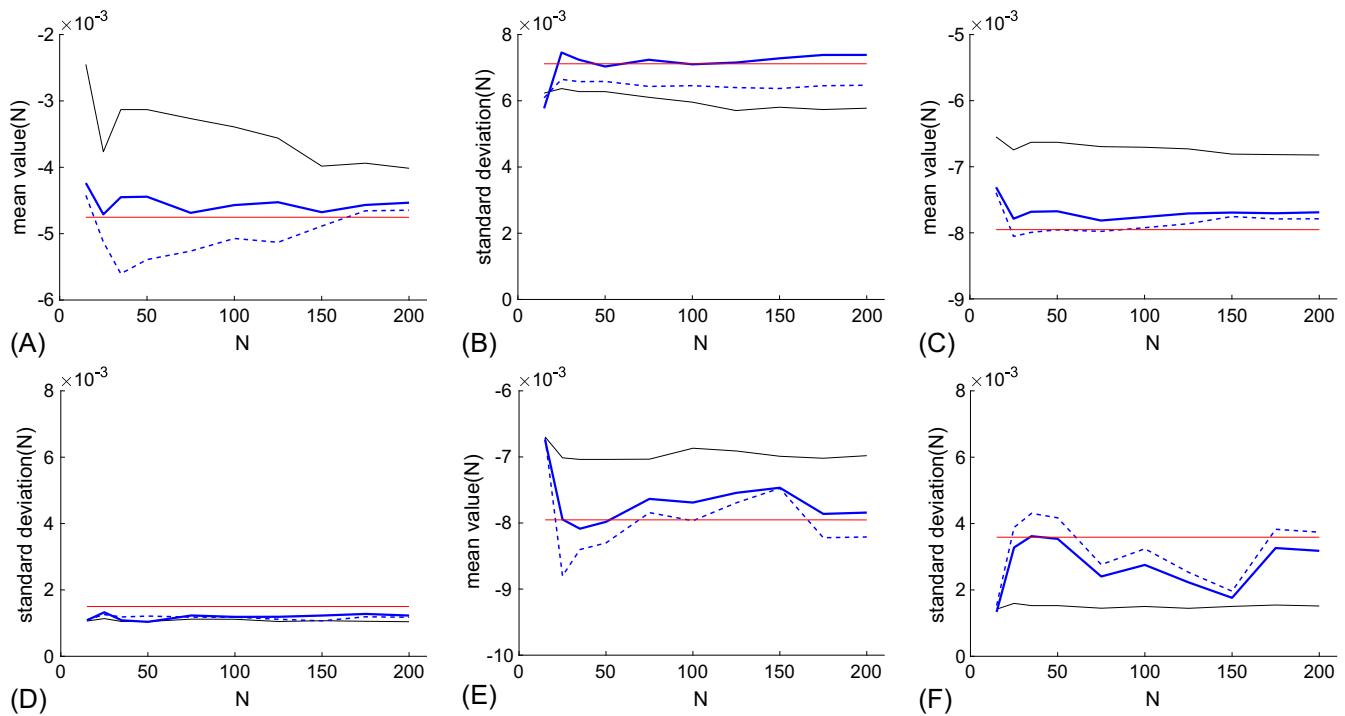


FIGURE 8 Application AP2: For component $k = 2$ (Figures A and B), $k = 31$ (Figures C and D), and $k = 33$ (Figures E and F), the left figures plot the graphs of the mean functions $N \mapsto m_{Q_k}(N)$ (black thin line), $N \mapsto m_{Q_k}^{\text{exp}}(N)$ (red horizontal line), $N \mapsto m_{Q_k^c}(N)$ (blue thick line), and $N \mapsto m_{Q_k^c,\text{model}}(N)$ (blue dashed line); the right figures plot the graphs of the standard deviation of functions $N \mapsto s_{Q_k}(N)$ (black dashed line), $N \mapsto s_{Q_k}^{\text{exp}}(N)$ (red line), $N \mapsto s_{Q_k^c}(N)$ (blue line), and $N \mapsto s_{Q_k^c,\text{model}}(N)$ (blue dashed line)

The computation has been carried out with $M = 50\,000$, $n_r = 1000$, and for 10 values of integer N such that $15 \leq N \leq 200$.

For $k \in \{2, 31, 33\}$, Figure 8 allows for analyzing the convergence with respect to N of the mean value $m(N)$ and of the standard deviation $s(N)$ of the following random variables: Q_k (estimated with the N points of the initial data set), Q_k^c (estimated with the M additional realizations of the generated data set computed with the PLoM under the $m_c = 2366$ constraints), Q_k^{exp} (corresponding to the experimental data set), and finally, for the quality assessment, $Q_k^{\text{c},\text{model}}$ that results from the transformation of \mathbb{W}^c using the stochastic computational model, that is to say, $Q_k^{\text{c},\text{model}} = f(\mathbb{W}^c, \mathbb{U})$. It can be seen that the effects of the constraints are significant: the thin lines (without constraints) are different from the thick lines (with constraints). The convergence with respect to N is good enough even for the quality assessment: The thick lines (with constraints) are close to the dashed lines (QoI computed with the stochastic model using \mathbb{W}^c learned with the constraints).

7.7 | Results obtained for $N=200$ and $M=50\,000$

Similarly to Section 7.6, we consider, for $k = 2, 31, 33$, Q_k defined by the initial data set, Q_k^{exp} associated with the experiments, Q_k^c constructed using the constraints, and $Q_k^{\text{c},\text{model}}$ transformed from \mathbb{W}^c using the stochastic computational model.

