Sampling of Bayesian posteriors with a non-Gaussian probabilistic learning on manifolds from a small dataset

Christian Soize, Roger Ghanem, Christophe Desceliers

Noname manuscript No.
(will be inserted by the editor)

Sampling of Bayesian posteriors with a non-Gaussian probabilistic learning on manifolds from a small dataset

Christian Soize · Roger G. Ghanem · Christophe Desceliers

Received: date / Accepted: date

Abstract This paper tackles the challenge presented by small-data to the task of Bayesian inference. A novel methodology, based on manifold learning and manifold sampling, is proposed for solving this computational statistics problem under the following assumptions: 1) neither the prior model nor the likelihood function are Gaussian and neither can be approximated by a Gaussian measure; 2) the number of functional input (system parameters) and functional output (quantity of interest) can be large; 3) the number of available realizations of the prior model is small, leading to the small-data challenge typically associated with expensive numerical simulations; the number of experimental realizations is also small; 4) the number of the posterior realizations required for decision is much larger than the available initial dataset. The method and its mathematical aspects are detailed. Three applications are presented for validation: The first two involve mathematical constructions aimed to develop intuition around the method and to explore its performance. The third example aims to demonstrate the operational value of the method using a more complex application related to the statistical inverse identification of the non-Gaussian matrix-valued random elasticity field of a damaged biological tissue (osteoporosis in a cortical bone) using ultrasonic waves.

Keywords Probabilistic learning · Bayesian posterior · Non-Gaussian · Manifolds · Machine learning · Data driven · Small dataset · Uncertainty quantification · UQ · Data sciences

C. Soize
Université Paris-Est Marne-la-Vallée, Laboratoire Modélisation et Simulation Multi-Echelle, MSME UMR 8208, 5 bd Descartes, 77454 Marne-la-Vallée, France
Tel.: +33-1-60957661, Fax: ++33-1-60957799
E-mail: christian.soize@u-pem.fr

R.G. Ghanem
University of Southern California, 210 KAP Hall, Los Angeles, CA 90089, United States
E-mail: ghanem@usc.edu

C. Desceliers
Université Paris-Est Marne-la-Vallée, Laboratoire Modélisation et Simulation Multi-Echelle, MSME UMR 8208, 5 bd Descartes, 77454 Marne-la-Vallée, France
E-mail: christophe.desceliers@u-pem.fr

1 Introduction

1.1 Overview of the Bayesian approach

The Bayesian approach is a very powerful statistical tool that provides a rigorous formulation for statistical inverse problems and about which numerous papers and treatises have been published [68,34,38,71,12,10,70,52,14,25,61]. In general, this approach requires the use of variants of the Markov Chain Monte Carlo (MCMC) methods [2] for generating realizations (samples) of the posterior model given a prior model and data typically derived either from numerical simulations or from experimental measurements. This probabilistic approach is extensively used in many fields of physical and life sciences, computational and engineering sciences, and also in machine learning [43,57,74] and in algorithms devoted to artificial intelligence [37,23].

In the supervised case, the most popular Bayesian inversion approach consists of constructing the likelihood function using a Gaussian model. For instance, using the output predictive error, the conditional probability density function (pdf) of the random quantity of interest, $\mathbb{Q}$, given a value $\mathbb{w}$ of the random parameter $\mathbb{W}$, is constructed using the equation $\mathbb{Q} = \mathbb{f}(\mathbb{W}) + \mathbb{B}$ in which $\mathbb{B}$ is a Gaussian random vector that accounts for modeling errors introduced during the construction of the mathematical/computational model of the system (represented by the deterministic mapping $\mathbb{f}$) and/or the experimental measurements errors. Although generally

more efficient than their alternatives, MCMC generators for sampling from the posterior distribution [30,61], still require a large number of calls to the computational model, which can present insurmountable difficulties for expensive models, specially when dealing with high-dimensional problems (functional inputs/outputs). Generally, this situation requires the introduction of a surrogate model for $\mathbb{f}$ in order to decrease the numerical cost such as the Gaussian-process surrogate model including Gaussian-process regression and linearization techniques (see for instance [35,36,55] for calibration of computer models, [6,48] for formulations using Gaussian processes, and [22,50,33,69,75] for algorithms adapted to large-scale inverse problems in the Gaussian likelihood framework).

Nevertheless, the additive Gaussian noise model for the likelihood is not always sufficient and embedded models have to be considered for the likelihood. Consequently, the Bayesian approach becomes much more computationally taxing, in particular for high-dimension where it can become outright prohibitive. This is the case if $\mathbb{Q} = \mathbb{f}(\mathbb{W})$ is replaced by $\mathbb{Q} = \mathbb{f}(\mathbb{W}, \mathbb{U})$ in which $\mathbb{U}$ is a random vector. For instance, $\mathbb{U}$ corresponds to the spatial discretization of a non-Gaussian tensor-valued random field appearing as a coefficient in a partial differential operator. In such a case, the conditional probability density function of $\mathbb{Q} = \mathbb{f}(\mathbb{w}, \mathbb{U})$, given $\mathbb{W} = \mathbb{w}$, involves solving the forward problem for several realizations of $\mathbb{U}$. A number of procedures have been proposed in recent years to tackle this challenge, ranging from adapted representations [72,73], to reduced-order models, and surrogate models (see for instance, [32,11,9,20,54] for reduced-order models and [42,45,8,62] for stochastic reduced-order models). Many methods based on the use of polynomial chaos expansions have also been developed (see for instance, [24,4,17,61] for the identification of stochastic system parameters and random fields in stochastic boundary value problems, [40,39,53,73] for Bayesian inference in inverse problems, and [44,29] for explicit construction of surrogate models).

The Bayesian approach for parameter estimation in the non-Gaussian embedded likelihood case has significantly been developed for low dimension [47,3] and using filtering techniques and functional approximations [31,46,41,1]. Recently, a nonparametric Bayesian approach for non-Gaussian cases has been proposed [49] for which the invertible covariance matrix of the Gaussian kernel-density estimation is optimized by taking into account the unknown block dependence structure.

The high-dimension case concerns functional inputs/outputs for which the numerical cost of the computational model is expensive (the one considered in this paper). Concerning the case of a high dimensional target for which a large number of realizations are also required, methods have been proposed [13,5].

1.2 Framework of the developments, difficulties involved, and novelties

This paper is devoted to the Bayesian inference for the small-data challenge using the probabilistic learning on manifolds (PLoM). The PLoM method has been introduced in [63]. Complementary developments can be found in [26,64–66]. Applications and validations can be found in [27,28,67,62]. The small-data character is related to the given initial dataset. We clarify hereinafter the role-played by the initial dataset in the proposed methodology, using for explaining it an example and detailing the difficulties involved that contribute to demonstrate the novelty of the approach. Let us consider the statistical inverse problem for identifying a posterior probability measure of a non-Gaussian random field. This random field is the functional input that is the parameter in infinite dimension of a partial differential operator of a stochastic boundary value problem (BVP). Partial observations are done on the functional output (the solution of the BVP), which is also in high dimension. Only a few experimental realizations are available (small number of experiments). Consequently, this Bayesian inference is in high dimension (functional input and functional output), with small experimental data for the observations. The problem consists of estimating the posterior probability measure of the functional input using the experimental data for the functional output while preserving the non-Gaussian character of the likelihood function. The fundamental assumption used, which justifies the proposed novel method, is as follows. The numerical cost of an evaluation of one realization of the random output, knowing one realization of the random input, using the computational model of the BVP, is very large. Only a very limited number of evaluations (tens, or at most hundreds) can be carried out. This number is too small to use the Bayes method in very high dimensions. Classically, a surrogate model must then be constructed. Taking into account the high-dimension character of this surrogate model (functional input – functional output) and the limitation of the number of evaluations, any algebraic representation cannot be constructed. We thus propose an alternative approach based on the use of the PLoM, which allows for generating a large number of additional realizations of the random couple (input – output) from the knowledge of an initial dataset of this couple. These realizations are calculated using an informative prior model of the random input and the computational model of the BVP. This initial dataset is small due to the limited number of evaluations. As it has been shown in [63], a small initial dataset allows for constructing a probabilistic surrogate model without algebraic representation. For solving this problem, the PLoM method published in [63] requires key modifications. It is necessary to introduce several novel ingredients in the methodology that are detailed in Section 2.

Introducing a mathematical framework, we now synthesize the novel approach dedicated to the high-dimension case in the context of the small-data challenge. We consider the case $\mathbb{Q} = \mathbb{f}(\mathbb{W}, \mathbb{U})$ in which $\mathbb{W}$, $\mathbb{U}$, $\mathbb{Q}$ are random variables with values in $\mathbb{R}^{n_w}$, $\mathbb{R}^{n_u}$, $\mathbb{R}^{n_q}$, and where $(\mathbb{w}, \mathbb{u}) \mapsto \mathbb{f}(\mathbb{w}, \mathbb{u})$ is a nonlinear mapping. In addition to the mapping $\mathbb{f}$, only two pieces of information are available. The first one consists of an *initial dataset* (the training set), $\mathbb{D}_{N_d}$, made up of N_d independent realizations (samples) $\{(\mathbb{q}^j, \mathbb{w}^j), j = 1, \ldots, N_d\}$ of the couple of random vectors, $(\mathbb{Q}, \mathbb{W})$. The second piece of information consists of an *experimental dataset*, $\mathbb{D}^{\text{exp}}_{n_r}$, used for updating, and consisting of n_r given independent experimental realizations (measures or simulations) $\{(\mathbb{q}^{\text{exp},r}, r = 1, \ldots, n_r\}$ of $\mathbb{Q}$. The objective then is to construct, using the Bayesian approach, a set of ν_{post} realizations, $\{\mathbb{w}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ of the posterior random vector, denoted by $\mathbb{W}^{\text{post}}$. The following requirements have guided the development of the proposed novel methodology.

1. The non-Gaussian case is considered. The conditional probability distribution of $\mathbb{Q}$ given $\mathbb{W} = \mathbb{w}$ is not Gaussian. For instance, mapping $\mathbb{f}$ is not additive with respect to the Gaussian random vector $\mathbb{U}$, contrarily to the case for which the output-predictive-error formulation is used, which consists in adding to $\mathbb{Q} = \mathbb{f}(\mathbb{W})$ a Gaussian noise $\mathbb{U}$.
2. The problem is in high dimension, n_w is large and n_q can also be large.
3. The number N_d of realizations of the prior model is small, which means that we are in the context of the small-data challenge. This situation can be induced, for instance, by the use of an expensive computer code for generating the set $\mathbb{D}_{N_d}$ of realizations.
4. The number n_r of experimental realizations is small.
5. The number ν_{post} of the posterior realizations required for decision is large.

1.3 Organization of the paper

In order to discuss and motivate the intricate interplay between the requirements presented in Section 1.2 necessary details concerning PLoM and the various modeling choices are included in the paper, which is organized as follows.
Section 3 is devoted to the mathematical statement of the problem. In Section 4, we introduce the scaling of the initial dataset. The scaling of the random vector $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$ is denoted by $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$. Section 5 deals with the generation of additional realizations for the prior probability model using the PLoM while the reduced-order representations for $\mathbf{Q}$ and $\mathbf{W}$ are constructed in Section 6 using the learned dataset. Section 7 is devoted to the Bayesian formulation for the posterior model and Section 8 deals with the nonparametric statistical estimation of the posterior pdf using the learned dataset, for which a regularization model is proposed. The dissipative Hamiltonian MCMC generator, which is a nonlinear Itô stochastic differential equation (ISDE), is detailed in Section 9 for the posterior pdf. The question relative to the choice of a value of the regularization parameter is analyzed in Section 10. In Section 11, we summarize the main steps for implementing the algorithm. Three applications are presented in Sections 12 and 13. The first two are relatively simple and can easily be reproduced. The third application is devoted to the ultrasonic wave propagation in biological tissues for which $\mathbb{W}$ is the random vector corresponding to the spatial discretization of a non-Gaussian tensor-valued random elasticity field of a cortical bone exhibiting osteoporosis. In order to retain clarity throughout the paper, several of the mathematical and algorithmic details have been relegated to 9 appendices. For reasons of limitation of the number of pages, the first two applications and all the appendices have been moved to the **Supplementary material**.

Notations

x: lower-case letters are deterministic real variables.
$\mathbf{x}$: boldface lower-case letters are deterministic vectors.
$\mathbb{x}$: lower-case blackboard letters are deterministic vectors.
X: upper-case letters are real-valued random variables.
$\mathbf{X}$: boldface upper-case letters are random vectors.
$\mathbb{X}$ upper-case blackboard letters are random vectors.
$[x]$: lower-case letters between brackets are matrices.
$[\mathbf{X}]$: boldface upper-case letters between brackets are random matrices.

n: dimension ($n = n_q + n_w$) of vector $\mathbf{x}$ or $\mathbb{X}$.
n_q: dimension of vectors $\mathbf{q}$, $\mathbb{q}$, $\mathbf{Q}$, and $\mathbb{Q}$.
n_r: number of independent experimental realizations.
n_w: dimension of vectors $\mathbf{w}$, $\mathbb{w}$, $\mathbf{W}$, and $\mathbb{W}$.
ν: dimension ($\nu = \nu_q + \nu_w$) of vectors $\widehat{\mathbf{x}}$ and $\widehat{\mathbf{X}}$.
ν_{ar}: number of additional realizations.
ν_{post}: number of independent realizations for the posterior.
ν_q: dimension of vectors $\widehat{\mathbf{q}}$ and $\widehat{\mathbf{Q}}$.
ν_w: dimension of vectors $\widehat{\mathbf{w}}$ and $\widehat{\mathbf{W}}$.
$[I_n]$: identity matrix in $\mathbb{M}_n$.
$\mathbb{M}_{n,N}$: set of the $(n \times N)$ real matrices.
$\mathbb{M}_n$: set of the square $(n \times n)$ real matrices.
$\mathbb{M}_n^+$: set of the positive-definite $(n \times n)$ real matrices.
$\mathbb{M}_n^{+0}$: set of the positive-semidefinite $(n \times n)$ real matrices.
$\mathbb{R}$: set of the real numbers.
$\mathbb{R}^n$: Euclidean vector space on $\mathbb{R}$ of dimension n.
$[y]_{kj}$: entry of matrix $[y]$.
$[y]^T$: transpose of matrix $[y]$.
E: Mathematical expectation.
$\|\mathbf{x}\|$: usual Euclidean norm in $\mathbb{R}^n$.

$<\mathbf{x},\mathbf{y}>$: usual Euclidean inner product in $\mathbb{R}^n$.
$\|[A]\|_F$: Frobenius norm of a real matrix $[A]$.
$\delta_{kk'}$: Kronecker's symbol.
pdf: probability density function.
MCMC: Markov Chain Monte Carlo.

2 Outline of the novel method

In order to improve numerical conditioning, the initial dataset $\mathbb{D}_{N_d}$ is scaled using an adapted affine transformation into a dataset D_{N_d} made up of the N_d independent realizations $\{(\mathbf{q}^j,\mathbf{w}^j), j=1,\ldots,N_d\}$ of the scaled random variables $(\mathbf{Q},\mathbf{W})$ with values in $\mathbb{R}^{n_q}\times\mathbb{R}^{n_w}$. Using this same affine transformation, experimental dataset $\mathbb{D}^{\exp}_{n_r}$ is transformed into a scaled experimental dataset $D^{\exp}_{n_r}$ made up of the n_r independent realizations $\{\mathbf{q}^{\exp,r}, r=1,\ldots,n_r\}$.

Each of the requirements listed in Section 1.2 presents its own significant challenges which are addressed throughout the paper.

(i)- For addressing the small-data challenge, the PLoM, which has been introduced in [63], is used. This PLoM allows for generating a *learned dataset* (big dataset) $D_{\nu_{\text{ar}}}$ of ν_{ar} additional realizations of the prior model of the scaled random vector $(\mathbf{Q},\mathbf{W})$ in which the number ν_{ar} can be arbitrarily large ($\nu_{\text{ar}}\gg N_d$), using only information defined by the scaled initial dataset D_{N_d}. The convergence in probability distribution of the learning with respect to N_d is investigated. This learned dataset $D_{\nu_{\text{ar}}}$ allows for constructing an accurate estimate of the posterior distribution.

(ii)- For addressing the high-dimension data challenge, a novel approach is proposed. Two reduced-order representations are separately constructed, one for random vector $\mathbf{Q}$ and another one for random vector $\mathbf{W}$, using for each one a principal component analysis (PCA) based on their covariance matrix estimated with the ν_{ar} additional realizations that are extracted from the learned dataset $D_{\nu_{\text{ar}}}$. Random vector $\mathbf{Q}$ (resp. $\mathbf{W}$) is then transformed into a random vector $\widehat{\mathbf{Q}}$ (resp. $\widehat{\mathbf{W}}$) with values in $\mathbb{R}^{\nu_q}$ (resp. in $\mathbb{R}^{\nu_w}$). In general, but depending on the application, the reduced dimensions are such that $\nu_q\ll n_q$ and $\nu_w\ll n_w$. It should be noted that a direct construction by PCA of a reduced-order representation of random vector $\mathbf{X}=(\mathbf{Q},\mathbf{W})$ cannot be done because we need to have a separate representation for the projected random variable $\widehat{\mathbf{Q}}$ and for its counterpart $\widehat{\mathbf{W}}$ in order to be able to write Bayes formula. Consequently, the random vector $\widehat{\mathbf{Q}}$ (resp. $\widehat{\mathbf{W}}$) is centered with an empirical-estimated covariance matrix that is the identity matrix $[I_{\nu_q}]$ (reps. $[I_{\nu_w}]$). The centered random variables $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{W}}$, which are statistically dependent, are then correlated. This means that the empirical-estimated covariance matrix $[C_{\widehat{X}}]$ of random vector $\widehat{\mathbf{X}}=(\widehat{\mathbf{Q}},\widehat{\mathbf{W}})$ is not a diagonal matrix. The (2×2) block writing of $[C_{\widehat{X}}]$ (with respect to $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{W}}$) exhibits two block diagonal identity matrices, namely $[I_{\nu_q}]$ and $[I_{\nu_w}]$, but there are extradiagonal block matrices that, in general, are not equal to zero. At this stage, there is an additional difficulty that is related to the fact that, in general, matrix $[C_{\widehat{X}}]$ is singular or is not sufficiently well conditioned to carry out the algebraic manipulations necessary for the construction of the posterior pdf based on the use of the Gaussian kernel-density estimation method, using the learned dataset $D_{\nu_{\text{ar}}}$. Most often, in the literature, either the rank of $[C_{\widehat{X}}]$ is assumed to be less than $\nu=\nu_q+\nu_w$ (in this case, adapted algebraic methods have been proposed) or matrix $[C_{\widehat{X}}]$ is assumed to be invertible (in that case, there is no difficulty). However, no adapted method seems to have been proposed for the "intermediate" case. Therefore, we had to develop a novel regularization $[\widehat{C}_\varepsilon]$ of $[C_{\widehat{X}}]$ in order to achieve the required robustness.

(iii)- To ensure the robustness of proposed methodology, several novel ingredients have been analyzed, tested, and validated.

- The first one (as explained above) is related, if necessary, to the construction of a regularization $[\widehat{C}_\varepsilon]$ in $\mathbb{M}^+_\nu$ of $[C_{\widehat{X}}]$ in order to obtain a positive-definite inverse matrix $[\widehat{C}_\varepsilon]^{-1}$ whose condition number is of order 1 and for which the value of the hyperparameter ε can be set, independently of applications.