(i) For $N = 200$ and $M = 50\,000$, for components $k = 2, 31, 33$, Table 1 gives the mean values m_{Q_k} , $m_{Q_k}^{\text{exp}}$, $m_{Q_k^c}$, $m_{Q_k^{\text{c},\text{model}}}$ and the standard deviations s_{Q_k} , $s_{Q_k}^{\text{exp}}$, $s_{Q_k^c}$, $s_{Q_k^{\text{c},\text{model}}}$. This table allows comparisons to be carried out. As already pointed out in Section 7.6, it can be seen that the constraints play an important role because the values relative to Q_k (without the constraints) are very different from the corresponding values relative to Q_k^c (with the constraints). The predictions obtained under the constraints are good enough when comparing the values of Q_k^c (with the constraints) with the corresponding target values for Q_k^{exp} . Finally, the quality assessment is good taking into account the high dimension of the problem and the constraints that are applied, which are only relative to a subset of all the dofs and related to second-order moments. This quality can be analyzed by comparing the values relative to Q_k^c (with the constraints) with the values relative to $Q_k^{\text{c},\text{model}}$.

TABLE 1 Applications (AP2): For $N = 200$ and $M = 50\,000$, and for $k = 2, 31, 33$, values of the mean $\mathbb{M}_{Q_k}, \mathbb{M}_{Q_k^{\text{exp}}}, \mathbb{M}_{Q_k^c}, \mathbb{M}_{Q_k^{\text{c},\text{model}}}$, and values of the standard deviations $\mathbb{S}_{Q_k}, \mathbb{S}_{Q_k^{\text{exp}}}, \mathbb{S}_{Q_k^c}, \mathbb{S}_{Q_k^{\text{c},\text{model}}}$

Millimeter ($10^{-3} m$)	Component 2	Component 31	Component 33
\mathbb{M}_{Q_k}	-4.01	-6.82	-6.98
$\mathbb{M}_{Q_k^c}$	-4.53	-7.69	-7.84
$\mathbb{M}_{Q_k^{\text{c},\text{model}}}$	-4.65	-7.79	-8.21
$\mathbb{M}_{Q_k^{\text{exp}}}$	-4.75	-7.95	-7.95
\mathbb{S}_{Q_k}	5.78	1.04	1.52
$\mathbb{S}_{Q_k^c}$	7.38	1.23	3.18
$\mathbb{S}_{Q_k^{\text{c},\text{model}}}$	6.47	1.17	3.74
$\mathbb{S}_{Q_k^{\text{exp}}}$	7.12	1.50	3.59

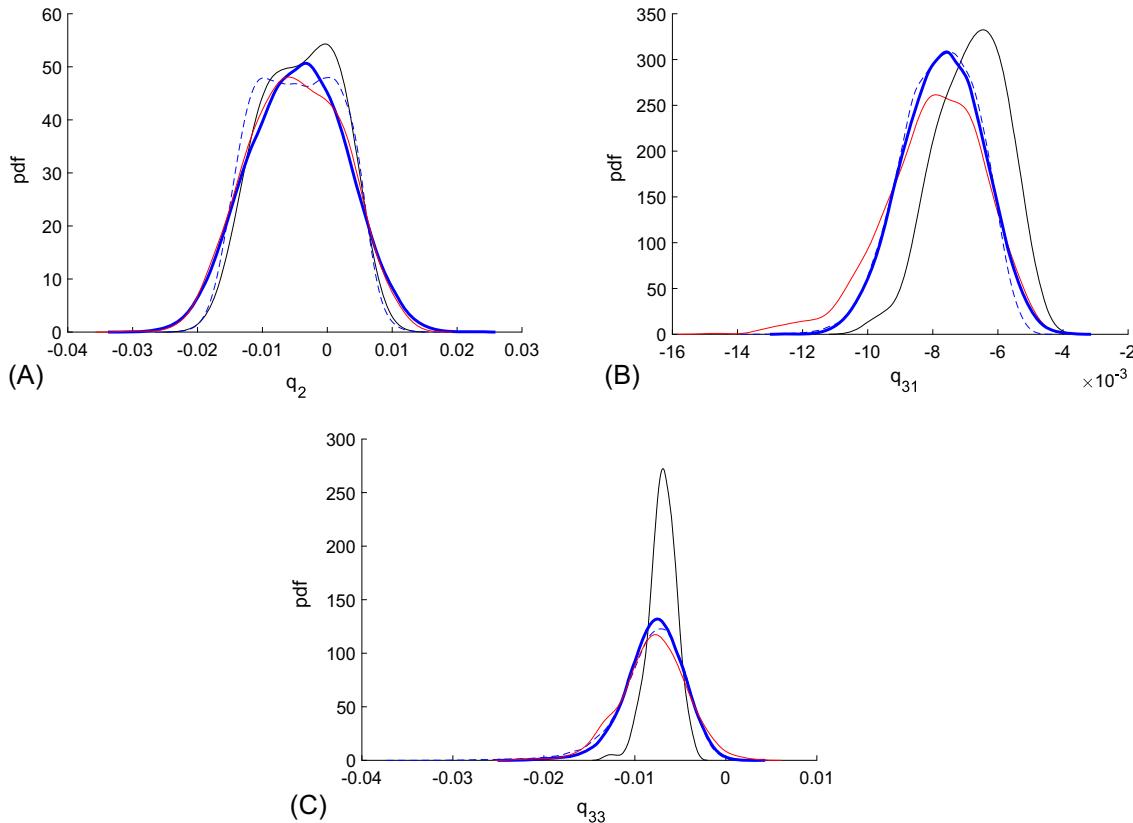


FIGURE 9 Application AP2: For $M = 50\,000$ and $N = 200$, the three figures show, for components $k = 2$ (a), $k = 31$ (b), and $k = 33$ (c), the pdf of Q_k (black thin line), Q_k^{exp} (red line), Q_k^c (blue thick line), and $Q_k^{\text{c},\text{model}}$ (blue dashed line)

(ii) To complete the analysis and validation, we present the results obtained for the probability density functions for components $k = 2$ (Figure 9A), $k = 31$ (Figure 9B), and $k = 33$ (Figure 9C). These figures display the graphs of the probability density functions of Q_k , Q_k^{exp} , Q_k^c , and $Q_k^{\text{c},\text{model}}$. The results are very good and even a little unexpected because only the mean and the standard deviation are imposed for a subset of dofs of the stochastic computational model. The pdf of Q_k^c is close to the pdf of $Q_k^{\text{c},\text{model}}$, which is itself close to the pdf of Q_k^{exp} (this experimental pdf is not used in the methodology, is not required, and is not available in the framework of the method proposed).