- The second one is related to the construction of the MCMC generator for obtaining a robust algorithm for the computation of the ν_{post} realizations of $\mathbb{W}^{\text{post}}$ whose posterior pdf is $p^{\text{post}}_{\widehat{\mathbf{W}}}$. This pdf is explicitly deduced from the Gaussian kernel-density representation of the joint pdf $p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}$ using the ν_{ar} additional realizations of $(\widehat{\mathbf{Q}},\widehat{\mathbf{W}})$ and the n_r experimental realizations of $\widehat{\mathbf{Q}}$. This MCMC generator is the one (but adapted to the posterior model) used for the PLoM. However, it has been seen through many numerical experiments that a normalization with respect to the covariance matrix of the posterior model $\widehat{\mathbf{W}}^{\text{post}}$ of $\widehat{\mathbf{W}}$ had to be made in order to improve the robustness of the algorithm. Unfortunately, although the expression of $p^{\text{post}}_{\widehat{\mathbf{W}}}$ is explicitly known, the algebraic calculation of this covariance matrix is not possible and, as it will be explained in the following, an approximation has to be constructed. Finally, a statistical reduction along the data axis is performed using a diffusion maps basis in order to avoid a possible scattering of the posterior realizations generated, which then allows for preserving the concentration of the posterior probability measure (when such a concentration exists).

3 Formulation

In this paper, any Euclidean space $\mathcal{E}$ (such as $\mathbb{R}^{n_w}$) 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. In this section, we first detail the mathematical formulation of the problem and we state the objective.

Defining the stochastic mapping $\mathbb{F}$ and the initial dataset $\mathbb{D}_{N_d}$. Let $\mathbb{w} \mapsto \mathbb{F}(\mathbb{w})$ be a stochastic mapping from $\mathbb{R}^{n_w}$ into the space $L^2(\Theta, \mathbb{R}^{n_q})$ of all the second-order random variables defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^{n_q}$. The vector $\mathbb{w}$ (the input) belongs to an admissible set $\mathcal{C}_{\mathbb{w}} \subset \mathbb{R}^{n_w}$ and is modeled by a second-order random variable $\mathbb{W} = (\mathbb{W}_1, \ldots, \mathbb{W}_{n_w})$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^{n_w}$, for which the support of its probability distribution $P_{\mathbb{W}}(d\mathbb{w})$ is $\mathcal{C}_{\mathbb{w}}$, and which is assumed to be statistically independent of $\mathbb{F}$. The quantity of interest (the output) is a random variable $\mathbb{Q} = (\mathbb{Q}_1, \ldots, \mathbb{Q}_{n_q})$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^{n_q}$, which is written as $\mathbb{Q} = \mathbb{F}(\mathbb{W})$, which is statistically dependent of $\mathbb{F}$ and $\mathbb{W}$, and which is assumed to be of second order. For the problem considered, the only available information consists of a given *initial dataset* (training set) constituted of N_d independent realizations $\{(\mathbb{q}_d^j, \mathbb{w}_d^j), j = 1, \ldots, N_d\}$ of random variable $(\mathbb{Q}, \mathbb{W})$ with values in $\mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$.

Example of stochastic mapping $\mathbb{F}$ and origin of the given initial dataset $\mathbb{D}_{N_d}$. The stochastic nature of the mapping $\mathbb{F}$ deserves a clarification. It is induced by the division of the input random parameters of a computational model into two separate subsets only one of which is initially observed, and the influence of the other subset is manifested as uncertainty about the mapping. Thus consider, for instance, a large-scale stochastic computational model of a discretized stochastic physical system for which the random quantity of interest is written as $\mathbb{Q} = \mathbb{f}(\mathbb{W}, \mathbb{U})$. The random variable $\mathbb{U} = (\mathbb{U}_1, \ldots, \mathbb{U}_{n_u})$ is construed as a hidden variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$, with values in $\mathbb{R}^{n_u}$, with probability distribution $P_{\mathbb{U}}(d\mathbb{u})$, and which is statistically independent of $\mathbb{W}$. The function $(\mathbb{w}, \mathbb{u}) \mapsto \mathbb{f}(\mathbb{w}, \mathbb{u})$ is a measurable mapping from $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$ into $\mathbb{R}^{n_q}$, which is a representation of the solution of the stochastic computational model. Consequently, the joint probability distribution $P_{\mathbb{W},\mathbb{U}}(d\mathbb{w}, d\mathbb{u})$ of $\mathbb{W}$ and $\mathbb{U}$ is $P_{\mathbb{W}}(d\mathbb{w}) \otimes P_{\mathbb{U}}(d\mathbb{u})$. For all $\mathbb{w}$ in $\mathbb{R}^{n_w}$, stochastic mapping $\mathbb{F}$ is such that $\mathbb{F}(\mathbb{w}) = \mathbb{f}(\mathbb{w}, \mathbb{U})$. The origin of the initial dataset $\mathbb{D}_{N_d}$ can come from the computation of N_d independent realizations $\{\mathbb{q}_d^j, j = 1, \ldots, N_d\}$ such that $\mathbb{q}_d^j = \mathbb{f}(\mathbb{w}_d^j, \mathbb{U}(\theta_j))$, in which $\{\mathbb{w}_d^j = \mathbb{W}(\theta_j)\}_j$ are N_d independent realizations of $\mathbb{W}$ generated with $P_{\mathbb{W}}(d\mathbb{w})$, and where $\{\mathbb{U}(\theta_j)\}_j$ are N_d independent realizations of $\mathbb{U}$ generated with $P_{\mathbb{U}}(d\mathbb{u})$. It should be noted that realizations $\{\mathbb{U}(\theta_j)\}_j$ are not explicitly included in the initial dataset.

Introducing the random variable $\mathbb{X}$ and its realizations. We then introduce the random variable $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$, with values in $\mathbb{R}^n$ ($n = n_q + n_w$), and for which the probability distribution, $P_{\mathbb{X}}(d\mathbb{x})$, on $\mathbb{R}^n$ is unknown, and for which the initial dataset defined by

$$\mathbb{D}_{N_d} = \{\mathbb{x}_d^j = (\mathbb{q}_d^j, \mathbb{w}_d^j), j = 1, \ldots, N_d\}\,, \tag{1}$$

is the only available information.

Existence hypothesis of probability density function for $\mathbb{X}$. It is assumed that the unknown probability distribution $P_{\mathbb{X}}(d\mathbb{x})$ admits a density $p_{\mathbb{X}}(\mathbb{x})$ with respect to the Lebesgue measure $d\mathbb{x}$ on $\mathbb{R}^n$. Therefore, the joint probability distribution $P_{\mathbb{Q},\mathbb{W}}(d\mathbb{q}, d\mathbb{w})$ on $\mathbb{R}^n$ of $\mathbb{Q}$ and $\mathbb{W}$ admits a density $p_{\mathbb{Q},\mathbb{W}}(\mathbb{q}, \mathbb{w})$ with respect to the Lebesgue measure $d\mathbb{q}\, d\mathbb{w}$ on $\mathbb{R}^n$. The probability distributions $P_{\mathbb{Q}}(d\mathbb{q})$ and $P_{\mathbb{W}}(d\mathbb{w})$ of $\mathbb{Q}$ and $\mathbb{W}$ admit the densities $p_{\mathbb{Q}}(\mathbb{q}) = \int p_{\mathbb{Q},\mathbb{W}}(\mathbb{q}, \mathbb{w})\, d\mathbb{w}$ and $p_{\mathbb{W}}(\mathbb{w}) = \int p_{\mathbb{Q},\mathbb{W}}(\mathbb{q}, \mathbb{w})\, d\mathbb{q}$ with respect to the Lebesgue measures $d\mathbb{q}$ on $\mathbb{R}^{n_q}$ and $d\mathbb{w}$ on $\mathbb{R}^{n_w}$, respectively. The conditional pdf $\mathbb{q} \mapsto p_{\mathbb{Q}|\mathbb{W}}(\mathbb{q}|\mathbb{w})$ on $\mathbb{R}^{n_q}$ of $\mathbb{Q}$ given $\mathbb{W} = \mathbb{w}$ in $\mathcal{C}_{\mathbb{w}} \subset \mathbb{R}^{n_w}$ is such that $p_{\mathbb{Q},\mathbb{W}}(\mathbb{q}, \mathbb{w}) = p_{\mathbb{Q}|\mathbb{W}}(\mathbb{q}|\mathbb{w})\, p_{\mathbb{W}}(\mathbb{w})$. Since the support of $p_{\mathbb{W}}$ is $\mathcal{C}_{\mathbb{w}} \subset \mathbb{R}^{n_w}$, if $\mathbb{w}$ is given in $\mathbb{R}^{n_w} \backslash \mathcal{C}_{\mathbb{w}}$, then $p_{\mathbb{W}}(\mathbb{w}) = 0$, and consequently, $\mathbb{q} \mapsto p_{\mathbb{Q},\mathbb{W}}(\mathbb{q}, \mathbb{w})$ is the zero function. It should be noted that hypothesis $P_{\mathbb{X}}(d\mathbb{x}) = p_{\mathbb{X}}(\mathbb{x})\, d\mathbb{x}$ would not be satisfied if $\mathbb{F}$ was a deterministic mapping, $\mathbb{F}(\mathbb{w}) = \mathbb{f}(\mathbb{w})$ independent of $\mathbb{U}$, because the support, $\mathcal{S}_{n_w} = \{(\mathbb{w}, \mathbb{f}(\mathbb{w})), \mathbb{w} \in \mathcal{C}_{\mathbb{w}} \subset \mathbb{R}^{n_w}\}$ of $P_{\mathbb{X}}(d\mathbb{x})$ on $\mathbb{R}^n$, would be the manifold of dimension n_w in $\mathbb{R}^n$, consisting of the graph of the deterministic mapping $\mathbb{f}$.

Specifying the experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$. An *experimental dataset* $\mathbb{D}_{n_r}^{\text{exp}}$ is given and is constituted of n_r independent experimental realizations of $\mathbb{Q}$,

$$\mathbb{D}_{n_r}^{\text{exp}} = \{\mathbb{q}^{\text{exp},r}, r = 1, \ldots, n_r\}\,, \tag{2}$$

that are also assumed to be independent of realizations $\{\mathbb{q}_d^j\}_j$.

Objective. As explained in Section 1, the objective is to generate realizations $\{\mathbb{w}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ of the posterior model of $\mathbb{W}$ for which the only available information consist of the initial dataset $\mathbb{D}_{N_d}$ associated with a prior model of $\mathbb{W}$ and of the experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$.

4 Scaling the initial dataset

Initial dataset $\mathbb{D}_{N_d}$ can be made up of heterogeneous numerical values and must be scaled for performing computational statistics. Let $\mathbb{x}^{\max} = \max_j\{\mathbb{x}_d^j\}$, $\mathbb{x}^{\min} = \min_j\{\mathbb{x}_d^j\}$, and $\boldsymbol{\beta}_x = \mathbb{x}^{\min}$ be a vector in $\mathbb{R}^n$. The diagonal $(n \times n)$ real matrix $[\alpha_x]_{kk'} = (\mathbb{x}_k^{\max} - \mathbb{x}_k^{\min})\delta_{kk'}$ is invertible. The scaling

of random vector $\mathbb{X}$ with values in $\mathbb{R}^n$ is the random vector $\mathbf{X}$ with values in $\mathbb{R}^n$ such that

$$\mathbb{X} = [\alpha_x]\,\mathbf{X} + \boldsymbol{\beta}_x \quad , \quad \mathbf{X} = [\alpha_x]^{-1}(\mathbb{X} - \boldsymbol{\beta}_x)\,. \tag{3}$$

From Eq. (3), the scaled random variables $\mathbf{Q}$ and $\mathbf{W}$ with values in $\mathbb{R}^{n_q}$ and $\mathbb{R}^{n_w}$ can directly be deduced,

$$\mathbb{Q} = [\alpha_q]\,\mathbf{Q} + \boldsymbol{\beta}_q \quad , \quad \mathbf{Q} = [\alpha_q]^{-1}(\mathbb{Q} - \boldsymbol{\beta}_q)\,, \tag{4}$$

$$\mathbb{W} = [\alpha_w]\,\mathbf{W} + \boldsymbol{\beta}_w \quad , \quad \mathbf{W} = [\alpha_w]^{-1}(\mathbb{W} - \boldsymbol{\beta}_w)\,. \tag{5}$$

The N_d realizations of $\mathbf{X}$ are then $\{\mathbf{x}_d^j\}_j$ with $\mathbf{x}_d^j = [\alpha_x]^{-1}(\mathbb{x}_d^j - \boldsymbol{\beta}_x)$. The scaled initial dataset is then defined by

$$D_{N_d} = \{\mathbf{x}_d^j = (\mathbf{q}_d^j, \mathbf{w}_d^j), j = 1, \ldots, N_d\}\,,$$

in which $\mathbf{q}_d^j = [\alpha_q]^{-1}(\mathbb{q}_d^j - \boldsymbol{\beta}_q)$ and $\mathbf{w}_d^j = [\alpha_w]^{-1}(\mathbb{w}_d^j - \boldsymbol{\beta}_w)$. The collection of these N_d vectors $\{\mathbf{x}_d^j\}_j$ in $\mathbb{R}^n$ is represented by the matrix $[x_d]$ such that

$$[x_d] = [\mathbf{x}_d^1 \ldots \mathbf{x}_d^{N_d}] \in \mathbb{M}_{n,N_d}\,.$$

In the following, we will use the scaled random variable $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ with values in $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$. The experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$ defined in Section 3 is scaled using Eq. (4), yielding the scaled experimental dataset,

$$D_{n_r}^{\text{exp}} = \{\mathbf{q}^{\text{exp},r}, r = 1, \ldots, n_r\}\,,$$

in which $\mathbf{q}^{\text{exp},r} = [\alpha_q]^{-1}(\mathbb{q}^{\text{exp},r} - \boldsymbol{\beta}_q)$. If $\mathbb{Q} = \mathbb{f}(\mathbb{W}, \mathbb{U})$ (see the example of stochastic function $\mathbb{F}$ presented in Section 3), then $\mathbf{Q}$ can be rewritten as

$$\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbb{U})\,, \tag{6}$$

in which $\mathbf{f}$ corresponds to the transformation of mapping $\mathbb{f}$.

5 Generating additional realizations for the prior probability model using the probabilistic learning on manifolds

As explained in Section 1.2, the framework of this paper is the Bayesian approach for the small-data challenge because N_d is assumed to be small. The Bayesian method allows for updating the prior pdf $p_{\mathbf{W}}$ on $\mathbb{R}^{n_w}$ of $\mathbf{W}$ using experimental dataset $D_{n_r}^{\text{exp}}$ relative to $\mathbf{Q}$ with values in $\mathbb{R}^{n_q}$ in order to obtain the posterior pdf $p_{\mathbf{W}}^{\text{post}}$ on $\mathbb{R}^{n_w}$. Clearly, the posterior pdf strongly depends on the joint pdf $p_{\mathbf{Q},\mathbf{W}}$ on $\mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$. Consequently, a bigger dataset $D_{\nu_{\text{ar}}}$ (that we have called "learned dataset" in Section 2),

$$D_{\nu_{\text{ar}}} = \{\mathbf{x}_{\text{ar}}^\ell = (\mathbf{q}_{\text{ar}}^\ell, \mathbf{w}_{\text{ar}}^\ell), \ell = 1, \ldots, \nu_{\text{ar}}\}\,,$$

which is made up of $\nu_{\text{ar}} \gg N_d$ independent realizations of $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$, is required for the two following reasons:
- a better estimate of prior pdf $p_{\mathbf{W}}$ has to be constructed using $D_{\nu_{\text{ar}}}$ instead of D_{N_d}.
- the non-Gaussian conditional pdf $\mathbf{q} \mapsto p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w})$ on $\mathbb{R}^{n_q}$ of $\mathbf{Q}$ for given $\mathbf{W} = \mathbf{w}$ in $\mathbb{R}^{n_w}$ has to be correctly estimated thus requiring a big dataset such as $D_{\nu_{\text{ar}}}$. The use of D_{N_d} for such an estimation would not be sufficiently "good" because N_d is assumed to be small.

In this paper, only D_{N_d} and $D_{n_r}^{\text{exp}}$ are known. In addition, D_{N_d} is assumed to be constituted of numerical simulations performed with a large-scale computational model represented by $\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbb{U})$ (see Eq. (6)) in which $\mathbb{U}$ is not an "observation noise and model discrepancy", but is for instance (as explained in Section 1.2), the spatial discretization of a non-Gaussian tensor-valued random field that appears as a coefficient in a partial differential operator in a stochastic boundary value problem. In this framework, it is important to preserve the non-Gaussian character of the conditional pdf $p_{\mathbf{Q}|\mathbf{W}}(\cdot|\mathbf{w})$, which is the pdf of random vector $\mathbf{f}(\mathbf{w}, \mathbb{U})$. Since $\mathbf{f}$ and $\mathbb{U}$ are unknown (only D_{N_d} is assumed to be known), we propose to construct the big dataset (learned dataset) $D_{\nu_{\text{ar}}}$ of additional realizations using the probabilistic learning on manifolds [63]. In order to facilitate the reading of this paper, a summary of this algorithm is given in **Supplementary material**, Appendix A, in which we propose numerical values and identification methods for the parameters involved in the algorithm.

6 Reduced-order representations for Q and W using the learned dataset

As explained in Section 1.2, dimension $n = n_q + n_w$ of random vector $\mathbf{X}$ can be high. It is thus necessary to decrease the numerical cost of the MCMC generator of $p_{\mathbf{W}}^{\text{post}}$. For that and as explained in Section 2-(ii), a statistical reduction of $\mathbf{Q}$ and $\mathbf{W}$ is performed using a PCA for which the learned dataset $D_{\nu_{\text{ar}}}$ is used.

6.1 PCA of random vector Q

Let $\underline{\mathbf{q}}_{\text{ar}} \in \mathbb{R}^{n_q}$ and $[C_{\mathbf{Q},\text{ar}}] \in \mathbb{M}_{n_q}^{+0}$ be the empirical estimates of the mean vector and the covariance matrix of $\mathbf{Q}$, constructed using the additional realizations $\{\mathbf{q}_{\text{ar}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$. The PCA representation, $\mathbf{Q}^{(\nu_q)}$, of $\mathbf{Q}$ at order $1 \le \nu_q \le \nu_{\text{ar}}$ is written as

$$\mathbf{Q}^{(\nu_q)} = \underline{\mathbf{q}}_{\text{ar}} + [\varphi_q]\,[\mu_q]^{1/2}\,\widehat{\mathbf{Q}}\,, \tag{7}$$

in which $[\varphi_q] \in \mathbb{M}_{n_q,\nu_q}$ is the matrix of the eigenvectors of $[C_{\mathbf{Q},\text{ar}}]$ associated with its ν_q largest eigenvalues $\mu_{q,1} \ge \mu_{q,2} \ge \ldots \ge \mu_{q,\nu_q} > 0$, represented by the diagonal matrix $[\mu_q] \in \mathbb{M}_{\nu_q}$. The value of ν_q is classically calculated in order that the L^2-error function $\nu_q \mapsto \text{err}_{\mathbf{Q}}(\nu_q)$ defined by

$$\text{err}_{\mathbf{Q}}(\nu_q) = \frac{E\{\|\mathbf{Q} - \mathbf{Q}^{(\nu_q)}\|^2\}}{E\{\|\mathbf{Q} - \underline{\mathbf{q}}_{\text{ar}}\|^2\}} = 1 - \frac{\sum_{\alpha=1}^{\nu_q} \mu_{q,\alpha}}{\text{tr}\,[C_{\mathbf{Q},\text{ar}}]}\,, \tag{8}$$

be smaller than $\varepsilon_q > 0$. In Eq. (8), $\mathbf{Q}$ stands for $\mathbf{Q}^{(n_q)}$. Since $[\varphi_q]^T[\varphi_q] = [I_{\nu_q}]$, the random variable $\widehat{\mathbf{Q}}$ with values in $\mathbb{R}^{\nu_q}$ and its ν_{ar} independent realizations are written as

$$\widehat{\mathbf{Q}} = [\mu_q]^{-1/2}\,[\varphi_q]^T(\mathbf{Q} - \underline{\mathbf{q}}_{\text{ar}})\,,$$

$$\widehat{\mathbf{q}}^{\,\ell} = [\mu_q]^{-1/2}\,[\varphi_q]^T(\mathbf{q}^{\ell}_{\text{ar}} - \underline{\mathbf{q}}_{\text{ar}}) \quad , \quad \ell = 1, \ldots, \nu_{\text{ar}}\,. \tag{9}$$

It can then be deduced that the empirical estimate $\underline{\widehat{\mathbf{q}}} \in \mathbb{R}^{\nu_q}$ of the mean vector of $\widehat{\mathbf{Q}}$, and the empirical estimate $[C_{\widehat{\mathbf{Q}}}] \in \mathbb{M}^+_{\nu_q}$ of its covariance matrix are such that

$$\underline{\widehat{\mathbf{q}}} = \mathbf{0} \quad , \quad [C_{\widehat{\mathbf{Q}}}] = [I_{\nu_q}]\,. \tag{10}$$

Therefore, the components $\widehat{Q}_1, \ldots, \widehat{Q}_{\nu_q}$ of $\widehat{\mathbf{Q}}$ are centered and uncorrelated but they are statistically dependent because, in general, $\widehat{\mathbf{Q}}$ is not a Gaussian vector.

6.2 Projection of experimental dataset $D^{\text{exp}}_{n_r}$

Using the representation of $\mathbf{Q}$ (at convergence) defined by Eq. (7), the experimental dataset $D^{\text{exp}}_{n_r}$ is transformed into the data set $\widehat{D}^{\text{exp}}_{n_r}$ such that

$$\widehat{D}^{\text{exp}}_{n_r} = \{\widehat{\mathbf{q}}^{\text{exp},r}, r = 1, \ldots, n_r\}\,, \tag{11}$$

in which $\widehat{\mathbf{q}}^{\text{exp},r} \in \mathbb{R}^{\nu_q}$ is given by

$$\widehat{\mathbf{q}}^{\text{exp},r} = [\mu_q]^{-1/2}\,[\varphi_q]^T(\mathbf{q}^{\text{exp},r} - \underline{\mathbf{q}}_{\text{ar}}). \tag{12}$$

6.3 PCA of random vector $\mathbf{W}$

Similarly to the PCA of $\mathbf{Q}$, let $\underline{\mathbf{w}}_{\text{ar}} \in \mathbb{R}^{n_w}$ and $[C_{\mathbf{W},\text{ar}}] \in \mathbb{M}^{+0}_{n_w}$ be the empirical estimates of the mean vector and the covariance matrix of $\mathbf{W}$, which are constructed using the additional realizations $\{\mathbf{w}^{\ell}_{\text{ar}}, \ell = 1, \ldots, \nu_{\text{ar}}\}$. The PCA representation, $\mathbf{W}^{(\nu_w)}$, of $\mathbf{W}$ at order $1 \leq \nu_w \leq \nu_{\text{ar}}$ is written as

$$\mathbf{W}^{(\nu_w)} = \underline{\mathbf{w}}_{\text{ar}} + [\varphi_w]\,[\mu_w]^{1/2}\,\widehat{\mathbf{W}}\,, \tag{13}$$

in which $[\varphi_w] \in \mathbb{M}_{n_w,\nu_w}$ is the matrix of the eigenvectors of $[C_{\mathbf{W},\text{ar}}]$ associated with its ν_w largest strictly positive eigenvalues $\mu_{w,1} \geq \mu_{w,2} \geq \ldots \geq \mu_{w,\nu_w} > 0$, represented by the diagonal matrix $[\mu_w] \in \mathbb{M}_{\nu_w}$. The value of ν_w is calculated in order that the L^2-error function $\nu_w \mapsto \text{err}_{\mathbf{W}}(\nu_w)$ defined by

$$\text{err}_{\mathbf{W}}(\nu_w) = \frac{E\{\|\mathbf{W} - \mathbf{W}^{(\nu_w)}\|^2\}}{E\{\|\mathbf{W} - \underline{\mathbf{w}}_{\text{ar}}\|^2\}} = 1 - \frac{\sum_{\alpha=1}^{\nu_w}\mu_{w,\alpha}}{\text{tr}\,[C_{\mathbf{W},\text{ar}}]}\,, \tag{14}$$

be smaller than $\varepsilon_w > 0$. As previously, in Eq. (14), $\mathbf{W}$ stands for $\mathbf{W}^{(n_w)}$. Since $[\varphi_w]^T[\varphi_w] = [I_{\nu_w}]$, the random variable $\widehat{\mathbf{W}}$ with values in $\mathbb{R}^{\nu_w}$ and its ν_{ar} independent realizations are written as

$$\widehat{\mathbf{W}} = [\mu_w]^{-1/2}\,[\varphi_w]^T(\mathbf{W} - \underline{\mathbf{w}}_{\text{ar}})\,, \tag{15}$$

$$\widehat{\mathbf{w}}^{\,\ell} = [\mu_w]^{-1/2}\,[\varphi_w]^T(\mathbf{w}^{\ell}_{\text{ar}} - \underline{\mathbf{w}}_{\text{ar}}) \quad , \quad \ell = 1, \ldots, \nu_{\text{ar}}\,. \tag{16}$$

The empirical estimates $\underline{\widehat{\mathbf{w}}} \in \mathbb{R}^{\nu_w}$ and $[C_{\widehat{\mathbf{W}}}] \in \mathbb{M}^+_{\nu_w}$ of the mean vector and the covariance matrix of $\widehat{\mathbf{W}}$ are such that

$$\underline{\widehat{\mathbf{w}}} = \mathbf{0} \quad , \quad [C_{\widehat{\mathbf{W}}}] = [I_{\nu_w}]\,. \tag{17}$$

As for $\widehat{\mathbf{Q}}$, the components $\widehat{W}_1, \ldots, \widehat{W}_{\nu_w}$ of $\widehat{\mathbf{W}}$ are centered, uncorrelated, and statistically dependent (in the general case).

6.4 Mean-square convergence of the sequence $\{\mathbf{X}^{(\nu_q,\nu_w)}\}_{\nu_q,\nu_w}$

Let $\mathbf{X}^{(\nu_q,\nu_w)} = (\mathbf{Q}^{(\nu_q)}, \mathbf{W}^{(\nu_w)})$ be the random variable with values in $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ and let

$$\text{err}_{\mathbf{X}}(\nu_q, \nu_w) = E\{\|\mathbf{X} - \mathbf{X}^{(\nu_q,\nu_w)}\|^2\}/E\{\|\mathbf{X} - \underline{\mathbf{x}}_{\text{ar}}\|^2\}$$

be the L^2-error function in which $\underline{\mathbf{x}}_{\text{ar}} = (\underline{\mathbf{q}}_{\text{ar}}, \underline{\mathbf{w}}_{\text{ar}}) \in \mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$. In **Supplementary material**, Appendix B, it is proven that if ν_q and ν_w are such that $\text{err}_{\mathbf{Q}}(\nu_q) \leq \varepsilon_q$ and $\text{err}_{\mathbf{W}}(\nu_w) \leq \varepsilon_w$, then

$$\text{err}_{\mathbf{X}}(\nu_q, \nu_w) \leq \varepsilon_q + \varepsilon_w\,.$$

6.5 Learned dataset for the random vector $\widehat{\mathbf{X}} = (\widehat{\mathbf{Q}}, \widehat{\mathbf{W}})$

For a fixed value of $\varepsilon_q + \varepsilon_w$, which defines the level of the mean-square convergence of the PCA of $\mathbf{Q}$ and $\mathbf{W}$, we introduce the learned dataset $\widehat{D}_{\nu_{\text{ar}}}$ constituted of the ν_{ar} independent realizations defined by Eqs. (9) and (16) for the random vector $\widehat{\mathbf{X}} = (\widehat{\mathbf{Q}}, \widehat{\mathbf{W}})$ with values in $\mathbb{R}^\nu$ ($\nu = \nu_q + \nu_w$), such that

$$\widehat{D}_{\nu_{\text{ar}}} = \{\widehat{\mathbf{x}}^{\,\ell} = (\widehat{\mathbf{q}}^{\,\ell}, \widehat{\mathbf{w}}^{\,\ell})\ , \ell = 1, \ldots, \nu_{\text{ar}}\}\,. \tag{18}$$

The methodology consists in constructing a MCMC generator of independent realizations $\{\widehat{\mathbf{w}}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ (for a given ν_{post} as big as we want) of the posterior model $\widehat{\mathbf{W}}^{\text{post}}$ of $\widehat{\mathbf{W}}$, for which the pdf is $p^{\text{post}}_{\widehat{\mathbf{W}}}$, using the learned dataset $\widehat{D}_{\nu_{\text{ar}}}$ defined by Eq. (18) and the experimental dataset $\widehat{D}^{\text{exp}}_{n_r}$ defined by Eq. (11). As soon as these ν_{post} realizations have been generated, the corresponding independent realizations $\{\mathbb{w}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ of $\mathbb{W}^{\text{post}}$, given experimental dataset $\mathbb{D}^{\text{exp}}_{n_r}$ for $\mathbb{Q}$, are calculated using Eq. (13) and (5), by

$$\mathbf{w}^{\text{post},\ell} = \underline{\mathbf{w}}_{\text{ar}} + [\varphi_w]\,[\mu_w]^{1/2}\,\widehat{\mathbf{w}}^{\text{post},\ell}\,, \tag{19}$$

$$\mathbb{w}^{\text{post},\ell} = [\alpha_w]\,\mathbf{w}^{\text{post},\ell} + \boldsymbol{\beta}_w\,. \tag{20}$$

7 Bayesian formulation for the posterior

The classical Bayes formula is used for constructing the pdf $p^{\text{post}}_{\widehat{\mathbf{W}}}$ of the posterior model $\widehat{\mathbf{W}}^{\text{post}}$ of $\widehat{\mathbf{W}}$ with values in $\mathbb{R}^{\nu_w}$ given the datasets $\widehat{D}_{\nu_{\text{ar}}}$ defined by Eq. (18) and $\widehat{D}^{\text{exp}}_{n_r}$ defined by Eq. (11). It is assumed that the mean-square convergence level of $\mathbf{X}^{(\nu_q,\nu_w)}$ is sufficient for substituting $\mathbf{X}^{(\nu_q,\nu_w)}$ by $\mathbf{X}$ or equivalently, substituting $\mathbf{Q}^{(\nu_q)}$ by $\mathbf{Q}$ and $\mathbf{W}^{(\nu_w)}$ by $\mathbf{W}$. The pdf $p_{\widehat{\mathbf{X}}}$ of $\widehat{\mathbf{X}}$ with respect to the Lebesgue measure $d\widehat{\mathbf{x}}$ on $\mathbb{R}^{\nu}$ is replaced by its nonparametric estimate using the learned dataset $\widehat{D}_{\nu_{\text{ar}}}$. The use of Eqs. (7) and (13) allows for deducing the measurable mapping $\widehat{\mathbf{f}}$ from $\mathbb{R}^{\nu_w} \times \mathbb{R}^{n_u}$ into $\mathbb{R}^{\nu_q}$ such that

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

in which $\mathbb{U}$ is the $\mathbb{R}^{n_u}$-valued random variable defined in Section 3, which is statistically independent of $\widehat{\mathbf{W}}$. Let $\mathbb{w} \mapsto \widehat{\mathbf{w}} = \mathbf{h}(\mathbb{w})$ be the continuous mapping from $\mathbb{R}^{n_w}$ into $\mathbb{R}^{\nu_w}$ defined by Eqs. (5) and (15), that is to say, $\mathbf{h}(\mathbb{w}) = [\mu_w]^{-1/2}\,[\varphi_w]^T(\mathbf{w} - \underline{\mathbf{w}}_{\text{ar}})$ with $\mathbf{w} = [\alpha_w]^{-1}(\mathbb{w} - \boldsymbol{\beta}_w)$. Let $\mathcal{C}_{\widehat{\mathbf{w}}} = \mathbf{h}(\mathcal{C}_{\mathbb{w}})$ be the subset of $\mathbb{R}^{\nu_w}$ such that

$$\mathcal{C}_{\widehat{\mathbf{w}}} = \{\widehat{\mathbf{w}} \in \mathbb{R}^{\nu_w} \, ; \, \widehat{\mathbf{w}} = \mathbf{h}(\mathbb{w}) \, , \, \mathbb{w} \in \mathcal{C}_{\mathbb{w}} \subset \mathbb{R}^{n_w}\} \, .$$

Consequently, the support of the prior pdf $\widehat{\mathbf{w}} \mapsto p_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}})$ on $\mathbb{R}^{\nu_w}$ of random variable $\widehat{\mathbf{W}}$ is $\mathcal{C}_{\widehat{\mathbf{w}}} \subset \mathbb{R}^{\nu_w}$. The conditional pdf $\widehat{\mathbf{q}} \mapsto p_{\widehat{\mathbf{Q}}|\widehat{\mathbf{W}}}(\widehat{\mathbf{q}}|\widehat{\mathbf{w}})$ of $\widehat{\mathbf{Q}}$ given $\widehat{\mathbf{W}} = \widehat{\mathbf{w}}$ is defined for $\widehat{\mathbf{w}} \in \mathcal{C}_{\widehat{\mathbf{w}}}$. Taking into account all the hypotheses previously introduced, pdf $p^{\text{post}}_{\widehat{\mathbf{W}}}$ is given by the Bayes formula that is written, for all $\widehat{\mathbf{w}}$ in $\mathcal{C}_{\widehat{\mathbf{w}}}$, as

$$p^{\text{post}}_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}}) = c_0 \, \{\prod_{r=1}^{n_r} p_{\widehat{\mathbf{Q}}|\widehat{\mathbf{W}}}(\widehat{\mathbf{q}}^{\text{exp},r}|\widehat{\mathbf{w}})\} \, p_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}}) \, , \tag{21}$$

in which c_0 is a positive constant of normalization. Let $p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}$ be the joint pdf of $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{W}}$ with respect to the Lebesgue measure $d\widehat{\mathbf{q}}\, d\widehat{\mathbf{w}}$ on $\mathbb{R}^{\nu_q} \times \mathbb{R}^{\nu_w}$. Then, for all $\widehat{\mathbf{w}}$ in $\mathcal{C}_{\widehat{\mathbf{w}}}$, Eq. (21) can be rewritten as

$$p^{\text{post}}_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}}) = c_0 \, \{\prod_{r=1}^{n_r} p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}(\widehat{\mathbf{q}}^{\text{exp},r}, \widehat{\mathbf{w}})\} \, p_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}})^{1-n_r} \, . \tag{22}$$

8 Nonparametric statistical estimation of the posterior

Many works have been published concerning the multidimensional Gaussian kernel-density estimation method [19, 18,21,76]. However, for the high dimensional case, we propose to use a constant covariance matrix that is parameterized by the Silverman bandwidth.

8.1 Formulation proposed and its difficulties

Taking into account Eq. (22), we have to characterize the joint pdf $p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}$ that can be deduced from an estimation of the pdf $p_{\widehat{\mathbf{X}}}$ of $\widehat{\mathbf{X}} = (\widehat{\mathbf{Q}}, \widehat{\mathbf{W}})$. The estimate of $p_{\widehat{\mathbf{X}}}$ is constructed using the Gaussian kernel-density estimation method with the learned dataset $\widehat{D}_{\nu_{\text{ar}}}$ defined by Eq. (18). The construction proposed involves the empirical covariance matrix $[C_{\widehat{\mathbf{X}}}]$ of $\widehat{\mathbf{X}}$ given by

$$[C_{\widehat{\mathbf{X}}}] = \frac{1}{\nu_{\text{ar}} - 1} \sum_{\ell=1}^{\nu_{\text{ar}}} (\widehat{\mathbf{x}}^{\ell} - \underline{\widehat{\mathbf{x}}}) \, (\widehat{\mathbf{x}}^{\ell} - \underline{\widehat{\mathbf{x}}})^T \, , \; \underline{\widehat{\mathbf{x}}} = \frac{1}{\nu_{\text{ar}}} \sum_{\ell=1}^{\nu_{\text{ar}}} \widehat{\mathbf{x}}^{\ell} . \tag{23}$$

Taking into account Eqs. (10) and (17), it can be deduced that $\underline{\widehat{\mathbf{x}}} = (\underline{\widehat{\mathbf{q}}}, \underline{\widehat{\mathbf{w}}}) = \mathbf{0}$. Matrix $[C_{\widehat{\mathbf{X}}}]$ is an element of $\mathbb{M}^{+0}_{\nu}$ or in $\mathbb{M}^{+}_{\nu}$, and can be expressed in block decomposition as,

$$[C_{\widehat{\mathbf{X}}}] = \begin{bmatrix} [\, I_q \,] & [C_{qw}] \\ [C_{qw}]^T & [I_w] \end{bmatrix} \, , \tag{24}$$

in which $[C_{qw}] \in \mathbb{M}_{\nu_q,\nu_w}$ is the covariance matrix of random vectors $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{W}}$. By the Cauchy-Schwarz inequality, we have

$$|\, [C_{qw}]_{jk} \,| \leq 1 \; , \; j \in \{1, \ldots, \nu_q\} \; , \; k \in \{1, \ldots, \nu_w\} \, . \tag{25}$$

Random vectors $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{W}}$ are statistically dependent and are also correlated because we have introduced independent PCA decompositions for $\mathbf{Q}$ and $\mathbf{W}$. The following two comments are appropriate at this point.