(iii) In order to illustrate how the constraints qualitatively impact the concentration of the probability measure of random vector $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$, the left part of Figure 10 shows the 50 000 additional realizations computed with the PLoM for random vector $(\mathbb{W}_{21000}, \mathbb{W}_{121000}, \mathbb{Q}_{33})$ (without the constraints) and $(\mathbb{W}_{21000}, \mathbb{W}_{121000}, \mathbb{Q}_{33}^c)$ (with the constraints), whereas the right part of Figure 10 shows the results for $(\mathbb{W}_{121000}, \mathbb{W}_{321000}, \mathbb{Q}_{33})$ (without the constraints) and $(\mathbb{W}_{121000}, \mathbb{W}_{321000}, \mathbb{Q}_{33}^c)$ (with the constraints). It should be noted that the two submanifolds that are considered are not deterministic due to the presence of the random vector \mathbb{U} and to the other components of \mathbb{W} , which are not used in the chosen representation.

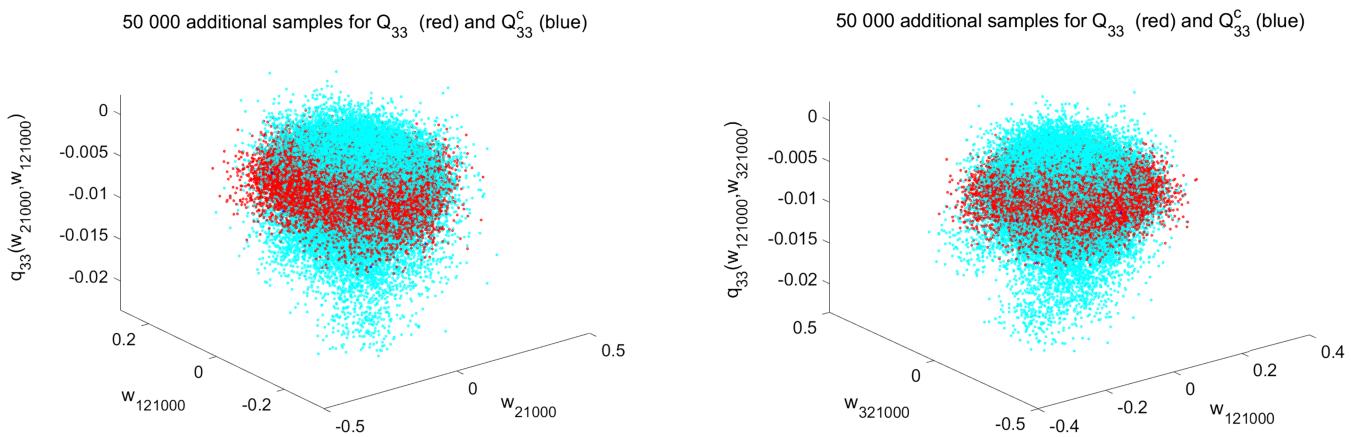


FIGURE 10 Application AP2. Left figure: 50 000 realizations of $(\mathbb{W}_{21000}, \mathbb{W}_{121000}, Q_{33})$ (red symbols) and of $(\mathbb{W}_{21000}, \mathbb{W}_{121000}, Q_{33}^c)$ (blue symbols). Right figure: 50 000 realizations of $(\mathbb{W}_{121000}, \mathbb{W}_{321000}, Q_{33})$ (red symbols) and of $(\mathbb{W}_{121000}, \mathbb{W}_{321000}, Q_{33}^c)$ (blue symbols)

The relatively large dispersion that seems to be obtained is only an appearance that is related to the choice of the scale that has been taken for Q_{33} . On the other hand, according to Table 1, Q_{33}^c is more scattered than Q_{33} due to the imposed constraints, so the concentration is less for Q_{33}^c than for Q_{33} , which is normal.

8 | CONCLUSION

In this paper, we have presented a methodology that extends the probabilistic leaning on manifolds from a small data set to the case for which constraints are imposed, during the learning process, to a subset of QoI. These constraints are, for instance, experimental second-order statistical moments (mean and standard deviation) that are given. The method has the capability to consider more general constraints than the second-order statistical moments. The proposed method allows for analyzing non-Gaussian cases in high dimension related to functional inputs and outputs. The iteration algorithm presented is very robust and seems to be exponentially convergent with respect to the number of iterations. Two applications have been presented, which allow for analyzing the behavior of the method and for validating it.

NOTATIONS

Lowercase letters such as q or η are deterministic real variables.

Boldface lowercase letters such as \mathbf{q} or $\boldsymbol{\eta}$ are deterministic vectors.

Lowercase letters \mathbf{q} , \mathbf{w} , and \mathbf{x} are deterministic vectors.

Uppercase letters such as X or H are real-valued random variables.

Boldface uppercase letters such as \mathbf{X} or \mathbf{H} are vector-valued random variables.

Uppercase letters \mathbb{Q} , \mathbb{U} , \mathbb{W} , and \mathbb{X} are vector-valued random variables.

Lowercase letters between brackets such as $[x]$ or $[\eta]$ are deterministic matrices.

Boldface uppercase letters between brackets such as $[\mathbf{X}]$ or $[\mathbf{H}]$ are matrix-valued random variables.

m_c : number of constraints.

n : dimension ($n = n_q + n_w$) of vectors \mathbf{x} , \mathbf{X} , and \mathbf{X}^c .

n_q : dimension of vectors \mathbf{q} , \mathbf{Q} , and \mathbf{Q}^c .

n_w : dimension of vectors \mathbf{w} , \mathbf{W} , and \mathbf{W}^c .

v : dimension of \mathbf{H} and \mathbf{H}^c .

$\mathbf{q}^j, \mathbf{q}^{c,\ell}$: realization of \mathbf{Q} , \mathbf{Q}^c .

$\mathbf{w}^j, \mathbf{w}^{c,\ell}$: realization of \mathbf{W} , \mathbf{W}^c .

$\mathbf{x}^j, \mathbf{x}^{c,\ell}$: realization of \mathbf{X} , \mathbf{X}^c .

M : number of points in the generated data set D_M^g .

N : number of points in initial data set D_N .

D_N : initial data set for \mathbf{X} .

- D_M^g : generated data set for \mathbf{X}^c .
 \mathbb{D}_N : initial data set for \mathbb{X} .
 \mathbf{Q}, \mathbf{Q}^c : random QoI (output).
 \mathbf{W}, \mathbf{W}^c : random system parameter (input).
 \mathbf{X}, \mathbf{X}^c : random vector $(\mathbf{Q}, \mathbf{W}), (\mathbf{Q}^c, \mathbf{W}^c)$.
 $[I_n]$: identity matrix in \mathbb{M}_n .
 $\mathbb{M}_{n,N}$: set of all the $(n \times N)$ real matrices.
 \mathbb{M}_n : set of all the square $(n \times n)$ real matrices.
 \mathbb{M}_n^{+0} : set of all the positive symmetric $(n \times n)$ real matrices.
 \mathbb{M}_n^+ : set of all the positive-definite symmetric $(n \times n)$ real matrices.
 \mathbb{R} : set of all the real numbers.
 \mathbb{R}^n : Euclidean vector space on \mathbb{R} of dimension n .
 $[x]_{kj}$: entry of matrix $[x]$.
 $[x]^T$: transpose of matrix $[x]$.
 $\delta_{kk'}$: Kronecker's symbol such that $\delta_{kk'} = 0$ if $k \neq k'$ and $= 1$ if $k = k'$.
 $\|\mathbf{x}\|$: usual Euclidean norm in \mathbb{R}^n .
 $\langle \mathbf{x}, \mathbf{y} \rangle$: usual Euclidean inner product in \mathbb{R}^n .
 $\delta_{kk'}$: Kronecker's symbol.

ACKNOWLEDGEMENT

Part of this research was supported by the PIRATE project funded under DARPA's AIRA program.

ORCID

- Christian Soize  <https://orcid.org/0000-0002-1083-6771>
Roger Ghanem  <https://orcid.org/0000-0002-1890-920X>

REFERENCES

1. Kaipio J, Somersalo E. *Statistical and Computational Inverse Problems*. New York, NY: Springer-Verlag; 2005.
2. Marzouk YM, Najm HN, Rahn LA. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J Comput Phys*. 2007;224(2):560-586. <https://doi.org/10.1016/j.jcp.2006.10.010>
3. Carlin BP, Louis TA. *Bayesian Methods for Data Analysis*. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC Press; 2009.
4. Stuart AM. Inverse problems: a Bayesian perspective. *Acta Numerica*. 2010;19:451-559. <https://doi.org/10.1017/S0962492910000061>
5. Spantini A, Cui T, Willcox K, Tenorio L, Marzouk Y. Goal-oriented optimal approximations of Bayesian linear inverse problems. *SIAM J Sci Comput*. 2017;39(5):S167-S196. <https://doi.org/10.1137/16M1082123>
6. Marin J-M, Pudlo P, Robert CP, Ryder RJ. Approximate Bayesian computational methods. *Stat Comput*. 2012;22(6):1167-1180. <https://doi.org/10.1007/s11222-011-9288-2>
7. Marjoram P, Molitor J, Plagnol V, Tavaré S. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci USA*. 2003;100(26):15324-15328. <https://doi.org/10.1073/pnas.0306899100>
8. Jaynes ET. Information theory and statistical mechanics. *Physical Review*. 1957;106(4):620-630.
9. Bhattacharyya A. On the measures of divergence between two statistical populations defined by their probability distributions. *Bull Calcutta Math Soc*. 1943;35:99-109.
10. Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat*. 1951;22(1):79-86.
11. Kapur JN, Kesavan HK. *Entropy Optimization Principles with Applications*. San Diego, CA: Academic Press; 1992.
12. Gish H. A probabilistic approach to the understanding and training of neural network classifiers. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing; 1990; Albuquerque, NM.
13. Shatkay H, Kaelbling LP. Learning topological maps with weak local odometric information. In: Proceedings of Fifteenth International Joint Conference on Artificial Intelligence (IJCAI'97); 1997; Nagoya, Japan.
14. Barber D, Bishop CM. Ensemble learning for multi-layer networks. In: Proceedings of the Advances in Neural Information Processing Systems; 1998; Denver, CO.
15. Hollmén J, Tresp V, Simula O. A learning vector quantization algorithm for probabilistic models. In: Proceedings of the 2000 10th European Signal Processing Conference; 2000; Tampere, Finland.
16. Galata A, Johnson N, Hogg D. Learning variable-length Markov models of behavior. *Comput Vis Image Underst*. 2001;81(3):398-413. <https://doi.org/10.1006/cviu.2000.0894>