(i)- If $[C_{\widehat{\mathbf{X}}}]$ was invertible, the estimate $p^{(\nu_{\text{ar}})}_{\widehat{\mathbf{X}}}$ of $p_{\widehat{\mathbf{X}}}$ would be written, for all $\widehat{\mathbf{x}}$ in $\mathbb{R}^{\nu}$, as [7,56],

$$p^{(\nu_{\text{ar}})}_{\widehat{\mathbf{X}}}(\widehat{\mathbf{x}}) = \frac{c_1}{\nu_{\text{ar}}} \sum_{\ell=1}^{\nu_{\text{ar}}} \exp\{-\frac{1}{2 s^2_{\text{ar}}} < [C_{\widehat{\mathbf{X}}}]^{-1}(\widehat{\mathbf{x}} - \widehat{\mathbf{x}}^{\ell}), (\widehat{\mathbf{x}} - \widehat{\mathbf{x}}^{\ell}) > \}, \tag{26}$$

in which $c_1 = (\, (2\pi)^{\nu/2} s^{\nu}_{\text{ar}} \sqrt{\det[C_{\widehat{\mathbf{X}}}]}\,)^{-1}$ and where s_{ar} is the Silverman bandwidth that is written as

$$s_{\text{ar}} = \left(\frac{4}{\nu_{\text{ar}}(\nu + 2)} \right)^{1/(\nu+4)} . \tag{27}$$

With such a hypothesis, from Eq. (26), it is easy to deduce $p^{(\nu_{\text{ar}})}_{\widehat{\mathbf{Q}}|\widehat{\mathbf{W}}}$ and $p^{(\nu_{\text{ar}})}_{\widehat{\mathbf{W}}}$.

(ii)- Unfortunately, in high dimensions, matrix $[C_{\widehat{\mathbf{X}}}]$ can sometimes be singular. More critically, and also more commonly, $[C_{\widehat{\mathbf{X}}}]$ is invertible in the computational sense but it is slightly ill-conditioned. All the numerical experiments that have been conducted have shown that, if $[C_{\widehat{\mathbf{X}}}]$ is slightly ill-conditioned (for instance, with a condition number of the order 10^3 or 10^4, which is much smaller that the usual tolerance on the condition number for computing the inverse of a matrix),

and if its inverse $[C_{\widehat{\mathbf{X}}}]^{-1}$ is still used, then the estimate of $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}$ defined by Eq. (26) induces some difficulties for the MCMC generator of the posterior pdf defined by Eq. (21). Consequently, we propose to introduce a regularization of $[C_{\widehat{\mathbf{X}}}]$ that should be viewed as an essential part of the construction of the estimation $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}$ of $p_{\widehat{\mathbf{X}}}$.

8.2 Construction of a regularization model of $[C_{\widehat{\mathbf{X}}}]$

Let $[\widehat{C}_\varepsilon]$ be a regularization model in $\mathbb{M}_\nu^+$ of $[C_{\widehat{\mathbf{X}}}]$ such that its condition number is of order 1. Therefore, $[\widehat{C}_\varepsilon]^{-1}$ is in $\mathbb{M}_\nu^+$ and its condition number is also of order 1. This regularization depends on a hyperparameter $\varepsilon \in [\varepsilon_{\min}, 1[$ were $\varepsilon_{\min} > 0$ controls the regularization and whose value will be of close to 0.5. The methodology for choosing the value of ε will be presented in Section 10. The proposed regularization is constructed as follows and additional explanations can be found in **Supplementary material**, Appendix C. Let us consider the following classical spectral representation of matrix $[C_{\widehat{\mathbf{X}}}]$,

$$[C_{\widehat{\mathbf{X}}}] = [\Phi]\,[\lambda]\,[\Phi]^T\,, \tag{28}$$

in which the real eigenvalues are in decreasing order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_\nu \geq 0$ and where $[\Phi]$ is the matrix in $\mathbb{M}_\nu$ of the corresponding eigenvectors. Due to Eqs. (24) and (25), it is proven that these eigenvalues are such that

$$0 \leq \lambda_j \leq 2 \quad , \quad j \in \{1, \ldots, \nu\}\,. \tag{29}$$

If $[C_{qw}]$ was the zero matrix in $\mathbb{M}_{\nu_q,\nu_w}$, then matrix $[C_{\widehat{\mathbf{X}}}]$ would be the identity matrix and therefore, all the eigenvalues would be equal to 1. Since $[C_{qw}]$ is not the zero matrix and taking into account Eq. (29), there exists and we define (by construction of the regularization model) the integer ν_1, such that,

$$\lambda_{\nu_1} \geq 1 \quad , \quad \lambda_{\nu_1+1} < 1 \quad , \quad \nu_1 + 1 \leq \nu\,. \tag{30}$$

The regularization, $[\widehat{C}_\varepsilon]$ of $[C_{\widehat{\mathbf{X}}}]$ is defined by

$$[\widehat{C}_\varepsilon] = [\Phi]\,[\Lambda_\varepsilon]\,[\Phi]^T\,, \tag{31}$$

in which the diagonal matrix $[\Lambda_\varepsilon]$ is such that

$$[\Lambda_\varepsilon]_{jj} = \lambda_j,\ 1 \leq j \leq \nu_1\ ;\ [\Lambda_\varepsilon]_{jj} = \varepsilon^2\,\lambda_{\nu_1},\ \nu_1+1 \leq j \leq \nu, \tag{32}$$

in which $\varepsilon \in [\varepsilon_{\min}, 1[$ were $\varepsilon_{\min} > 0$ is a hyperparameter that controls the regularization and whose value will be of close to 0.5. The methodology for choosing the value of ε will be presented in Section 10. The following properties can then easily be deduced:

$$[\widehat{C}_\varepsilon] \in \mathbb{M}_\nu^+ \quad , \quad [\widehat{C}_\varepsilon]^{-1} = [\Phi]\,[\Lambda_\varepsilon]^{-1}\,[\Phi]^T \in \mathbb{M}_\nu^+\,. \tag{33}$$

The condition numbers of $[\widehat{C}_\varepsilon]$ and $[\widehat{C}_\varepsilon]^{-1}$ are thus equal to $\lambda_1/(\varepsilon^2\lambda_{\nu_1})$, and satisfy the following equation,

$$\text{cond}([\widehat{C}_\varepsilon]) = \text{cond}([\widehat{C}_\varepsilon]^{-1}) \leq \frac{2}{\varepsilon^2}\,.$$

For ε close to 0.5, the condition number is less that 8. We next make four observations relevant to the proposed regularization.

8.3 Construction of the regularized estimate $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}$ of the pdf $p_{\widehat{\mathbf{X}}}$ of $\widehat{\mathbf{X}}$

The regularized estimate of $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}$ defined by Eq. (26) is obtained by using the procedures detailed in Section 8.2. For ε fixed in $[\varepsilon_{\min}, 1[$, let $[G]$ be the $(\nu \times \nu)$ real matrix such that

$$[G] = [\widehat{C}_\varepsilon]^{-1} \in \mathbb{M}_\nu^+ \quad , \quad [G]^{-1} = [\widehat{C}_\varepsilon] \in \mathbb{M}_\nu^+\,, \tag{34}$$

In these conditions, the regularized expression of $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}$ defined by Eq. (26) is written (keeping the same notation) as

$$p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}(\widehat{\mathbf{x}}) = \frac{c_2}{\nu_{\text{ar}}} \sum_{\ell=1}^{\nu_{\text{ar}}} \exp\{-\frac{1}{2s_{\text{ar}}^2} < [G](\widehat{\mathbf{x}} - \widehat{\mathbf{x}}^\ell), (\widehat{\mathbf{x}} - \widehat{\mathbf{x}}^\ell) > \}\,, \tag{35}$$

in which s_{ar} is the Silverman bandwidth defined by Eq. (27) and where

$$c_2 = \frac{\sqrt{\det[G]}}{s_{\text{ar}}^\nu\,(2\pi)^{\nu/2}}\,. \tag{36}$$

In Eqs. (34) and (36), matrix $[G]$ and pdf $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}$ depend on ε, which will be omitted for notational clarity. Let $\widehat{\mathbf{X}}^1, \ldots, \widehat{\mathbf{X}}^{\nu_{\text{ar}}}$ be ν_{ar} independent copies of random variable $\widehat{\mathbf{X}}$ whose pdf is $p_{\widehat{\mathbf{X}}}$. For all $\widehat{\mathbf{x}}$ fixed in $\mathbb{R}^\nu$, let $P_{\nu_{\text{ar}}}(\widehat{\mathbf{x}})$ be the estimator (positive-valued random variable) corresponding to the estimation $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}(\widehat{\mathbf{x}})$ defined by Eq. (35), such that

$$P_{\nu_{\text{ar}}}(\widehat{\mathbf{x}}) = \frac{c_2}{\nu_{\text{ar}}} \sum_{\ell=1}^{\nu_{\text{ar}}} \exp\{-\frac{1}{2s_{\text{ar}}^2} < [G](\widehat{\mathbf{X}}^\ell - \widehat{\mathbf{x}}), (\widehat{\mathbf{X}}^\ell - \widehat{\mathbf{x}}) > \}. \tag{37}$$

It is proven in **Supplementary material**, Appendix E that,

$$E\{(P_{\nu_{\text{ar}}}(\widehat{\mathbf{x}}) - \underline{P}_{\nu_{\text{ar}}}(\widehat{\mathbf{x}}))^2\} \leq \left\{\frac{1}{\nu_{\text{ar}}}\right\}^{4/(\nu+4)} \left\{\frac{\nu+2}{4}\right\}^{\nu/(\nu+4)} \frac{\sqrt{\det[G]}}{(2\pi)^{\nu/2}}\,\underline{P}_{\nu_{\text{ar}}}(\widehat{\mathbf{x}})\,, \tag{38}$$

in which $\underline{P}_{\nu_{\text{ar}}}(\widehat{\mathbf{x}}) = E\{P_{\nu_{\text{ar}}}(\widehat{\mathbf{x}})\}$ is the mean value that tends to $p_{\widehat{\mathbf{X}}}(\widehat{\mathbf{x}})$ when ν_{ar} goes to infinity and consequently, the estimator is asymptotically unbiased and consistent. Due to the mean-square convergence of the sequence of random variables $\{P_{\nu_{\text{ar}}}(\widehat{\mathbf{x}})\}_{\nu_{\text{ar}}}$, as implied by Eq. (38), this sequence of

estimators converges in probability to $p_{\widehat{\mathbf{X}}}(\widehat{\mathbf{x}})$.

Remark. Below, for notational clarity, $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}(\widehat{\mathbf{x}})$ will simply be denoted by $p_{\widehat{\mathbf{X}}}(\widehat{\mathbf{x}})$, which also means that ν_{ar} is chosen sufficiently large for writing that $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})} \simeq p_{\widehat{\mathbf{X}}}$. The ν_{ar}-dependence of $p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}$, $p_{\widehat{\mathbf{Q}}|\widehat{\mathbf{W}}}$, and $p_{\widehat{\mathbf{W}}}$ will also be omitted.

8.4 Deducing the pdf $p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}$ of $(\widehat{\mathbf{Q}},\widehat{\mathbf{W}})$ and the pdf $p_{\widehat{\mathbf{W}}}$ of $\widehat{\mathbf{W}}$

Vector $\widehat{\mathbf{x}}$ and realization $\widehat{\mathbf{x}}^{\ell}$ in $\mathbb{R}^{\nu}$ can be decomposed as $\widehat{\mathbf{x}} = (\widehat{\mathbf{q}},\widehat{\mathbf{w}})$ and $\widehat{\mathbf{x}}^{\ell} = (\widehat{\mathbf{q}}^{\ell},\widehat{\mathbf{w}}^{\ell})$ in which $(\widehat{\mathbf{q}},\widehat{\mathbf{w}})$ and $(\widehat{\mathbf{q}}^{\ell},\widehat{\mathbf{w}}^{\ell})$ belong to $\mathbb{R}^{\nu_q}\times\mathbb{R}^{\nu_w}$ with $\nu=\nu_q+\nu_w$. The $(\nu_q\times\nu_w)$ block notation of matrix $[G]$ is introduced as

$$[G] = \begin{bmatrix} [G_q] & [G_{qw}] \\ [G_{qw}]^T & [G_w] \end{bmatrix} . \tag{39}$$

Since $[G]\in\mathbb{M}_{\nu}^{+}$, we have

$$[G_q]\in\mathbb{M}_{\nu_q}^{+} \quad , \quad [G_w]\in\mathbb{M}_{\nu_w}^{+} . \tag{40}$$

From Eq. (35) and taking into account Eqs. (39)-(40), the joint pdf $p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}$ of $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{W}}$ (with respect to the Lebesgue measure $d\widehat{\mathbf{q}}\,d\widehat{\mathbf{w}}$ on $\mathbb{R}^{\nu_q}\times\mathbb{R}^{\nu_w}$) can be written, for all $\widehat{\mathbf{q}}\in\mathbb{R}^{\nu_q}$ and $\widehat{\mathbf{w}}\in\mathbb{R}^{\nu_w}$, as

$$p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}(\widehat{\mathbf{q}},\widehat{\mathbf{w}}) = \frac{c_2}{\nu_{\text{ar}}}\sum_{\ell=1}^{\nu_{\text{ar}}}\exp\{-\frac{1}{2s_{\text{ar}}^2}\,\psi(\widehat{\mathbf{q}}-\widehat{\mathbf{q}}^{\ell},\widehat{\mathbf{w}}-\widehat{\mathbf{w}}^{\ell})\}, \tag{41}$$

in which the real-valued function $(\widehat{\mathbf{q}},\widehat{\mathbf{w}})\mapsto\psi(\widehat{\mathbf{q}},\widehat{\mathbf{w}})$ defined on $\mathbb{R}^{\nu_q}\times\mathbb{R}^{\nu_w}$ is defined as

$$\psi(\widehat{\mathbf{q}},\widehat{\mathbf{w}}) = <[G_q]\,\widehat{\mathbf{q}}\,,\widehat{\mathbf{q}}> + 2<[G_{qw}]^T\,\widehat{\mathbf{q}}\,,\widehat{\mathbf{w}}> + <[G_w]\,\widehat{\mathbf{w}}\,,\widehat{\mathbf{w}}> . \tag{42}$$

Moreover, the prior pdf $p_{\widehat{\mathbf{W}}}$ of $\widehat{\mathbf{W}}$ (with respect to $d\widehat{\mathbf{w}}$) can be expressed as,

$$p_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}}) = \int_{\mathbb{R}^{\nu_q}} p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}(\widehat{\mathbf{q}},\widehat{\mathbf{w}})\,d\widehat{\mathbf{q}}\,. \tag{43}$$

From Eqs. (41) to (43), since matrix $[G]$ is positive definite, the right-hand side of Eq. (43) can be explicitly calculated [51] and yields,

$$p_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}}) = \frac{c_3}{\nu_{\text{ar}}}\sum_{\ell=1}^{\nu_{\text{ar}}}\exp\{-\frac{1}{2s_{\text{ar}}^2}<[G_0](\widehat{\mathbf{w}}-\widehat{\mathbf{w}}^{\ell}),(\widehat{\mathbf{w}}-\widehat{\mathbf{w}}^{\ell})>\}\,, \tag{44}$$

in which c_3 is the constant of normalization and where $[G_0]$ is a positive-definite matrix that is constructed as the following Schur complement,

$$[G_0] = [G_w] - [G_{qw}]^T\,[G_q]^{-1}\,[G_{qw}]\in\mathbb{M}_{\nu_w}^{+}\,. \tag{45}$$

9 Hamiltonian MCMC generator for the posterior

In Section 9.2, an MCMC generator of the posterior model $\widehat{\mathbf{W}}^{\text{post}}$ of $\widehat{\mathbf{W}}$ is presented, which is based on a nonlinear Itô stochastic differential equation (ISDE) corresponding to a stochastic dissipative Hamiltonian dynamical system for a stochastic process $\{[\mathbf{U}(t)], t\in\mathbb{R}^{+}\}$ with values in $\mathbb{M}_{\nu_w,N_s}$. The number, N_s, of columns of $[\mathbf{U}(t)]$ is chosen sufficiently large (but such that $N_s\le\nu_{\text{ar}}$) in order to increase the exploration of space $\mathbb{R}^{\nu_w}$ by the MCMC algorithm and to facilitate the construction of a reduced-order nonlinear ISDE using the diffusion-maps basis.

The posterior pdf $p_{\widehat{\mathbf{W}}}^{\text{post}}$ defined by Eq. (22) with Eqs. (35) and (41) could require a large number of increments in the MCMC generator if the "distance" of experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$ to initial dataset $\mathbb{D}_{N_d}$ is too large. For decreasing the computational burden, the nonlinear ISDE has to be adapted with respect to the covariance matrix of $\widehat{\mathbf{W}}^{\text{post}}$. Nevertheless, this covariance matrix is unknown and consequently, an appropriate method has to be developed for estimating an approximation of it. Such a relatively classical problem has been addressed for the case of Gaussian likelihoods (see for instance [22]) and more recently, for non-Gaussian likelihoods in [3] within the parametric framework. In the present work devoted to the non-Gaussian likelihood in high dimension and in a nonparametric framework, the proposed approach consists in constructing a nonlinear ISDE adapted to the mean value and to the covariance matrix of $\widehat{\mathbf{W}}^{\text{post}}$, which we will call, *adapted nonlinear ISDE*. The use of an affine transformation, $\widehat{\mathbf{W}}^{\text{post}} = \mathbf{u}_T + [A]^{-T}\,\mathbf{S}^{\text{post}}$ (constructed in Section 9.2), which introduces the matrix-valued stochastic process $\{[\mathbf{S}(t)], t\in\mathbb{R}^{+}\}$ such that $[\mathbf{U}(t)] = [u_T] + [A]^{-T}\,[\mathbf{S}(t)]$, will transform the adapted nonlinear ISDE related to the MCMC generator of $\widehat{\mathbf{W}}^{\text{post}}$ into a nonlinear ISDE for the MCMC generator of $\mathbf{S}^{\text{post}}$ that is a non-Gaussian $\mathbb{R}^{\nu_w}$-valued random variable $\mathbf{S}^{\text{post}}$, "close to" a centered random vector with an identity covariance matrix.

Finally, in order to avoid the data scattering during the generation of independent realizations of $[\mathbf{S}]$, in Section 9.3, the nonlinear ISDE related to stochastic process $\{[\mathbf{S}(t)], t\in\mathbb{R}^{+}\}$ will be projected on a diffusion-maps basis similarly to the methodology of PLoM summarized in **Supplementary material**, Appendix A. The final generation of realizations $\widehat{\mathbf{W}}^{\text{post}}$ is summarized in Section 9.4.

9.1 Criteria for choosing a value of N_s

A natural choice would be $N_s=\nu_{\text{ar}}$. Nevertheless, in general, the number ν_{ar} of additional realizations generated by the PLoM is chosen very large in order to obtain a good convergence of the statistical estimate of the probability distribution of the posterior model. Although such a choice is

always possible, it will always induce a significant increase in computational requirements, often without attaining commensurate gains for the MCMC generator. The choice, $N_s = N_d$, is logical and efficient because the generation of the additional realizations is done with this value by the PLoM. The choice can also be highlighted by the following criterion. The empirical estimate $[C_{\widehat{\mathbf{W}}}]$ of the covariance matrix of $\widehat{\mathbf{W}}$, performed with $\{\widehat{\mathbf{w}}^{\ell}, \ell = 1, \ldots, \nu_{\text{ar}}\}$, is the identity matrix (see Eq. (17)). Let $[C^{N_s}_{\widehat{\mathbf{W}}}]$ be the empirical covariance matrix estimated with $\{\widehat{\mathbf{w}}^{\nu_{\text{ar}}-j+1}, j = 1, \ldots, N_s\}$. Integer N_s can then be chosen such that $\|[C^{N_s}_{\widehat{\mathbf{W}}}]-[I_{\nu_w}]\|_F/\|[I_{\nu_w}]\|_F < \varepsilon_{N_s}$. It can easily be seen that there exists $0 < \varepsilon_{N_s} < 1$ such that $N_s = N_d$ (for instance when $N_d = 200$ and $\nu_{\text{ar}} = 30\,000$, $\varepsilon_{N_s} = 0.05$). Alternatively, a value of N_s can be assessed, using this same criterion, for a predetermined value of ε_{N_s}.