17. Bigi B. Using Kullback-Leibler distance for text categorization. In: Proceedings of the European Conference on Information Retrieval; 2003; Pisa, Italy.
18. Kaski S, Sinkkonen J. Principle of learning metrics for exploratory data analysis. *J VLSI Signal Process Syst Signal Image Video Technol*. 2004;37(2-3):177-188.
19. Lange T, Law MHC, Jain AK, Buhmann JM. Learning with constrained and unlabelled data. In: Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); 2005; San Diego, CA.
20. Peperkamp S, Le Calvez R, Nadal J-P, Dupoux E. The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition*. 2006;101(3):B31-B41. <https://doi.org/10.1016/j.cognition.2005.10.006>
21. Filippi S, Cappé O, Garivier A. Optimism in reinforcement learning and Kullback-Leibler divergence. In: Proceedings of the 2010 48th Annual Allerton Conference on Communication, Control, and Computing; 2010; Allerton, IL.
22. Vasconcelos N, Ho P, Moreno P. The Kullback-Leibler Kernel as a framework for discriminant and localized representations for visual recognition. In: Proceedings of the European Conference on Computer Vision; 2004; Prague, Czech Republic.
23. Zhang W, Shan S, Chen X, Gao W. Local Gabor binary patterns based on Kullback-Leibler divergence for partially occluded face recognition. *IEEE Signal Process Lett*. 2007;14(11):875-878. <https://doi.org/10.1109/LSP.2007.903260>
24. Köstinger M, Hirzer M, Wohlhart P, Roth PM, Bischof H. Large scale metric learning from equivalence constraints. In: Proceedings of the 2012 IEEE Conference on Computer vision and Pattern Recognition; 2012; Providence, RI.
25. Cappé O, Garivier A, Maillard O-A, Munos R, Stoltz G. Kullback-Leibler upper confidence bounds for optimal sequential allocation. *Ann Stat*. 2013;41(3):1516-1541. <https://doi.org/10.1214/13-AOS119>
26. Ran Z-Y, Hu B-G. Determining structural identifiability of parameter learning machines. *Neurocomputing*. 2014;127:88-97. <https://doi.org/10.1016/j.neucom.2013.08.039>
27. Sun M, Li Y, Gemmeke JF, Zhang X. Speech enhancement under low SNR conditions via noise estimation using sparse and low-rank NMF with Kullback-Leibler divergence. *IEEE/ACM Trans Audio Speech Lang Process*. 2015;23(7):1233-1242. <https://doi.org/10.1109/TASLP.2015.2427520>
28. Nowozin S, Cseke B, Tomioka R. f-gan: training generative neural samplers using variational divergence minimization. In: Proceedings of the Conference on Advances in Neural Information Processing Systems; 2016; Barcelona, Spain.
29. Hanusanto GA, Roitch V, Kuhn D, Wiesemann W. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*. 2017;65(3):751-767. <https://doi.org/10.1287/opre.2016.1583>
30. Najm HN, Berry RD, Safta C, Sargsyan K, Debusschere BJ. Data-free inference of uncertain parameters in chemical models. *Int J Uncertain Quantification*. 2014;4(2):111-132. <https://doi.org/10.1615/Int.J.UncertaintyQuantification.2013005679>
31. Najm HN, Chowdhary K. Inference given summary statistics. In: Ghanem R, Higdon D, Owhadi H, eds. *Handbook of Uncertainty Quantification*. Cham, Switzerland: Springer; 2017:33-67.
32. Nguyen B, Morell C, De Baets B. Supervised distance metric learning through maximization of the Jeffrey divergence. *Pattern Recognition*. 2017;64:215-225. <https://doi.org/10.1016/j.patcog.2016.11.010>
33. Wetzel SJ. Unsupervised learning of phase transitions: from principal component analysis to variational autoencoders. *Phys Rev E*. 2017;96(2):022140. <https://doi.org/10.1103/PhysRevE.96.022140>
34. Fan M, Zhang X, Du L, Chen L, Tao D. Semi-supervised learning through label propagation on geodesics. *IEEE Trans Cybern*. 2018;48(5):1486-1499. <https://doi.org/10.1109/TCYB.2017.2703610>
35. Saleem N, Ijaz G. Low rank sparse decomposition model based speech enhancement using gammatone filterbank and Kullback-Leibler divergence. *Int J Speech Technol*. 2018;21(2):217-231. <https://doi.org/10.1007/s10772-018-9500-2>
36. Xu M, Han M, Chen CP, Qiu T. Recurrent broad learning systems for time series prediction. *IEEE Trans Cybern*. 2018;99:1-13. <https://doi.org/10.1109/TCYB.2018.2863020>
37. Soize C, Ghanem R. Data-driven probability concentration and sampling on manifold. *J Comput Phys*. 2016;321:242-258. <https://doi.org/10.1016/j.jcp.2016.05.044>
38. Soize C, Ghanem R. Polynomial chaos representation of databases on manifolds. *J Comput Phys*. 2017;335:201-221. <https://doi.org/10.1016/j.jcp.2017.01.031>
39. Ghanem R, Soize C. Probabilistic nonconvex constrained optimization with fixed number of function evaluations. *Int J Numer Methods Eng*. 2018;113(4):719-741.
40. Soize C, Ghanem RG, Safta C, et al. Entropy-based closure for probabilistic learning on manifolds. *J Comput Phys*. 2019;388:528-533. <https://doi.org/10.1016/j.jcp.2018.12.029>
41. Ghanem R, Soize C, Thimmisetty C. Optimal well-placement using probabilistic learning. *Data Enabled Discov Appl*. 2018;2(1):4-16. <https://doi.org/10.1007/s41688-017-0014-x>
42. Ghanem RG, Soize C, Safta C, et al. Design optimization of a scramjet under uncertainty using probabilistic learning on manifolds. *J Comput Phys*. 2019.
43. Soize C, Farhat C. Probabilistic learning for modeling and quantifying model-form uncertainties in nonlinear computational mechanics. *Int J Numer Methods Eng*. 2019;117:819-843.
44. Soize C. Polynomial chaos expansion of a multimodal random vector. *SIAM/ASA J Uncertain Quantification*. 2015;3(1):34-60. <https://doi.org/10.1137/140968495>
45. Bowman A, Azzalini A. *Applied Smoothing Techniques for Data Analysis*. Oxford, UK: Oxford University Press; 1997.
46. Shannon CE. A mathematical theory of communication. *Bell Syst Technol J*. 1948;27(14):379-423 & 623-659.
47. Jaynes E. Information theory and statistical mechanics II. *Physical Review*. 1957;108(2):171-190.

48. Cover TM, Thomas JA. *Elements of Information Theory*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2006.
49. Soize C. *Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering*. New York, NY: Springer; 2017.
50. Guilleminot J, Soize C. On the statistical dependence for the components of random elasticity tensors exhibiting material symmetry properties. *J Elast.* 2013;111(2):109-130. <https://doi.org/10.1007/s10659-012-9396-z>
51. Agmon N, Alhassid Y, Levine RD. An algorithm for finding the distribution of maximal entropy. *J Comput Phys.* 1979;30(2):250-258.
52. Batou A, Soize C. Calculation of Lagrange multipliers in the construction of maximum entropy distributions in high stochastic dimension. *SIAM/ASA J Uncertain Quantification.* 2013;1(1):431-451.
53. Soize C. Construction of probability distributions in high dimension using the maximum entropy principle: applications to stochastic processes, random fields and random matrices. *Int J Numer Methods Eng.* 2008;76(10):1583-1611.
54. Soize C. *The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions*. Singapore: World Scientific Publishing; 1994.
55. Coifman R, Lafon S, Lee A, et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc Natl Acad Sci.* 2005;102(21):7426-7431.
56. Soize C. Non-Gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators. *Comput Methods Appl Mech Eng.* 2006;195(1-3):26-64. <https://doi.org/10.1016/j.cma.2004.12.014>
57. Shinozuka M. Simulation of multivariate and multidimensional random processes. *J Acoust Soc Am.* 1971;49(1B):357-368. <https://doi.org/10.1121/1.1912338>

How to cite this article: Soize C, Ghanem R. Physics-constrained non-Gaussian probabilistic learning on manifolds. *Int J Numer Methods Eng.* 2020;121:110–145. <https://doi.org/10.1002/nme.6202>

APPENDIX A

EXAMPLES OF CONSTRAINTS

In this appendix, two examples of constraints are given in addition to the those given in Section 4.

A.1 | Learning from a training data set under constraints defined by experimental second-order moments given for components of \mathbf{Q}^c and by the mean value of a physical constraint

This example corresponds to the concatenation of the two examples presented in Sections 4.1 and 4.2. Consequently, using the notations introduced in these two examples, $m_c = 2\kappa_{\text{mom}} + n_q$, function $\mathbf{x} \mapsto \mathbf{g}^c(\mathbf{x}) = (g_1^c(\mathbf{x}), \dots, g_{m_c}^c(\mathbf{x}))$ from \mathbb{R}^n into \mathbb{R}^{m_c} , and vector $\beta^c = (\beta_1^c, \dots, \beta_{m_c}^c) \in \mathbb{R}^{m_c}$, which define the constraints (see Equation (10)), are such that, for $\kappa = 1, \dots, \kappa_{\text{mom}}$ and for $k = 1, \dots, n_q$,

$$g_\kappa^c(\mathbf{X}^c) = Q_{k_\kappa}^c, \quad \beta_\kappa^c = m_\kappa^{\text{exp}}, \quad (\text{A1})$$

$$g_{\kappa+\kappa_{\text{mom}}}^c(\mathbf{X}^c) = (Q_{k_\kappa}^c)^2, \quad \beta_{\kappa+\kappa_{\text{mom}}}^c = \mathbb{R}_\kappa^{\text{exp}}, \quad (\text{A2})$$

$$g_{k+2\kappa_{\text{mom}}}^c(\mathbf{X}^c) = Q_k^c - ([\underline{B}] \mathbf{W}^c)_k, \quad \beta_{k+2\kappa_{\text{mom}}}^c = \beta_k^{\text{const}}. \quad (\text{A3})$$

A.1.1 | Probability density functions given for components of \mathbf{Q}^c

Let $\mathcal{J}_{\text{pdf}} = \{k_1, \dots, k_{\kappa_{\text{pdf}}}\}$ be $\kappa_{\text{pdf}} \leq n_q$ integers such that $\mathcal{J}_{\text{pdf}} \subseteq \{1, \dots, n_q\}$. For $\kappa \in \{1, \dots, \kappa_{\text{pdf}}\}$, let $q \mapsto p_\kappa^{\text{exp}}(q)$ be the given experimental pdf assigned to component $Q_{k_\kappa}^c$ of random vector \mathbf{Q}^c , which is assumed to be defined by its values $\{p_\kappa^{\text{exp}}(q_{\kappa i}), i = 1, \dots, m_\kappa\}$ that are given in m_κ points $q_{\kappa 1}, \dots, q_{\kappa m_\kappa}$ in \mathbb{R} . We then have to take into account the $m_c = \sum_{\kappa=1}^{\kappa_{\text{pdf}}} m_\kappa$ constraints that could be written, for $\kappa = 1, \dots, \kappa_{\text{pdf}}$, as

$$p_{X_{k_\kappa}^c}(q_{\kappa i}) = p_\kappa^{\text{exp}}(q_{\kappa i}), \quad i = 1, \dots, m_\kappa. \quad (\text{A4})$$