9.2 Adapted nonlinear ISDE as the MCMC generator

The nonlinear ISDE of the MCMC generator of $\widehat{\mathbf{W}}^{\text{post}}$ is constructed as proposed in [59,60], which is based on the works [58] (in which more general stochastic Hamiltonian dynamical systems are analyzed, in particular with a general mass operator that we use hereinafter). The adapted nonlinear ISDE is deduced from it using a similar normalization as the one proposed by Arnst [3]. Nevertheless, in the present non-Gaussian case, the drift vector of the nonlinear ISDE is completely different and the affine transformation for centering and normalizing the posterior model is not the same. We then introduce the matrix $[A]$ that appears in the affine transformation $\widehat{\mathbf{W}}^{\text{post}} = \mathbf{u}_T + [A]^{-T}\,\mathbf{S}$ mentioned above. The method presented in Section 9.3 for constructing $[K]$ (and thus, $[A]$) is also different. Let $[K]$ be a given matrix in $\mathbb{M}^+_{\nu_w}$ and let us consider its Cholesky factorization

$$[K] = [A]\,[A]^T\,. \tag{46}$$

Consequently, the inverse matrices $[K]^{-1}$ and $[A]^{-1}$ exist. As explained above, matrix $[K]$, which is constructed in Section 9.3, will be an approximation of the inverse of the covariance matrix of $\widehat{\mathbf{W}}^{\text{post}}$. We consider, for $t > 0$, the nonlinear stochastic dissipative Hamiltonian dynamical system represented by the following nonlinear ISDE,

$$d[\mathbf{U}(t)] = [K]^{-1}\,[\mathbf{V}(t)]\,dt\,, \tag{47}$$

$$d[\mathbf{V}(t)] = [L([\mathbf{U}(t)])]\,dt - \frac{1}{2}\,f_0^{\text{post}}[\mathbf{V}(t)]\,dt + \sqrt{f_0^{\text{post}}}\,[A]\,d[\mathbf{W}^{\text{wien}}(t)]\,, \tag{48}$$

with the initial condition at $t = 0$,

$$[\mathbf{U}(0)] = [\widehat{w}_0] \quad , \quad [\mathbf{V}(0)] = [\widehat{v}_0] \quad , a.s.\,, \tag{49}$$

in which:

(i) $f_0^{\text{post}} > 0$ is a free parameter allowing the dissipation to be controlled in the stochastic dynamical system. This parameter is chosen such that $f_0^{\text{post}} < 4$. The value, 4, of the upper bound corresponds to the critical damping rate for the linearized ISDE in terms of stochastic process [**S**] (see Section 9.3.3).

(ii) $\{[\mathbf{W}^{\text{wien}}(t)], t \in \mathbb{R}^+\}$ is the stochastic process, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by $\mathbb{R}^+$, with values in $\mathbb{M}_{\nu_w,N_s}$, for which the columns of $[\mathbf{W}^{\text{wien}}(t)]$ are N_s independent copies of the $\mathbb{R}^{\nu_w}$-valued normalized Wiener process $\{\mathbf{W}^{\text{wien}}(t), t \in \mathbb{R}^+\}$ whose matrix-valued autocorrelation function is such that $[R_{\mathbf{W}^{\text{wien}}}(t,t')] = E\{\mathbf{W}^{\text{wien}}(t)\,\mathbf{W}^{\text{wien}}(t')^T\} = \min(t,t')\,[I_{\nu_w}]$.

(iii) $[u] \mapsto [L([u])]$ is a mapping from $\mathbb{M}_{\nu_w,N_s}$ into $\mathbb{M}_{\nu_w,N_s}$, which depends on $p^{\text{post}}_{\widehat{\mathbf{W}}}$ and which is defined as follows. The posterior pdf $p^{\text{post}}_{\widehat{\mathbf{W}}}$ defined by Eq. (22) is written as

$$p^{\text{post}}_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}}) = c_0\,p(\widehat{\mathbf{w}})\,,$$
$$p(\widehat{\mathbf{w}}) = \{\prod_{r=1}^{n_r} p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}(\widehat{\mathbf{q}}^{\text{exp},r}, \widehat{\mathbf{w}})\}\,p_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}})^{1-n_r}\,. \tag{50}$$

Let $\widehat{\mathbf{w}} \mapsto \mathcal{V}(\widehat{\mathbf{w}})$ be the potential function on $\mathbb{R}^{\nu_w}$, which is such that

$$p(\widehat{\mathbf{w}}) = e^{-\mathcal{V}(\widehat{\mathbf{w}})} \quad , \quad \mathcal{V}(\widehat{\mathbf{w}}) = -\log p(\widehat{\mathbf{w}})\,. \tag{51}$$

The matrix $[u]$ is written as $[\mathbf{u}^1 \ldots \mathbf{u}^{N_s}]$ with $\mathbf{u}^j = (u^j_1, \ldots u^j_{\nu_w}) \in \mathbb{R}^{\nu_w}$. Thus, mapping $[L]$ is defined, for all $[u]$ in $\mathbb{M}_{\nu_w,N_s}$, for all $k = 1, \ldots, \nu_w$, and for all $j = 1, \ldots, N_s$, as

$$[L([u])]_{kj} = -\frac{\partial}{\partial u^j_k}\mathcal{V}(\mathbf{u}^j)\,, \tag{52}$$

which can be rewritten as

$$[L([u])]_{kj} = \frac{1}{p(\mathbf{u}^j)}\{\boldsymbol{\nabla}_{\mathbf{u}^j} p(\mathbf{u}^j)\}_k\,. \tag{53}$$

For j fixed in $\{1, \ldots, N_s\}$, the Hamiltonian of the associated conservative homogeneous dynamical system related to stochastic process $\{(\mathbf{U}^j(t), \mathbf{V}^j(t)), t \in \mathbb{R}^+\}$ is thus written as $\mathbb{H}(\mathbf{u}^j, \mathbf{v}^j) = \frac{1}{2} < [K]^{-1}\mathbf{v}^j, \mathbf{v}^j > + \mathcal{V}(\mathbf{u}^j)$.

(iv) $[\widehat{w}_0] \in \mathbb{M}_{\nu_w,N_s}$ is defined by $[\widehat{w}_0] = [\widehat{\mathbf{w}}^{\nu_{\text{ar}}} \ldots \widehat{\mathbf{w}}^{\nu_{\text{ar}}-N_s+1}]$, in which the N_s columns correspond to the N_s last additional realizations $\{\widehat{\mathbf{w}}^{\nu_{\text{ar}}-j+1}, j = 1, \ldots, N_s\}$ generated by the PLoM (See Section 5).

(v) $[\widehat{v}_0] \in \mathbb{M}_{\nu_w,N_s}$ is any realization of a random matrix $[\widehat{\mathbf{V}}_0]$ independent of process $[\mathbf{W}^{\text{wien}}]$, for which the columns $\{\widehat{\mathbf{V}}^j_0, j = 1, \ldots, N_s\}$ are N_s independent Gaussian centered $\mathbb{R}^{\nu_w}$-valued random variables such that the covariance matrix of $\widehat{\mathbf{V}}^j_0$ is $[K]^{-1}$ for all j.

It can be proven (see Theorems 4 to 7 in Pages 211 to 216 and the invariant measure Page 240 of [58]) that the nonlinear ISDE defined by Eqs. (47) to (49) admits the unique invariant measure,

$$\otimes_{j=1}^{N_s} \{p_{\widehat{\mathbf{W}}}^{\text{post}}(\mathbf{u}^j)\, p_{\widehat{\mathbf{V}}}(\mathbf{v}^j)\, d\mathbf{u}^j\, d\mathbf{v}^j\}\,, \tag{54}$$

with $p_{\widehat{\mathbf{V}}}(\mathbf{v}^j) = (2\pi)^{-\nu_w/2} \exp\{-1/2 < [K]^{-1}\mathbf{v}^j, \mathbf{v}^j >\}$. In addition, these Theorems can be used to show that Eqs. (47) to (49) have a unique solution $\{([\mathbf{U}(t)],\, [\mathbf{V}(t)]), t \in \mathbb{R}^+\}$, which is a second-order diffusion stochastic process that is asymptotic for $t \to +\infty$ to the stationary stochastic process $\{([\mathbf{U}_{\text{st}}(t_{\text{st}})], [\mathbf{V}_{\text{st}}(t_{\text{st}})]),\, t_{\text{st}} \in \mathbb{R}^+\}$ for the right-shift semigroup on $\mathbb{R}^+$. For all fixed t_{st}, the joint probability distribution of the random matrices $[\mathbf{U}_{\text{st}}(t_{\text{st}})]$ and $[\mathbf{V}_{\text{st}}(t_{\text{st}})]$ is the invariant measure defined by Eq. (54) and the probability distribution of random matrix $[\mathbf{U}_{\text{st}}(t_{\text{st}})]$ is

$$\otimes_{j=1}^{N_s} p_{\widehat{\mathbf{W}}}^{\text{post}}(\mathbf{u}^j)\, d\mathbf{u}^j\,,$$

that is to say, is the probability distribution of the random matrix $[\widehat{\mathbf{W}}^{\text{post}}]$ with values in $\mathbb{M}_{\nu_w,N_s}$, for which the columns $\widehat{\mathbf{W}}^{\text{post},1}, \ldots, \widehat{\mathbf{W}}^{\text{post},N_s}$ are N_s independent copies of random vector $\widehat{\mathbf{W}}^{\text{post}}$ whose pdf is $p_{\widehat{\mathbf{W}}}^{\text{post}}$ defined by Eq. (50). It can then be deduced that, for any fixed t_{st},

$$[\widehat{\mathbf{W}}^{\text{post}}] = [\mathbf{U}_{\text{st}}(t_{\text{st}})] = \lim_{t\to+\infty} [\mathbf{U}(t)]\,. \tag{55}$$

Equation (55) implies that Eqs. (47) to (49) represent an MCMC generator for $p_{\widehat{\mathbf{W}}}^{\text{post}}$. The free parameter f_0^{post} allows for controlling the transient response generated by the initial condition for quickly reaching the stationary solution (the invariant measure is independent of f_0^{post}). It can also be proven that the asymptotic stationary solution is ergodic.

Expression of the mapping $[L]$ adapted to computation. An explicit algebraic expression is constructed in **Supplementary material**, Appendix F, for the mapping $[u] \mapsto [L([u])]$ defined by Eq. (53). This expression shows the presence of a summation of exponential terms (summation over the number ν_{ar} of realizations $\widehat{\mathbf{q}}^\ell$ and $\widehat{\mathbf{w}}^\ell$ of $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{W}}$). Consequently, an adapted algebraic representation must be developed in order to minimize the numerical cost for each evaluation of $[L([u])]$ and to avoid numerical noise, overflow, and underflow during the computation.

9.3 Transformation of the adapted nonlinear ISDE for the generation of $\widehat{\mathbf{W}}^{\text{post}}$

In this section, we construct the transformation introduced at the beginning of Section 9, we deduce the nonlinear ISDE from the adapted nonlinear ISDE, we verify that the construction proposed satisfies the criteria, and finally, we present the numerical aspects for the computation.

9.3.1 Construction of the transformation

The covariance matrix of $\widehat{\mathbf{W}}^{\text{post}}$ can neither explicitly be calculated using pdf $p_{\widehat{\mathbf{W}}}^{\text{post}}$ nor estimated by computational statistics. Indeed such an estimation would require an integration on $\mathbb{R}^{n_w}$, integration that has to be estimated using the Monte Carlo method [30,61] with respect to a pdf for which a large number of realizations would be drawn (for instance using the ν_{ar} additional realizations of the prior model of $\mathbf{W}$, or using a uniform pdf). The use of $p_{\widehat{\mathbf{W}}}^{\text{post}}$ is not possible since the generator is under construction and as of yet unknown. Even in relatively high dimension, this approach can be prohibitive since the normalization constant c_0 of $p_{\widehat{\mathbf{W}}}^{\text{post}}$ is unknown and has to be numerically estimated. Consequently, an approximation of the covariance matrix of $\widehat{\mathbf{W}}^{\text{post}}$ is performed using a linearization of mapping $[u] \mapsto [L([u])]$ around an approximation, denoted by $\underline{\widehat{\mathbf{w}}}^{\text{exp}}$, of the mean value $E\{\widehat{\mathbf{W}}^{\text{post}}\}$ of $\widehat{\mathbf{W}}^{\text{post}}$ that is also unknown (because only experimental realizations $\{\widehat{\mathbf{q}}^{\text{exp},r}, r = 1, \ldots, n_r\}$ of $\widehat{\mathbf{Q}}$ are assumed to be available. Let us assume that $\underline{\widehat{\mathbf{w}}}^{\text{exp}}$ is a given vector in $\mathbb{R}^{n_w}$, which will be identified in Section 9.3.5. For given vector $\mathbf{u}$ in $\mathbb{R}^{\nu_w}$, let $\mathbf{L}(\mathbf{u}) = (L_1(\mathbf{u}), \ldots, L_{\nu_w}(\mathbf{u}))$ be the vector in $\mathbb{R}^{\nu_w}$ such that, for $k = 1, \ldots, \nu_w$ and for $j = 1, \ldots, N_s$, the component $L_k(\mathbf{u}^j)$ of $\mathbf{L}(\mathbf{u}^j)$ is

$$L_k(\mathbf{u}^j) = [L([u])]_{kj} \quad , \quad [u] = [\mathbf{u}^1 \ldots \mathbf{u}^{N_s}]\,, \tag{56}$$

in which $[L([u])]$ is defined by Eq. (52). Matrix $[K] \in \mathbb{M}_{\nu_w}^+$ introduced in Section 9.2, for which the factorization $[K] = [A]\,[A]^T$ is given by Eq. (46), is then defined for all k and k' in $\{1, \ldots, \nu_w\}$, as

$$[K]_{kk'} = \left\{\frac{\partial^2 \mathcal{V}(\mathbf{u})}{\partial u_k \partial u_{k'}}\right\}_{\mathbf{u}=\underline{\widehat{\mathbf{w}}}^{\text{exp}}} = -\left\{\frac{\partial}{\partial u_{k'}} L_k(\mathbf{u})\right\}_{\mathbf{u}=\underline{\widehat{\mathbf{w}}}^{\text{exp}}}\,. \tag{57}$$

From this definition, matrix $[K]$ is symmetric, but it is not necessarily positive definite for any value of $\underline{\widehat{\mathbf{w}}}^{\text{exp}}$, because function $\mathbf{u} \mapsto \mathcal{V}(\mathbf{u})$ defined by Eq. (51), is not, *a priori*, convex on $\mathbb{R}^{\nu_w}$ for the non-Gaussian pdf $p_{\widehat{\mathbf{W}}}^{\text{post}}$. However, it can be assumed that $\mathbf{u} \mapsto \mathcal{V}(\mathbf{u})$ is locally convex in the neighborhood of $\mathbf{u} = \underline{\widehat{\mathbf{w}}}^{\text{exp}}$ if this vector is correctly estimated (see Section 9.3.5). Therefore, $[K]$ will be in $\mathbb{M}_{\nu_w}^+$ (this property will effectively be checked numerically in the algorithm (see Section 9.3.4)). The first-order Taylor development of $\mathbf{u} \mapsto \mathbf{L}(\mathbf{u})$ around $\mathbf{u} = \underline{\widehat{\mathbf{w}}}^{\text{exp}}$ is written as

$$\mathbf{L}(\mathbf{u}) = \mathbf{L}(\underline{\widehat{\mathbf{w}}}^{\text{exp}}) + [\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{L}(\mathbf{u})]_{\mathbf{u}=\underline{\widehat{\mathbf{w}}}^{\text{exp}}} (\mathbf{u} - \underline{\widehat{\mathbf{w}}}^{\text{exp}}) + o(\|\mathbf{u} - \underline{\widehat{\mathbf{w}}}^{\text{exp}}\|)\,,$$

which yields the following linearized expression,

$$\mathbf{L}^{\text{lin}}(\mathbf{u}) = \mathbf{L}(\underline{\widehat{\mathbf{w}}}^{\text{exp}}) - [K]\,(\mathbf{u} - \underline{\widehat{\mathbf{w}}}^{\text{exp}})\,,$$
$$[K] = -[\boldsymbol{\nabla}_{\mathbf{u}}\mathbf{L}(\mathbf{u})]_{\mathbf{u}=\underline{\widehat{\mathbf{w}}}^{\text{exp}}}\,. \tag{58}$$

Let $\mathbf{u}_T \in \mathbb{R}^{\nu_w}$ be the solution of the equation $\mathbf{L}^{\text{lin}}(\mathbf{u}_T) = \mathbf{0}$, which is such that

$$\mathbf{u}_T = \underline{\widehat{\mathbf{w}}}^{\text{exp}} + [K]^{-1}\,\mathbf{L}(\underline{\widehat{\mathbf{w}}}^{\text{exp}})\,. \tag{59}$$

The transformation of stochastic process $\{[\mathbf{U}(t)], [\mathbf{V}(t)], t \in \mathbb{R}^+\}$, involved in the adapted nonlinear ISDE defined by Eqs. (47) to (49), is written as

$$[\mathbf{U}(t)] = [u_T] + [A]^{-T} [\mathbf{S}(t)] , \tag{60}$$

$$[\mathbf{V}(t)] = [A] [\mathbf{R}(t)] , \tag{61}$$

in which $[u_T] = [\mathbf{u}_T \dots \mathbf{u}_T] \in \mathbb{M}_{\nu_w, N_s}$, where $[A]$ is defined by Eq. (46), and where $\{([\mathbf{S}(t)], [\mathbf{R}(t)]), t \in \mathbb{R}^+\}$ is the new stochastic process with values in $\mathbb{M}_{\nu_w, N_s} \times \mathbb{M}_{\nu_w, N_s}$, such that

$$[\mathbf{S}(t)] = [A]^T ([\mathbf{U}(t)] - [u_T]) , \tag{62}$$

$$[\mathbf{R}(t)] = [A]^{-1} [\mathbf{V}(t)] . \tag{63}$$

9.3.2 Nonlinear ISDE for stochastic process $\{([\boldsymbol{S}(t)], [\boldsymbol{R}(t)]), t \in \mathbb{R}^+\}$

Substituting Eqs. (60) and (61) into Eqs. (47) and (48), using Eqs. (62) and (63) for transforming the initial conditions defined by Eq. (49), simple algebraic manipulations yield, for $t > 0$, the following nonlinear ISDE,

$$d[\mathbf{S}(t)] = [\mathbf{R}(t)] \, dt , \tag{64}$$

$$d[\mathbf{R}(t)] = [\widetilde{L}([\mathbf{S}(t)])] \, dt - \frac{1}{2} f_0^{\text{post}} [\mathbf{R}(t)] \, dt + \sqrt{f_0^{\text{post}}} \, d[\mathbf{W}^{\text{wien}}(t)] , \tag{65}$$

with the almost-sure initial condition at $t = 0$,

$$[\mathbf{S}(0)] = [s_0] \quad , \quad [\mathbf{R}(0)] = [r_0] .$$

The matrices $[s_0]$ and $[r_0]$ in $\mathbb{M}_{\nu_w, N_s}$ are given by

$$[s_0] = [A]^T ([\widehat{w}_0] - [u_T]) , \tag{66}$$

$$[r_0] = [A]^{-1} [\widehat{v}_0] . \tag{67}$$

The mapping $[s] \mapsto [\widetilde{L}([s])]$ from $\mathbb{M}_{\nu_w, N_s}$ into $\mathbb{M}_{\nu_w, N_s}$ is written as

$$[\widetilde{L}([s])] = [A]^{-1} [L([u_T] + [A]^{-T} [s])] . \tag{68}$$

9.3.3 Verifying that the linearized ISDE is well adapted for stochastic process $\{([\boldsymbol{S}(t)], [\boldsymbol{R}(t)]), t \in \mathbb{R}^+\}$

This verification is performed in **Supplementary material**, Appendix G, and shows that the nonlinear ISDE defined by Eqs. (64) to (68) is well adapted to the covariance matrix of the asymptotic stochastic process $\{[\mathbf{S}(t_{\text{st}})], t_{\text{st}} \in \mathbb{R}^+\}$ and therefore, to the covariance of $[\mathbf{W}^{\text{post}}]$ *via* the transformation defined by Eqs. (60) and (61). It can be seen that $f_0^{\text{post}} = 4$ corresponds to the critical damping rate of the linearized dynamical system.

9.3.4 Numerical aspects for computing matrix $[K]$

Assuming that $\underline{\widehat{\mathbf{w}}}^{\text{exp}}$ is given in $\mathbb{R}^{\nu_w}$, we must calculate $[K]$ defined by Eq. (57). Although the algebraic calculation can actually be carried out, the corresponding numerical implementation carries a numerical cost that is greater than the direct numerical calculation of the gradient. This last approach will thus be pursued. Let $\{\Delta t_\alpha, \alpha = 1, 2, \dots\}$ be a decreasing sequence of positive real numbers that goes to zero. Let $[K_\alpha]$ be the sequence of matrices in $\mathbb{M}_{\nu_w}$ such that

$$[K_\alpha]_{kk'} = -\frac{1}{\Delta t_\alpha} (L_k(\underline{\widehat{\mathbf{w}}}^{\text{exp}} + \frac{\Delta t_\alpha}{2} \mathbf{e}^{k'}) - L_k(\underline{\widehat{\mathbf{w}}}^{\text{exp}} - \frac{\Delta t_\alpha}{2} \mathbf{e}^{k'})),$$

in which $\{\mathbf{e}^1, \dots, \mathbf{e}^{\nu_w}\}$ is the canonical basis of $\mathbb{R}^{\nu_w}$ and where L_k is defined by Eq. (56). Matrix $[K]$ is then defined as $[K_{\alpha^{\text{opt}}}]$ in which, for all $\alpha > \alpha^{\text{opt}}$, the symmetrization error is sufficiently small for the Frobenius norm and all the eigenvalues are strictly positive.

9.3.5 Estimating $\underline{\widehat{\boldsymbol{w}}}^{\text{exp}}$

The algorithm proposed for estimating $\underline{\widehat{\mathbf{w}}}^{\text{exp}}$ is based on a predictor-corrector method. The predictor is based on the fact that the size ν_{ar} of the learned dataset, $\mathbb{D}_{\nu_{\text{ar}}}$, constructed in Section 5 using the PLoM, can be chosen as large as required.

(i)- Predictor. The predictor of $\underline{\widehat{\mathbf{w}}}^{\text{exp}}$ is the vector $\underline{\widehat{\mathbf{w}}}^{\text{exp,pred}} \in \mathbb{R}^{\nu_w}$ such that

$$\underline{\widehat{\mathbf{w}}}^{\text{exp,pred}} = E\{\widehat{\mathbf{W}} \,|\, \widehat{\mathbf{Q}} = \underline{\widehat{\mathbf{q}}}^{\text{exp}}\} ,$$

in which $\underline{\widehat{\mathbf{q}}}^{\text{exp}} = (1/n_r) \sum_{r=1}^{n_r} \underline{\widehat{\mathbf{q}}}^{\text{exp},r}$ is the vector in $\mathbb{R}^{\nu_q}$ where $\underline{\widehat{\mathbf{q}}}^{\text{exp},r}$ is defined by Eq. (12). Therefore, we have

$$\underline{\widehat{\mathbf{w}}}^{\text{exp,pred}} = \frac{\int_{\mathbb{R}^{\nu_w}} \widehat{\mathbf{w}} \, p_{\widehat{\mathbf{Q}}, \widehat{\mathbf{W}}}(\underline{\widehat{\mathbf{q}}}^{\text{exp}}, \widehat{\mathbf{w}}) \, d\widehat{\mathbf{w}}}{\int_{\mathbb{R}^{\nu_w}} p_{\widehat{\mathbf{Q}}, \widehat{\mathbf{W}}}(\underline{\widehat{\mathbf{q}}}^{\text{exp}}, \widehat{\mathbf{w}}) \, d\widehat{\mathbf{w}}} , \tag{69}$$

where $p_{\widehat{\mathbf{Q}}, \widehat{\mathbf{W}}}$ is defined by Eqs. (41) and (42). The calculation of the integrals in Eq. (69) can be explicitly evaluated yielding,

$$\underline{\widehat{\mathbf{w}}}^{\text{exp,pred}} = \frac{\sum_{\ell=1}^{\nu_{\text{ar}}} \widetilde{\mathbf{w}}_2^\ell \, \zeta_2^\ell}{\sum_{\ell=1}^{\nu_{\text{ar}}} \zeta_2^\ell} , \tag{70}$$

in which $\widetilde{\mathbf{w}}_2^\ell$ belongs to $\mathbb{R}^{\nu_w}$ and is written as

$$\widetilde{\mathbf{w}}_2^\ell = \widehat{\mathbf{w}}^\ell - [G_w]^{-1} [G_{qw}]^T (\underline{\widehat{\mathbf{q}}}^{\text{exp}} - \widehat{\mathbf{q}}^\ell) ,$$

and where ζ_2^ℓ is positive and such that

$$\zeta_2^\ell = \exp\{-\frac{1}{2 s_{\text{ar}}^2} < [G_1] (\underline{\widehat{\mathbf{q}}}^{\text{exp}} - \widehat{\mathbf{q}}^\ell) , \underline{\widehat{\mathbf{q}}}^{\text{exp}} - \widehat{\mathbf{q}}^\ell > \} .$$

The matrix $[G_1]$ is the Schur complement defined by

$$[G_1] = [G_q] - [G_{qw}] [G_w]^{-1} [G_{qw}]^T \in \mathbb{M}_{\nu_q}^+ .$$

(ii)- Corrector. We introduce the maximum log-likelihood of the posterior model,

$$\underline{\widehat{\mathbf{w}}}^{\text{exp}} = \max_{\widehat{\mathbf{w}}\in\mathbb{R}^{\nu_w}} \log p^{\text{post}}_{\widehat{\mathbf{W}}}(\widehat{\mathbf{w}})\,.$$

Using Eq. (50) for $p^{\text{post}}_{\widehat{\mathbf{W}}}$ with Eqs. (41) and (42) for $p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}$, and Eq. (44) for $p_{\widehat{\mathbf{W}}}$, the non convex optimization problem can be rewritten as

$$\underline{\widehat{\mathbf{w}}}^{\text{exp}} = \max_{\widehat{\mathbf{w}}\in\mathbb{R}^{\nu_w}} J(\widehat{\mathbf{w}})\,, \tag{71}$$

in which $J(\widehat{\mathbf{w}})$ is written as

$$J(\widehat{\mathbf{w}}) = (1-n_r)\,\log\{\sum_{\ell=1}^{\nu_{\text{ar}}} \zeta_0^{\ell}(\widehat{\mathbf{w}})\} + \sum_{r=1}^{n_r} \log\{\sum_{\ell=1}^{\nu_{\text{ar}}} \zeta_1^{r\ell}(\widehat{\mathbf{w}})\}\,,$$

where $\zeta_0^{\ell}(\widehat{\mathbf{w}})$ and $\zeta_1^{r\ell}(\widehat{\mathbf{w}})$ are defined by Eqs. (29) and (30) of Appendix F in **Supplementary material**. The corrector of $\underline{\widehat{\mathbf{w}}}^{\text{exp,pred}}$ is the vector $\underline{\widehat{\mathbf{w}}}^{\text{exp}}$ that is constructed by solving the nonconvex optimization problem defined by Eq. (71) using the interior-point algorithm for which the initial point is chosen as $\widehat{\mathbf{w}}_0 = \underline{\widehat{\mathbf{w}}}^{\text{exp,pred}}$ that is computed using Eq. (70).

9.4 Projection of the nonlinear ISDE for stochastic process $\{([\mathbf{S}(t)],[\mathbf{R}(t)]), t\in\mathbb{R}^+\}$ using a diffusion-maps basis

In order to avoid a possible scattering of the generated realizations constructed by solving the nonlinear ISDE defined by Eqs. (64) to (68) and in order to preserve a possible concentration of the measure $P_{\widehat{\mathbf{W}}^{\text{post}}}(d\widehat{\mathbf{w}}) = p_{\widehat{\mathbf{W}}^{\text{post}}}(\widehat{\mathbf{w}})\,d\widehat{\mathbf{w}}$ on $\mathbb{R}^{\nu_w}$, a projection of the ISDE is carried out using the diffusion-maps basis following the methodology of the PLoM. We then obtain a reduced-order nonlinear ISDE.

9.4.1 Construction of the diffusion-maps basis for the posterior model

The diffusion-maps basis is represented by the matrix

$$[g_s] = [\mathbf{g}_s^1 \ldots \mathbf{g}_s^{m_{\text{post}}}] \in \mathbb{M}_{N_s,m_{\text{post}}}\,,$$

with $1 < m_{\text{post}} \le N_s \le \nu_{\text{ar}}$, which is constructed using the set of independent realizations $\{\mathbf{s}^j, j=1,\ldots,N_s\}$ that result from the transformation defined by Eq. (62) of the set $\{\widehat{\mathbf{w}}^{\nu_{\text{ar}}-j+1}, j=1,\ldots,N_s\}$ extracted form the learned dataset $\widehat{D}_{\nu_{\text{ar}}}$ (see Eq. (18)). We then have,

$$\mathbf{s}^j = [A]^T\,(\widehat{\mathbf{w}}^{\nu_{\text{ar}}-j+1} - \mathbf{u}_T) \in \mathbb{R}^{\nu_w}\,,\ j=1,\ldots,N_s, \tag{72}$$

in which $\mathbf{u}_T \in \mathbb{R}^{\nu_w}$ is defined by Eq. (59). The construction of this diffusion-maps basis is summarized in **Supplementary material**, Appendix H.

9.4.2 Reduced-order nonlinear ISDE

The reduced-order nonlinear ISDE is obtained by projection on diffusion-maps basis $[g_s] \in \mathbb{M}_{N_s,m_{\text{post}}}$ of the nonlinear ISDE relative to the $(\mathbb{M}_{\nu_w,N_s}\times\mathbb{M}_{\nu_w,N_s})$-valued stochastic process $\{([\mathbf{S}(t)],[\mathbf{R}(t)]), t\in\mathbb{R}^+\}$ defined by Eqs. (64) to (68). We then introduced the $(\mathbb{M}_{\nu_w,m_{\text{post}}}\times\mathbb{M}_{\nu_w,m_{\text{post}}})$-valued stochastic process $\{([\boldsymbol{\mathcal{Z}}(t)],[\boldsymbol{\mathcal{Y}}(t)]), t\in\mathbb{R}^+\}$ such that,

$$[\mathbf{S}(t)] = [\boldsymbol{\mathcal{Z}}(t)]\,[g_s]^T\,,\ [\mathbf{R}(t)] = [\boldsymbol{\mathcal{Y}}(t)]\,[g_s]^T,\ t\ge 0. \tag{73}$$

Stochastic process $\{([\boldsymbol{\mathcal{Z}}(t)],[\boldsymbol{\mathcal{Y}}(t)]), t\in\mathbb{R}^+\}$ is then the solution of the reduced-order nonlinear ISDE (obtained by projection) such that, for all $t>0$,

$$d[\boldsymbol{\mathcal{Z}}(t)] = [\boldsymbol{\mathcal{Y}}(t)]\,dt\,, \tag{74}$$

$$d[\boldsymbol{\mathcal{Y}}(t)] = [\widetilde{\mathcal{L}}([\boldsymbol{\mathcal{Z}}(t)])]\,dt - \frac{1}{2}\,f_0^{\text{post}}[\boldsymbol{\mathcal{Y}}(t)]\,dt + \sqrt{f_0^{\text{post}}}\,d[\boldsymbol{\mathcal{W}}^{\text{wien}}(t)]\,, \tag{75}$$

with the almost-sure initial condition at $t=0$,

$$[\boldsymbol{\mathcal{Z}}(0)] = [z_0] \quad,\quad [\boldsymbol{\mathcal{Y}}(0)] = [y_0]\,. \tag{76}$$

The matrices $[z_0]$ and $[y_0]$ in $\mathbb{M}_{\nu_w,m_{\text{post}}}$ are written as

$$[z_0] = [s_0]\,[a_s] \quad,\quad [y_0] = [r_0]\,[a_s]\,,$$

in which matrices $[s_0]$ and $[r_0]$ in $\mathbb{M}_{\nu_w,N_s}$ are defined by Eqs. (66) and (67), and where $[a_s]$ is the matrix such that

$$[a_s] = [g_s]\,([g_s]^T\,[g_s])^{-1} \in \mathbb{M}_{N_s,m_{\text{post}}}\,.$$

In Eq. (75), $[\widetilde{\mathcal{L}}([\boldsymbol{\mathcal{Z}}(t)])]$ is such that

$$[\widetilde{\mathcal{L}}([\boldsymbol{\mathcal{Z}}(t)])] = [\widetilde{L}([\mathbf{Z}(t)]\,[g_s]^T)]\,[a_s]\,,$$

in which $[\widetilde{L}([s])]$ is defined by Eq. (68), and where

$$[\boldsymbol{\mathcal{W}}^{\text{wien}}(t)] = [\mathbf{W}^{\text{wien}}(t)]\,[a_s]\,.$$

9.5 Construction of realizations of $\widehat{\mathbf{W}}^{\text{post}}$

The independent realizations $\{\widehat{\mathbf{w}}^{\text{post},\ell}, \ell=1,\ldots,\nu_{\text{post}}\}$ (used in Eq. (19)) of $\widehat{\mathbf{W}}^{\text{post}}$ whose pdf is $p^{\text{post}}_{\widehat{\mathbf{W}}}$ defined by Eq. (50), are constructed using the discretization of the reduced-order ISDE defined by Eqs. (74) to (76). The number, ν_{post}, of realizations is reparameterized as

$$\nu_{\text{post}} = n_{\text{MC}}^{\text{post}} \times N_s\,,$$

in which $n_{\text{MC}}^{\text{post}}$ is a given integer. Let $[\mathbf{W}^{\text{wien}}(\cdot;\theta)]$ with $\theta\in\Theta$ be a realization of the Wiener stochastic process $[\mathbf{W}^{\text{wien}}]$ defined in Section 9.2-(ii). Let $\{([\boldsymbol{\mathcal{Z}}(t;\theta)],[\boldsymbol{\mathcal{Y}}(t;\theta)]), t\in\mathbb{R}^+\}$ be one realization of the $(\mathbb{M}_{\nu_w,m_{\text{post}}}\times\mathbb{M}_{\nu_w,m_{\text{post}}})$-valued stochastic process $\{([\boldsymbol{\mathcal{Z}}(t)],[\boldsymbol{\mathcal{Y}}(t)]), t\in\mathbb{R}^+\}$, which is computed by solving Eqs. (74) to (76) with the Störmer-Verlet

scheme detailed in **Supplementary material**, Appendix I, for which the sampling step is Δt. Let ℓ_0^{post} be the integer such that, for $t \geq \ell_0^{\text{post}}\,\Delta t$, the solution of Eqs. (74) to (76) is asymptotic to the stationary solution. Therefore, the independent realizations of $\widehat{\mathbf{W}}^{\text{post}}$ can be generated as follows. Let M_0^{post} be a given positive integer. Using Eqs. (60) and (73), for $n = 1, \ldots, n_{\text{MC}}^{\text{post}}$ and for $t_{\ell'} = \ell'\,\Delta t$ with $\ell' = \ell_0^{\text{post}} + n\,M_0^{\text{post}}$, we have, for $j = 1, \ldots, N_s$ and for $k = 1, \ldots, \nu_w$,

$$\widehat{\mathbf{w}}_k^{\text{post},\ell} = [u^{\ell'}]_{kj} \quad , \quad \ell = j + (\ell' - 1)\,N_s\,,$$

$$[u^{\ell'}] = [u_T] + [A]^{-T}\,[s^{\ell'}] \quad , \quad [s^{\ell'}] = [\boldsymbol{\mathcal{Z}}(t_{\ell'};\theta)]\,[g_s]^T\,.$$

In this method of generation, only one realization θ is used and M_0^{post} is chosen sufficiently large in order that $[\boldsymbol{\mathcal{Z}}(t_{\ell'})]$ and $[\boldsymbol{\mathcal{Z}}(t_{\ell'+M_0^{\text{post}}})]$ be two random matrices that are approximatively independent.

10 Choice of the value of the regularization parameter ε

The regularization introduced in Section 8.2 and detailed in **Supplementary material**, Appendix C, was aimed to facilitate the nonparametric statistical estimation of the pdf $p_{\widehat{\mathbf{X}}}$ of random variable $\widehat{\mathbf{X}} = (\widehat{\mathbf{Q}}, \widehat{\mathbf{W}})$ with values in $\mathbb{R}^\nu = \mathbb{R}^{\nu_q} \times \mathbb{R}^{\nu_w}$, using the multidimensional Gaussian kernel-density estimation (see Eq. (35)). As already explained in that section, the proposed regularization depends on the parameter ε and on the criterion for selecting ν_1. Consequently, posterior pdf $p_{\widehat{\mathbf{W}}}^{\text{post}}$ of $\widehat{\mathbf{W}}$, which is directly deduced from $p_{\widehat{\mathbf{X}}}$, depends on ε. There is no *prior* information constraining ε chosen, which is typical when regularization is introduced. Further, a mathematical exploration of Eq. (38), aimed at deducing such constraints, seems intractable.