Because $p_{X_{k_\kappa}^c}(q_{\kappa i}) = E\{\delta_0(Q_{k_\kappa}^c - q_{\kappa i})\}$, the Dirac δ_0 can be regularized and Equation (A4) is replaced by the following approximation:

$$E\{\tilde{g}_{\kappa i}^c(\mathbf{X}^c)\} = \tilde{\beta}_{\kappa i}^c, \quad i = 1, \dots, m_\kappa, \quad (\text{A5})$$

with

$$\tilde{g}_{\kappa i}^c(\mathbf{X}^c) = \frac{1}{\sqrt{2\pi}\varepsilon_\kappa} \exp\left\{-\frac{1}{2\varepsilon_\kappa^2}(Q_{k_\kappa}^c - q_{\kappa i})^2\right\}, \quad \tilde{\beta}_{\kappa i}^c = p_\kappa^{\text{exp}}(q_{\kappa i}), \quad (\text{A6})$$

in which an estimation of the hyperparameter ε_κ is detailed below. Consequently, function \mathbf{g}^c and vector $\boldsymbol{\beta}^c$ is obtained by gathering $\tilde{g}_{\kappa i}^c$ and $\tilde{\beta}_{\kappa i}^c$ for all the κ in $\{1, \dots, \kappa_{\text{pdf}}\}$ and for all i in $\{1, \dots, m_\kappa\}$,

$$\mathbf{g}^c(\mathbf{X}^c) = \{\tilde{g}_{\kappa i}^c(\mathbf{X}^c)\}_{\kappa i}, \quad \boldsymbol{\beta}^c = \{\tilde{\beta}_{\kappa i}^c\}_{\kappa i}, \quad m_c = \sum_{\kappa=1}^{\kappa_{\text{pdf}}} m_\kappa. \quad (\text{A7})$$

In Equation (A6), the hyperparameter ε_κ can be estimated as follows. Because $p_{Q_{k_\kappa}^c}(q_{\kappa i}) = p_\kappa^{\text{exp}}(q_{\kappa i})$ is the value of the closest pdf to $p_{Q_{k_\kappa}}$, ε_κ can be estimated using the Gaussian kernel-density estimation method of $p_{Q_{k_\kappa}}(q_{\kappa i})$ constructed with the initial data set $D_N = \{(\mathbf{q}^j, \mathbf{w}^j), j = 1, \dots, N\}$ in which the component k_κ of \mathbf{q}^j is $q_{k_\kappa}^j$. We then have

$$p_{Q_{k_\kappa}}(q_{\kappa i}) = \frac{1}{N} \sum_{j=1}^N \frac{1}{\sqrt{2\pi}s_s\hat{\sigma}_\kappa} \exp\left\{-\frac{1}{2s_s^2\hat{\sigma}_\kappa^2}(q_{k_\kappa}^j - q_{\kappa i})^2\right\}, \quad (\text{A8})$$

in which $s_s = (4/(3N))^{1/5}$ is the Sylverman bandwidth and where $\hat{\sigma}_\kappa$ is the estimate of the standard deviation of Q_{k_κ} performed with D_N . For N sufficiently large, the right-hand side of Equation (A8) is

$$E\left\{\frac{1}{\sqrt{2\pi}s_s\hat{\sigma}_\kappa} \exp\left\{-\frac{1}{2s_s^2\hat{\sigma}_\kappa^2}(Q_{k_\kappa}^c - q_{\kappa i})^2\right\}\right\},$$

which compares to Equations (A5) and (A6), yielding $\varepsilon_\kappa = s_s\hat{\sigma}_\kappa$.

APPENDIX B

STÖRMER-VERLET ALGORITHM FOR SOLVING THE REDUCED-ORDER NONLINEAR ISDE DEFINED BY EQUATIONS (94) TO (98)

The number M of realizations is reparameterized as

$$M = n_{\text{MC}} \times N, \quad (\text{B1})$$

in which n_{MC} is the integer that has been introduced in Section 5.7.5. The reduced-order ISDE defined by Equations (94) to (98) is solved for $t \in [0, t_{\max}]$ with $t_{\max} = (\ell_0 + n_{\text{MC}} \times M_0) \Delta t$ in which Δt is the sampling step, where ℓ_0 is chosen in order that the solution of Equations (94) to (96) has reached the stationary regime, and where M_0 allows for controlling the number of sampling steps between two consecutive realizations for obtaining a quasi-independence of the computed realizations. For $\ell = 0, 1, \dots, n_{\text{MC}} \times M_0$, we consider the sampling points $t_\ell = \ell \Delta t$ and the following notations: $[\mathcal{Z}_\ell] = [\mathcal{Z}_{\lambda'}(t_\ell)]$, $[\mathcal{Y}_\ell] = [\mathcal{Y}_{\lambda'}(t_\ell)]$, and $[\mathcal{W}_\ell^{\text{wien}}] = [\mathcal{W}_{\lambda'}^{\text{wien}}(t_\ell)]$. The Störmer-Verlet scheme is used for solving the reduced-order ISDE, which is written, for $\ell = 0, 1, \dots, n_{\text{MC}} \times M_0$, as

$$\begin{aligned} \left[\mathcal{Z}_{\ell+\frac{1}{2}}\right] &= [\mathcal{Z}_\ell] + \frac{\Delta t}{2} [\mathcal{Y}_\ell], \\ [\mathcal{Y}_{\ell+1}] &= \frac{1-\beta}{1+\beta} [\mathcal{Y}_\ell] + \frac{\Delta t}{1+\beta} \left[\mathcal{L}_{\ell+\frac{1}{2}}\right] + \frac{\sqrt{f_0}}{1+\beta} \left[\Delta \mathcal{W}_{\ell+1}^{\text{wien}}\right], \\ [\mathcal{Z}_{\ell+1}] &= \left[\mathcal{Z}_{\ell+\frac{1}{2}}\right] + \frac{\Delta t}{2} [\mathcal{Y}_{\ell+1}], \end{aligned}$$

with the initial condition defined by Equation (96), where $\beta = f_0 \Delta t / 4$, and where $[\mathcal{L}_{\ell+\frac{1}{2}}]$ is the $\mathbb{M}_{v,m}$ -valued random variable such that

$$[\mathcal{L}_{\ell+\frac{1}{2}}] = [\mathcal{L}_{\lambda'} ([\mathcal{Z}_{\ell+\frac{1}{2}}])] = [L_{\lambda'} ([\mathcal{Z}_{\ell+\frac{1}{2}}] [g]^T)] [a].$$