It may seem possible to compute an optimal value of ε by minimizing the L^1-norm of the difference between the pdf of $\widehat{\mathbf{Q}}^{\text{post}}$ and the pdf of $\widehat{\mathbf{Q}}^{\text{exp}}$. This is not possible if n_r is small, because the quality of the nonparametric estimation of the pdf of $\widehat{\mathbf{Q}}^{\text{exp}}$ would not be sufficiently good. If n_r is sufficiently large for obtaining a good estimation of $\widehat{\mathbf{Q}}^{\text{exp}}$, then an algorithm could proceed as follows. For a given value of ε, the first stage would consist of using the algorithm presented in this paper for estimating the pdf of $\widehat{\mathbf{W}}^{\text{post}}$ and then generating the ν_{post} realizations $\{\widehat{\mathbf{w}}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ of $\widehat{\mathbf{W}}^{\text{post}}$ (which depend on ε). The second stage would consist of estimating the pdf of $\widehat{\mathbf{Q}}^{\text{post}}$ using the conditional pdf of $\widehat{\mathbf{W}}^{\text{post}}$ given $\widehat{\mathbf{Q}} = \widehat{\mathbf{q}}$, which has to be evaluated for the ν_{post} realizations $\{\widehat{\mathbf{w}}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ (and not, using the conditional pdf of $\widehat{\mathbf{W}}$ given $\widehat{\mathbf{Q}} = \widehat{\mathbf{q}}$, which should then be evaluated for the experimental realizations of $\widehat{\mathbf{W}}$, which are not available). Note that the complexity of such an approach would be similar to the one that we have used for estimating the pdf of $\widehat{\mathbf{W}}^{\text{post}}$. The pdf of $\widehat{\mathbf{Q}}^{\text{post}}$ that would be estimated would depend on ε. The third stage would consist of solving an optimization problem with respect to ε for which the objective function would be the L^1-norm of the error. Such a non convex optimization problem would be relatively tricky and numerically expensive.

Consequently, we propose to fix the value of ε to an "average value" that has been estimated by numerical experiments. In order to estimate this "average" value, the following method has been used. Let $p_{\mathbb{W}}^{\text{post}}$ be the posterior pdf of $\mathbb{W}$ that is estimated with the ν_{post} realizations $\{\mathbb{w}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ that are deduced from $\{\widehat{\mathbf{w}}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ computed in Section 9.5, using Eqs. (19) and (20).

The methodology used for validating the range of the values of ε consists in estimating an optimal value of ε, which minimizes a "distance" between the pdf $p_{\mathbb{W}_k}^{\text{post}}$ of component $\mathbb{W}_k^{\text{post}}$, for $k = 1, \ldots, n_w$ (which depends on ε), and an experimental reference, $p_{\mathbb{W}_k}^{\text{exp}}$, that is assumed to be known for the applications used for the validation. Obviously, in the framework of the Bayesian inference, the family of $\{p_{\mathbb{W}_k}^{\text{exp}}, k = 1, \ldots, n_w\}$ are unknown and consequently, cannot be used for estimating ε *a priori*. It is recalled that only n_r experimental realizations $\{\mathbb{q}^{\text{exp},r}, r = 1, \ldots, n_r\}$ of $\mathbb{Q}$ are available and that the corresponding experimental realizations $\{\mathbb{w}^{\text{exp},r}, r = 1, \ldots, n_r\}$ of $\mathbb{W}$ are not available. We thus introduce the error function, $\varepsilon \mapsto \underline{\text{OVL}}(\varepsilon)$, defined by

$$\underline{\text{OVL}}(\varepsilon) = \frac{1}{n_w} \sum_{k=1}^{n_w} \frac{\int_{\mathbb{R}} |p_{\mathbb{W}_k}^{\text{post}}(w) - p_{\mathbb{W}_k}^{\text{exp}}(w)|\,dw}{\int_{\mathbb{R}} p_{\mathbb{W}_k}^{\text{exp}}(w)\,dw}\,. \tag{77}$$

Let $\mathbf{p} = (p_1, \ldots, p_{n_w})$ be a function in the space $L^1(\mathbb{R}, \mathbb{R}^{n_w})$ equipped with the L^1-norm,

$$\|\mathbf{p}\|_{L^1} = \int_{\mathbb{R}} |\mathbf{p}(w)\|_1\,dw = \int_{\mathbb{R}} \sum_{j=1}^{n_w} |p_j(w)|\,dw\,.$$

Introducing the functions $\mathbf{p}^{\text{post}} = (p_{\mathbb{W}_1}^{\text{post}}, \ldots, p_{\mathbb{W}_{n_w}}^{\text{post}})$ and $\mathbf{p}^{\text{exp}} = (p_{\mathbb{W}_1}^{\text{exp}}, \ldots, p_{\mathbb{W}_{n_w}}^{\text{exp}})$ that belong to $L^1(\mathbb{R}, \mathbb{R}^{n_w})$, it can be seen that

$$\frac{\|\mathbf{p}^{\text{post}} - \mathbf{p}^{\text{exp}}\|_{L^1}}{\|\mathbf{p}^{\text{exp}}\|_{L^1}} \leq n_w\,\underline{\text{OVL}}(\varepsilon)\,,$$

because, for $j = 1, \ldots, n_w$, we have $A_j/(a_1 + \ldots + a_{n_w}) \leq A_j/a_j$ for $a_j > 0$ and $A_j > 0$. All the numerical experiments that have been conducted, in particular the applications presented in Section 12, show that the value 0.5 seems an appropriate value for ε. However, this value of 0.5 corresponds to an ad-hoc estimate deduced to a large number of numerical calculations. Further applications and/or new mathematical results are needed to strengthen the arguments concerning this point.

11 Summary of the main steps for implementing the algorithm

In order to help the understanding of the proposed method, we summarize hereinafter the main steps of the algorithm.

Data

1. Scaled initial dataset $D_{N_d} = \{\mathbf{x}_d^j = (\mathbf{q}_d^j, \mathbf{w}_d^j), j = 1, \ldots, N_d\}$ defined in Section 4.
2. Scaled experimental dataset $D_{n_r}^{\exp} = \{\mathbf{q}^{\exp,r}, r = 1, \ldots, n_r\}$ defined in Section 4.

Steps of computation

Step 1. For $\nu_{\text{ar}} \gg N_d$, generating the learned dataset $D_{\nu_{\text{ar}}} = \{(\mathbf{q}_{\text{ar}}^\ell, \mathbf{w}_{\text{ar}}^\ell) = \mathbf{x}_{\text{ar}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$ for the prior probability model using the PLoM with initial dataset D_{N_d} (see Section 5).

Step 2.1. Computing the realizations $\{\widehat{\mathbf{q}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$ of the PCA $\widehat{\mathbf{Q}}$ of $\mathbf{Q}$ using $\{\mathbf{q}_{\text{ar}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$ (see Eqs. (7) and (9)), and controlling the mean-square convergence.

Step 2.2. Computing the realizations $\{\widehat{\mathbf{w}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$ of the PCA $\widehat{\mathbf{W}}$ of $\mathbf{W}$ using $\{\mathbf{w}_{\text{ar}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$ (see Eqs. (13) and (16)), and controlling the mean-square convergence.

Step 2.3. Deducing the learned dataset $\{\widehat{\mathbf{x}}^\ell = (\widehat{\mathbf{q}}^\ell, \widehat{\mathbf{w}}^\ell), \ell = 1, \ldots, \nu_{\text{ar}}\}$ for the random vector $\widehat{\mathbf{X}} = (\widehat{\mathbf{Q}}, \widehat{\mathbf{W}})$ (see Eq. (18)).

Step 2.4. Computing the projection $\{\widehat{\mathbf{q}}^{\exp,r}, r = 1, \ldots, n_r\}$ of the experimental realizations (see Eq. (12)).

Step 3.1. Estimating the covariance matrix $[C_{\widehat{\mathbf{X}}}]$ of $\widehat{\mathbf{X}}$ using the realizations $\{\widehat{\mathbf{x}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$ (see Eq. (23)).

Step 3.2. For $\varepsilon = 0.5$, constructing the regularized matrix $[\widehat{C}_\varepsilon]$ of $[C_{\widehat{\mathbf{X}}}]$ using the method presented in **Supplementary material**, Appendix C.

Step 3.3. Computing matrix $[G] = [\widehat{C}_\varepsilon]^{-1}$ (see Eq. (34)). Deducing matrices $[G_q]$ and $[G_w]$ from matrix $[G]$ (see Eq. (39)). Computing matrix $[G_0]$ as a Schur complement using Eq. (45) and matrix $[G_{0w}] = (1 - n_r)[G_0] + n_r[G_w]$. Computing matrices $[\mathcal{L}_0]$, $[\mathcal{L}_q]$, and $[\mathcal{L}_w]$ by Cholesky factorization of matrices $[G_0]$, $[G_q]$, and $[G_w]$.

Step 4. Computing $\underline{\widehat{\mathbf{w}}}^{\exp}$ using the predictor-corrector method detailed in Section 9.3.5. Computing matrix $[K]$ and vector $\mathbf{u}_T$ for the normalization with respect to the covariance matrix of the posterior $\widehat{\mathbf{W}}^{\text{post}}$: matrix $[K]$ is calculated using the algorithm given in Section 9.3.4 and then, $\mathbf{u}_T$ is calculated with Eq. (59).

Step 5. For a fixed value of N_s such that $N_s \leq \nu_{\text{ar}}$ and chosen as explained in Section 9.1, computing the realizations $\{\mathbf{s}^j, j = 1, \ldots, N_s\}$ using Eq. (72). Then, computing the diffusion-maps basis $[g_s]$ for the posterior model with $1 < m_{\text{post}} \leq N_s \leq \nu_{\text{ar}}$ using the realizations $\{\mathbf{s}^j, j = 1, \ldots, N_s\}$ (see Section 9.4.1).

Step 6. For a fixed value ν_{post} (such that $\nu_{\text{post}} = n_{\text{MC}}^{\text{post}} \times N_s$ in which $n_{\text{MC}}^{\text{post}}$ is given), computation of the realizations $\{\widehat{\mathbf{w}}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ of the posterior model $\widehat{\mathbf{W}}^{\text{post}}$ using the algorithm detailed in Section 9.5, which is based on the solution of the reduced-order nonlinear ISDE defined in Section 9.4.2 for which the Störmer-Verlet scheme is used.

Step 6. From realizations $\{\widehat{\mathbf{w}}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$, computing the realizations $\{\mathbf{w}^{\text{post},\ell}, \ell = 1, \ldots, \nu_{\text{post}}\}$ of the posterior model $\mathbf{W}^{\text{post}}$ using Eq. (19).

12 Applications (AP1) and (AP2)

Two applications, referenced as (AP1) and (AP2), are presented in **Supplementary material**. These applications, which can easily be reproduced, allow for performing the validation of the methodology and algorithms presented.

13 Application (AP3)

In this section, the methodology is applied to the ultrasonic wave propagation in biological tissue for which $\mathbb{W}$ is the vector of the spatial discretization related to the non-Gaussian tensor-valued random elasticity field of a damaged cortical bone due to osteoporosis. This application will be referred to (AP3). All the data concerning this application are described in order that the application can be reproduced.

13.1 Stochastic boundary value problem

This application deals with the numerical simulation of the axial transmission technique that is used in biomechanics for the identification of the cortical bone microstructure. The principle of the axial transmission technique is illustrated in Fig. 1. An impulse is generated by a transmitter placed on the skin of a patient and then, the backscattered pressure field is recorded at distant receivers in the ultrasonic range.

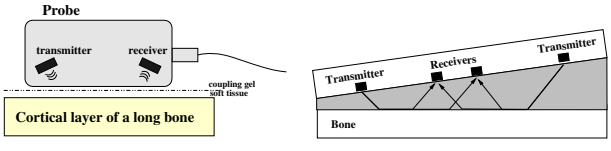

Fig. 1 Application AP3: scheme of the axial transmission technique.

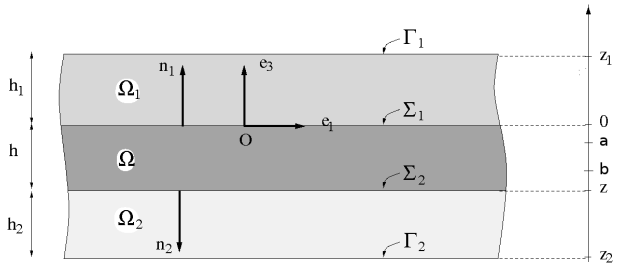

Fig. 2 Application AP3: Geometry of the multilayer system for the boundary value problem.

Boundary value problem The 2D physical space is equipped with a cartesian frame $(O, \mathbf{e}_1, \mathbf{e}_3)$ in which the coordinates of a point are denoted by (x_1, x_3). A boundary value problem has been introduced [15, 16] for modeling the ultrasonic wave propagation in human cortical bone. It consists of a 2D semi-infinite multilayer medium in the $\mathbf{e}_1$ longitudinal direction (see Fig. 2). The model consists of an elastic semi-infinite layer Ω (cortical bone) with thickness h in the $\mathbf{e}_3$ radial direction. This elastic semi-infinite layer is sandwiched between two acoustical fluid layers, Ω_1 (skin and soft tissues) and Ω_2 (bone marrow) with thicknesses h_1 and h_2 in the $\mathbf{e}_3$ radial direction. The media occupying domains Ω_1 and Ω_2 are homogeneous while the cortical bone that occupies domain Ω is heterogeneous in the $\mathbf{e}_3$ direction. The probe (transmitter and receivers) is located in Ω_1.

A mean (nominal) boundary value problem is written in time and space domains considering the three coupled layers: linear acoustic wave equation formulated with the pressure field $P_1(\mathbf{x}, t)$ and $P_2(\mathbf{x}, t)$ in domains Ω_1 and Ω_2, and linear elastodynamics equation formulated with the displacement field $\mathbf{D}(\mathbf{x}, t)$ in domain Ω. Such a formulation requires to define,
(1) for the heterogeneous cortical bone Ω, its mass density $\rho(x_3)$ and its (3×3) matrix-valued effective elasticity field $\{[C(x_3)]\,,\, x_3 \in]-h, 0[\}$;
(2) for the acoustic fluids Ω_1 and Ω_2, their mass densities $\rho_1 = \rho_2 = 1000\,\mathrm{kg.m^{-3}}$, and their sound velocities $c_1 = c_2 = 1500\,\mathrm{m.s^{-1}}$. Note that the acoustic fluid Ω_2 can also be viewed as an elastic solid for which the non zero components of its (3×3) elasticity matrix, denoted as $[C^F]$, are equal to $\rho_2\, c_2^2$.

Introducing a and b such that $-h < b < a < 0$ (see Fig. 2 in which $z = -h$). In case of osteoporosis, there is a gradient of porosity in domain Ω in the $\mathbf{e}_3$ direction:
(1) for $-h < x_3 < b$, the cortical bone is a damaged material mostly made up of an acoustic fluid, which has the same behavior as the acoustic fluid Ω_2.
(2) for $a < x_3 < 0$, the cortical bone is an elastic solid that is modeled by a homogeneous transverse isotropic elastic medium for which its (3×3) elasticity matrix is denoted by $[C^S]$: the transverse Young modulus and the Poisson coefficient are $E_T = 9.8\,\mathrm{GPa}$ and $\nu_T = 0.4$; the longitudinal Young modulus, the Poisson coefficient, and the shear modulus are respectively $E_L = 17.7\,\mathrm{GPa}$, $\nu_L = 0.38$, and $G_L = 4.79\,\mathrm{GPa}$; its mass density is $\rho_S = 1600\,\mathrm{Kg.m^{-3}}$.
(3) for $b \leq x_3 \leq a$, there is a gradient of porosity in the cortical bone.

The model proposed in [16] is used for constructing the mean (nominal) model, based on the hypotheses defined in paragraphs (1) to (3) above, which is written, for all $x_3 \in]-h, 0[$, as

$$\rho(x_3) = (1 - f(x_3))\,\rho_S + f(x_3)\,\rho_2\,,$$

$$[C(x_3)] = \beta_C \left((1 - f(x_3))\,[C^S] + f(x_3)\,[C^F] \right), \qquad (78)$$

in which β_C is a parameter that allows a bias to be introduced in the model, where $f(x_3) = 1$ if $x_3 < b$, $f(x_3) = 0$ if $x_3 > a$, and $f(x_3) = \alpha_0 + \alpha_1\, x_3 + \alpha_2\, x_3^2 + \alpha_3\, x_3^3$ if $b \leq x_3 \leq a$ in which $\alpha_0 = a^2\,(a - 3\,b)/(a - b)^3$, $\alpha_1 = 6\,a\,b/(a - b)^3$, $\alpha_2 = -3(a + b)/(a - b)^3$, and $\alpha_3 = 2/(a - b)^3$.

Prior stochastic model of the matrix-valued effective elasticity field of the cortical bone In practice, the effective elasticity field of the cortical bone, which occupies domain Ω, is a non-Gaussian random field and is modeled by a (3×3) matrix-valued random field $\{[\mathbf{C}(x_3)]\,, \in]-h, 0[\}$ whose mean value is the field $[C(x_3)]$ defined by Eq. (78),

$$E\{[\mathbf{C}(x_3)]\} = [C(x_3)] \quad , \quad \forall\, x_3 \in]-h, 0[\,.$$

The prior probabilistic model of this non-Gaussian random field is taken in the set SFE^+ introduced in [61]. The construction of this set of non-Gaussian matrix-valued random fields is based on the use of the Maximum Entropy principle for constructing a set of positive-definite random matrices. This prior probabilistic model depends only on two hyperparameters, a dispersion coefficient δ_S and a spatial correlation length ℓ_S.

Stochastic model for the acoustical source The transmitter is an acoustic point source located in domain Ω_1, which delivers a random acoustical impulse, and is modeled by a random acoustical source density Q such that

$$\frac{\partial Q}{\partial t}(\mathbf{x}, t) = \rho_1\, F(t) \delta_0(x_1) \delta_0(x_3)\,,$$

in which δ_0 is the Dirac function on the real line at the origin and $F(t) = f_0\, \sin(2\pi F_c\, t)\, e^{-4(t\,F_c - 1)^2}$ in which F_c is the random central frequency whose probability distribution is uniform on $[800, 1200]\,\mathrm{kHz}$ and where $f_0 = 100\,\mathrm{N}$. At time $t = 0$, the system is assumed to be at rest. For each given realization of random field $[\mathbf{C}]$, the corresponding realization of (1) the random displacement field $\mathbf{D}$ and its associated

Von Mises stresses fields $\mathbf{S}^{\mathrm{VM}}$ are computed in Ω, (2) the random pressure fields P_1 and P_2 are computed in Ω_1 and Ω_2. These numerical calculations are performed using the fast and efficient hybrid solver detailed in [15]. It involves a spatial Fourier transform of the random boundary value problems into the longitudinal direction ($\mathbf{e}_1$ direction) and a 1D finite element discretization into the radial direction ($\mathbf{e}_3$ direction).

13.2 Illustration of results obtained with the stochastic boundary value problem

This section deals with an illustration of the ultrasonic wave propagation in the three-layers system using the stochastic boundary value problem defined in Section 13.1, but for which the following particular configuration and parameterization are used. Note that, in presence of a gradient of porosity, the ultrasonic wave propagation is complex and the plot of such waves is difficult to interpret; consequently, for this illustration, it will be assumed that there is a damaged cortical bone without porosity gradient, which means that $-h < b = a < 0$. We consider the case $h_1 = 10^{-2}\,\mathrm{m}$, $h = 8 \times 10^{-3}\,\mathrm{m}$, $h_2 = 10^{-2}\,\mathrm{m}$, $b = a = -h/2$, and $z = -h$ (obviously, for generating the initial dataset used by the probabilistic learning and for generating the experimental dataset required for the Bayesian approach, we will consider a porosity gradient ($-h < b < a < 0$)). The calculation has been performed with $\beta_C = 1$ and $\rho_S = 1722\,\mathrm{Kg.m^{-3}}$. For this illustration, parameters δ_S, ℓ_S, and F_c are considered deterministic (that will not be the case for generating the initial dataset) and are such that $\delta_S = 0.2$, $\ell_S = 3 \times 10^{-3}\,\mathrm{m}$, and $F_c = 1000\,\mathrm{kHz}$. The acoustical point source (transmitter) is located in Ω_1 with coordinates $(x_1 = 0, x_3 = 0.001)\,\mathrm{m}$. In the $\mathbf{e}_3$ radial direction, the number of nodes used in the finite element interpolation of fields P_1, P_2, and $\mathbf{D}$ are 101, 101, and 82, respectively. The Monte Carlo numerical simulation method is used as stochastic solver. The sampling time step is $\Delta t = 2.94 \times 10^{-6}\,\mathrm{s}$ and the number of time sampling points is 330. The sampling spectral step is $\Delta k = 15.70\,\mathrm{rad.m^{-1}}$ and the number of spectral sampling points is 1024. At observation time $t = 9.72 \times 10^{-6}\,\mathrm{s}$, Fig. 3 displays the mean and variance of random fields P_1, $\mathbf{S}^{\mathrm{VM}}$, and P_2. Figure 3 shows the lateral wave (or head wave) propagating from the fluid-solid interface (plane wave front, which links the reflected P-wave front and the interface).