In the above equation, $[\Delta \mathbf{W}_{\ell+1}^{\text{wien}}] = [\Delta \mathbf{W}_{\ell+1}^{\text{wien}}][a]$ is a random variable with values in $\mathbb{M}_{v,m}$, in which the increment $[\Delta \mathbf{W}_{\ell+1}^{\text{wien}}] = [\mathbf{W}_{\lambda'}^{\text{wien}}(t_{\ell+1})] - [\mathbf{W}_{\lambda'}^{\text{wien}}(t_\ell)]$. The increments are statistically independent. For all $\alpha = 1, \dots, v$ and for all $j = 1, \dots, N$, the real-valued random variables $\{[\Delta \mathbf{W}_{\ell+1}^{\text{wien}}]_{\alpha j}\}_{\alpha j}$ are independent, Gaussian, second-order, and centered random variables such that

$$E \left\{ [\Delta \mathbf{W}_{\ell+1}^{\text{wien}}]_{\alpha j} [\Delta \mathbf{W}_{\ell+1}^{\text{wien}}]_{\alpha' j'} \right\} = \Delta t \delta_{\alpha \alpha'} \delta_{j j'}.$$

APPENDIX C

CONSTRUCTION OF THE NON-GAUSSIAN RANDOM FIELD $[G_0]$ AND ITS GENERATOR OF REALIZATIONS

In this appendix, we summarize the construction of the non-Gaussian random field $[G_0]$ and its generator of realizations (whose details can be found in the works of Soize^{49,56}). This construction starts with the introduction of a Gaussian random field.

C.1 | Gaussian random field U allowing the stochastic germ of random field $[G_0]$ to be constructed

Let $\{U(\zeta), \zeta \in \mathbb{R}^3\}$ be the real-valued second-order, centered, homogeneous, Gaussian, and normalized random field, defined on probability space $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by \mathbb{R}^3 . For all ζ in \mathbb{R}^3 , we then have

$$E\{U(\zeta)\} = 0, \quad E\{U(\zeta)^2\} = 1.$$

The random field U is then completely and uniquely defined by its autocorrelation function

$$\chi = (\chi_1, \chi_2, \chi_3) \mapsto R_U(\chi) = E\{U(\zeta + \chi) U(\zeta)\},$$

from \mathbb{R}^3 into \mathbb{R} , such that $R_U(0) = 1$. The spatial-correlation lengths of U , denoted as L_1, L_2, L_3 , are such that

$$L_k = \int_0^{+\infty} |R_U(\chi^k)| d\chi_k, \quad k = 1, 2, 3,$$

in which $\chi^1 = (\chi_1, 0, 0)$, $\chi^2 = (0, \chi_2, 0)$, and $\chi^3 = (0, 0, \chi_3)$. The autocorrelation function is written as

$$\begin{aligned} R_U(\chi) &= \rho_1(\chi_1) \times \rho_2(\chi_2) \times \rho_3(\chi_3), \\ \rho_k(\chi_k) &= \left\{ 4L_k^2 / (\pi^2 \chi_k^2) \right\} \sin^2(\pi \chi_k / (2L_k)), \quad k = 1, 2, 3. \end{aligned}$$

Random field U is mean-square continuous on \mathbb{R}^3 and its power spectral density function defined on \mathbb{R}^3 has a compact support that is written as

$$[-\pi/L_1, \pi/L_1] \times [-\pi/L_2, \pi/L_2] \times [-\pi/L_3, \pi/L_3].$$

Such a model allows for sampling random field U as a function of the spatial-correlation lengths by using the Shannon theorem and has only three real parameters. The details concerning the generation of realizations of random field U can be found in the work of Soize.⁵⁶ One possible method is based on the usual numerical simulation of homogeneous Gaussian vector-valued random field constructed with the stochastic integral representation of homogeneous stochastic fields.⁵⁷

C.1.1 | Defining the random field $[G_0]$ and its generator of realizations

Let δ_{G_0} be such that $0 < \delta_{G_0} < \sqrt{7/11}$. For $1 \leq j \leq k \leq 6$, let $\{U_{jk}(\zeta), \zeta \in \mathbb{R}^3\}_{jk}$ be 21 independent copies of Gaussian random field $\{U(\zeta), \zeta \in \mathbb{R}^3\}$ that is defined in 0.1. The non-Gaussian random field $\{[G_0(\zeta)], \zeta \in \mathbb{R}^3\}$ is then constructed as follows.

- For all ζ fixed in \mathbb{R}^3 , the random matrix $[G_0(\zeta)]$ is written as

$$[G_0(\zeta)] = [\mathbf{L}(\zeta)]^T [\mathbf{L}(\zeta)], \quad (\text{C1})$$

in which $[\mathbf{L}(\zeta)]$ is an upper (6×6) real triangular random matrix.

- For $1 \leq j \leq k \leq 6$, the random fields $\{[\mathbf{L}(\zeta)]_{jk}, \zeta \in \mathbb{R}^3\}$ are independent.
- For $j < k$, the real-valued random field $\{[\mathbf{L}(\zeta)]_{jk}, \zeta \in \mathbb{R}^3\}$ is defined by $[\mathbf{L}(\zeta)]_{jk} = \sigma_6 U_{jk}(\zeta)$ in which σ_6 is such that $\sigma_6 = \delta_{G_0}/\sqrt{6+1}$.
- For $j = k$, the positive-valued random field $\{[\mathbf{L}(\zeta)]_{jj}, \zeta \in \mathbb{R}^3\}$ is defined by $[\mathbf{L}(\zeta)]_{jj} = \sigma_6 \sqrt{2h(U_{jj}(\zeta), a_j)}$ in which $a_j = (6+1)/(2\delta_{G_0}^2) + (1-j)/2$. Function $u \mapsto h(\alpha, u)$ is such that $\Gamma_\alpha = h(\alpha, \mathcal{U})$ is a gamma random variable with parameter α (\mathcal{U} being a normalized Gaussian random variable).