13.3 Generation of the initial dataset

Using the stochastic boundary value problem defined in Section 13.1, the objective is to generate the initial dataset $\mathbb{D}_{N_d} = \{\mathbb{x}_d^j = (\mathbb{q}_d^j, \mathbb{w}_d^j), j = 1, \ldots, N_d\}$ (see Eq. 1) relative to random variable $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$ in which $\mathbb{Q} = (\mathbb{Q}_1, \ldots, \mathbb{Q}_{n_q})$ and

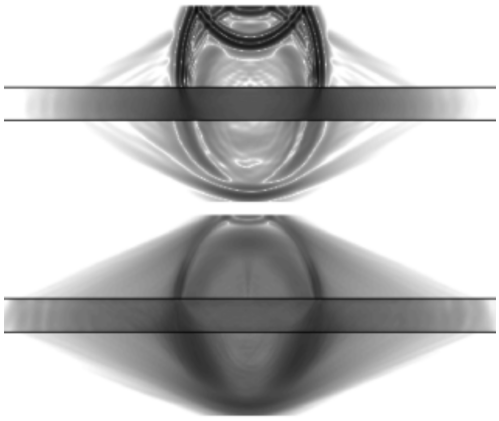

Fig. 3 Application AP3: Propagation of the mean (top) and variance (down) of the random wave in the three layers at $t = 9.72 \times 10^{-6}\mathrm{s}$ with for $h_1 = 10^{-2}\,\mathrm{m}$, $h = 8 \times 10^{-3}\,\mathrm{m}$, $h_2 = 10^{-2}\,\mathrm{m}$. Vertical direction: x_3. In grey color, mean and variance of wave fields $(x_1, x_3) \mapsto P_1(x_1, x_3, t)$ (upper layer), $(x_1, x_3) \mapsto S^{\mathrm{VM}}(x_1, x_3, t)$ (sandwiched layer), $(x_1, x_3) \mapsto P_2(x_1, x_3, t)$ (bottom layer).

$\mathbb{W} = (\mathbb{W}_1, \ldots, \mathbb{W}_{n_w})$. We then have to define the vector-valued random QoI, $\mathbb{Q}$, the vector-valued random system parameter, $\mathbb{W}$, and the $\mathbb{R}^{n_u}$-valued random variable $\mathbb{U}$, which are such that

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

The mapping $\mathbb{f}$ cannot explicitly be described because this mapping is associated with the solution of the boundary value problem. The generation is carried out with $N_d = 200$.
(i) Initial dataset $\mathbb{D}_{N_d}$ is constructed for the case analyzed in [16], that is to say, for a damaged cortical bone with a gradient of porosity such that $h_1 = 2 \times 10^{-3}\,\mathrm{m}$, $h = 8 \times 10^{-3}\,\mathrm{m}$, $h_2 = 10^{-2}\,\mathrm{m}$, $a = -h/2$, $b = -h$, $z = -h$ and $\rho_S = 1600\,\mathrm{Kg.m^{-3}}$. The acoustical point source (transmitter) is located in Ω_1 with coordinates $(x_1 = 0, x_3 = 0.001)\,\mathrm{m}$. The dispersion coefficient δ_S and the spatial correlation length ℓ_S are modeled by random variables with uniform probability distributions on $[0,\ 0.7977]$ and $[1,\ 8] \times 10^{-3}\,\mathrm{m}$ respectively. The central frequency F_c is the uniform random variable defined in Section 13.1. The is no bias introduced in the model and consequently, $\beta_C = 1$ in Eq. (78).
(ii) The number of nodes for the finite element discretization of P_1, P_2, and $\mathbf{D}$ in the $\mathbf{e}_3$ radial direction are 21, 101, and 162, respectively. The sampling time step is $\Delta t = 4.2565 \times 10^{-6}\,\mathrm{s}$ and the sampling spectral step is $\Delta k = 44.88 \mathrm{rad.m^{-1}}$. The number of time sampling points is 300 and the number of spectral sampling points is 2048.
(iii) Let $\mathbb{Q}$ be the random vector of the 300 time sampling points of the random pressure field P_1 at positions $\{(x_1^\ell, x_3 = 10^{-3}\,\mathrm{m}), \ell = 1, \ldots, 14\}$ (14 receivers) in which $x_1^\ell = 13.1 \times 10^{-3} + \ell\,\Delta x_1$ with $\Delta x_1 = 0.8 \times 10^{-3}\,\mathrm{m}$. Thus, $\mathbb{Q}$ is a $\mathbb{R}^{n_q}$-valued random vector with $n_q = 4200$.

(iv) Let $\mathbb{W}$ be the random vector of all the random variables

$$\{[\mathbf{L}(x_3^\ell)]_{ij}\,, 1 \leq i < j \leq 3\} \cup \{\log([\mathbf{L}(x_3^\ell)]_{jj})\,, 1 \leq j \leq 3\}\,,$$

in which $\{[\mathbf{L}(x_3^\ell)]_{ij}\}_{ij}$ are the entries of the random upper triangular matrix $[\mathbf{L}(x_3^\ell)]$ constructed by the following Cholesky factorization, $[\mathbf{C}(x_3^\ell)] = [\mathbf{L}(x_3^\ell)]^T[\mathbf{L}(x_3^\ell)]$. The points $\{x_3^\ell, 1 \leq \ell \leq 120\}$ are the coordinates for the ℓ-th integration points of the finite element mesh of interval $[-h, 0]$ (see Fig. 1). Thus, $\mathbb{W}$ is a $\mathbb{R}^{n_w}$-valued random vector with $n_w = 720$.
(v) The $\mathbb{R}^3$-valued random variable $\mathbb{U}$ is written as $\mathbb{U} = (\delta_s, \ell_s, F_c)$.

13.4 Generation of the experimental dataset

The experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$ is generated with $n_r = 200$ independent experimental realizations $\{\mathbb{q}^{\text{exp},r}, r = 1, \ldots n_r\}$ of $\mathbb{Q}^{\text{exp}} = (\mathbb{Q}_1^{\text{exp}}, \ldots, \mathbb{Q}_{n_q})$, which are such that

$$\mathbb{Q}^{\text{exp}} = \mathbb{f}(\mathbb{U}^{\text{exp}}, \mathbb{W}^{\text{exp}})\,, \tag{79}$$

in which the deterministic mapping $\mathbb{f}$ is the same as the one used in Eq. (78) and corresponds to the use of the boundary value problem defined in Sections 13.1 and 13.3. The random vectors $\mathbb{U}^{\text{exp}}$ and $\mathbb{W}^{\text{exp}}$ are constructed as independent copies of random vectors $\mathbb{U}$ and $\mathbb{W}$ define in Sections13.1 and 13.3, but the bias on the mean model of the random effective elasticity matrix introduced in Eq. (78) is chosen as $\beta_C = 0.9$. It should be noted that the n_r independent realizations $\{\mathbb{w}^{\text{exp},r}, r = 1, \ldots n_r\}$ of random vector $\mathbb{W}^{\text{exp}}$ are generated in order to construct the simulated experiments $\{\mathbb{q}^{\text{exp},r}, r = 1, \ldots n_r\}$ using Eq. (79), but these realizations are not used in the Bayesian approach proposed. Nevertheless, these realizations of $\mathbb{W}_{\text{exp}}$ will be used for estimating the probability density functions $\{w \mapsto p_{\mathbb{W}_k}^{\text{exp}}(w)\}_k$ of the components $\{\mathbb{W}_k^{\text{exp}}\}_k$ of $\mathbb{W}^{\text{exp}}$ in order to validate the methodology proposed (comparing $p_{\mathbb{W}_k}^{\text{exp}}$ to the posterior pdf $p_{\mathbb{W}_k}^{\text{post}}$).

13.5 Values of the numerical parameters, observed quantities for convergence analyses, and validation

The values of the numerical parameters introduced in the algorithm are summarized in Table 1. All the given values of the numerical parameters have been obtained by using the criteria introduced in the theory or have been estimated by performing a local convergence analysis. The quantities used for validating the choice of the value of the regularization parameter ε, for studying the convergence in probability distribution of the probabilistic learning with respect to N_d, and for validating the method proposed, are similar to those introduced in Section 12 for Applications (AP1) and (AP2).

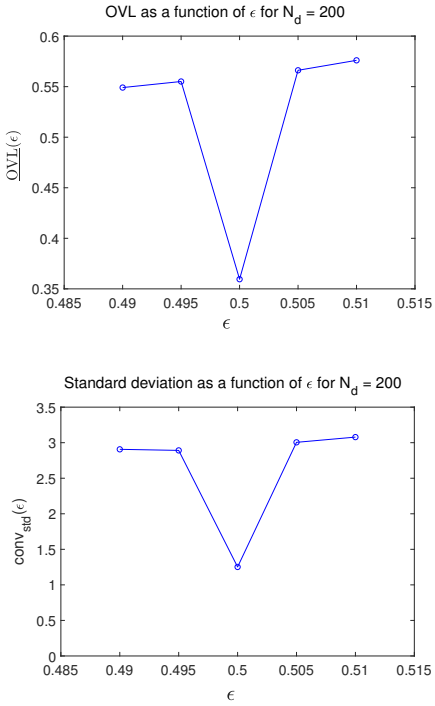

Fig. 4 Application AP3: validation of the choice $\varepsilon = 0.5$. For $N_d = 200$, graph of $\varepsilon \mapsto \underline{\text{OVL}}(\varepsilon)$ (top) and graph of $\varepsilon \mapsto \text{conv}_{\text{std}}(\varepsilon)$ (down).

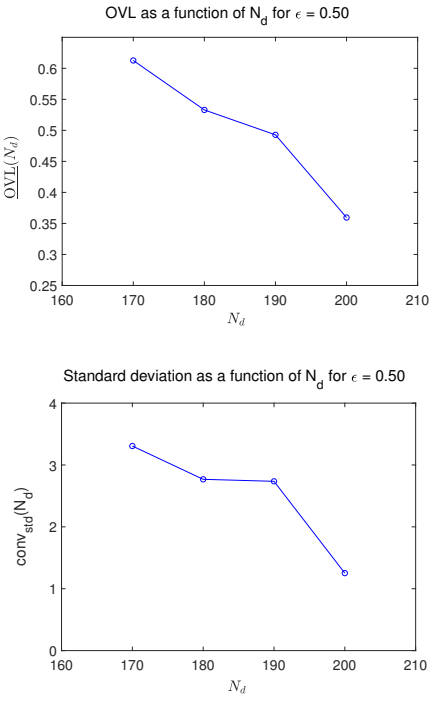

Fig. 5 Application AP3: convergence of the probabilistic learning with respect to N_d. For $\varepsilon = 0.5$, graph of $N_d \mapsto \underline{\text{OVL}}(N_d)$ (top) and graph of $N_d \mapsto \text{conv}_{\text{std}}(N_d)$ (down).

13.6 Results and comments for application (AP3)

The results are presented in Figs. 4 to 6.

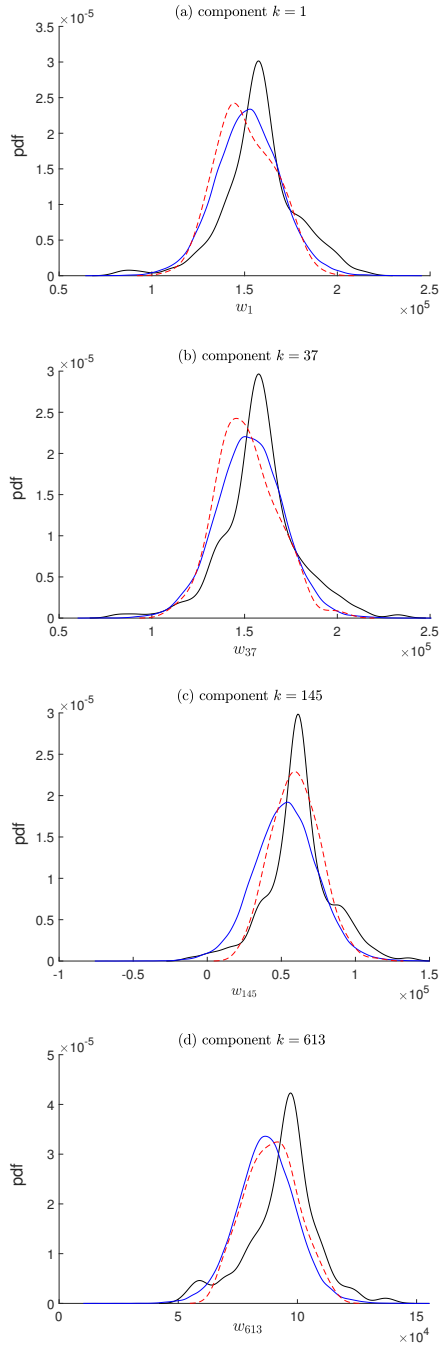

Fig. 6 Application AP3: pdf $w \mapsto p^d_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the initial dataset $\mathbb{D}_{N_d}$ of $N_d = 200$ realizations (thin black line), pdf $w \mapsto p^{\text{exp}}_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the experimental dataset $\mathbb{D}^{\text{exp}}_{n_r}$ of $n_r = 200$ realizations (thick red dashed line), pdf $w \mapsto p^{\text{post}}_{\mathbb{W}_k}(w)$ of $\mathbb{W}^{\text{post}}_k$ estimated with $\varepsilon = 0.5$, $N_d = 200$, and $\nu_{\text{post}} = 40\,000$ realizations (thick blue line), for $k = 5$ (a), $k = 6$ (b), $k = 13$ (c), and $k = 14$ (d).

(i) Concerning the validation of the choice $\varepsilon = 0.5$ of the regularization parameter, Fig. 4 shows that function $\varepsilon \mapsto \underline{\text{OVL}}(\varepsilon)$ has effectively a minimum in the neighborhood of $\varepsilon = 0.5$, as obtained for applications (AP1) and (AP2) (see Section 12).

(ii) Concerning the convergence in probability distribution of the probabilistic learning with respect to size N_d of the initial dataset that is used in all the calculations detailed in Sections 4 to 9, Fig. 5 shows the results obtained for the functions $N_d \mapsto \underline{\text{OVL}}(N_d)$ and $N_d \mapsto \text{conv}_{\text{std}}(N_d)$ with $\varepsilon = 0.5$. The convergence of the learning is slower and a best convergence could certainly be obtained by increasing the maximum value of N_d that should be considered, but as already explained, this slower convergence of the learning with respect to N_d does not interfere with the validation of the proposed methodology (see Section 12).

(iii) Concerning the validation of the method proposed, it can be seen in Fig.4 that, for $N_d = 200$ and $\varepsilon = 0.5$, the norm $\text{conv}_{\text{std}}(\varepsilon)$ of the vector of the standard deviations, normalized by its counterpart for the experiments, is close to 1. Figure 6 shows, for selected components $\mathbb{W}_k$ of random vector $\mathbb{W}$, the comparison of three probability density functions: the pdf $w \mapsto p^d_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the initial dataset $\mathbb{D}_{N_d}$ with $N_d = 200$, the pdf $w \mapsto p^{\text{exp}}_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the experimental dataset $\mathbb{D}^{\text{exp}}_{n_r}$ with $n_r = 200$, and the pdf $w \mapsto p^{\text{post}}_{\mathbb{W}_k}(w)$ of the posterior $\mathbb{W}^{\text{post}}_k$ estimated with $\varepsilon = 0.5$, $N_d = 200$, and $\nu_{\text{post}} = 40\,000$. For each value of k that is considered, the comparison between $p^d_{\mathbb{W}_k}$ and $p^{\text{exp}}_{\mathbb{W}_k}$ shows that there are significant differences (mean value, standard deviation, non-Gaussianity) between these two pdf's, which justifies the use of the Bayesian approach for improving $p^d_{\mathbb{W}_k}$ with $p^{\text{post}}_{\mathbb{W}_k}$. An important element for the validation is the comparison between $p^{\text{post}}_{\mathbb{W}_k}$ and $p^{\text{exp}}_{\mathbb{W}_k}$.

We can see that the results are good considering the intrinsic difficulty of this inverse statistical problem that is in very high dimension. The small differences that can be seen are analyzed from a general point of view in the discussion presented in Section 14. In addition to the general arguments given in this discussion, which explain part of the differences induced by the proposed methodology, there is an additional difficulty intrinsic to the inverse statistical problem for this application (AP3) and which partly affects the quality of the results. For the inverse statistical problem of this physical system, this difficulty is independent of the method used to solve it. This is the sensitivity of the quantities of interest with respect to certain components of the elastic tensor field at certain spatial points of the elastic medium of the dynamical fluid-structure coupled system. If these observed quantities of interest are not very sensitive to the realizations of $\mathbb{W}_k$, which represent random values at a certain spatial point of a certain component of the tensor of elasticity, then the identification of their probability density functions is a little more difficult.

Table 1 For application (AP3), the table defines the values of all the parameters introduced in the algorithm, and which are used in computations.

	Parameters	(AP3)
	n_q	4 200
	n_w	720
	n	4 920
	N_d	≤ 200
	$\max N_d$	200
	n_r	200
Learning (Appendix A in Supplementary material)	ν_x	164
	$\varepsilon_{\text{diff}}$	350
	m	166
	f_0	1.5
	n_{MC}	150
	ν_{ar}	30 000
	Δt	0.2163
	M_0	100
	ℓ_0	100
PCA of **Q** and **W**	ν_q	125
	$\text{err}_{\mathbf{Q}}(\nu_q)$	$3.9e{-4}$
	ν_w	72
	$\text{err}_{\mathbf{W}}(\nu_w)$	$2.7e{-4}$
Posterior	ν	197
	ν_1	125
	ε	0.5
	N_s	200
	ε_{N_s}	0.0433
	f_0^{post}	10^{-3}
	$\varepsilon_{\text{diff}}^{\text{post}}$	300 000
	m_{post}	69
	$n_{\text{MC}}^{\text{post}}$	200
	ν_{post}	40 000
	Δt	0.05854
	M_0^{post}	100
	ℓ_0^{post}	120

14 Discussion and conclusion

In this paper, we have presented a methodology for implementing the Bayesian inference in the framework of the small-data challenge using the probabilistic learning on manifolds under the following hypotheses: the likelihood probability distribution is not Gaussian and cannot be approximated by a Gaussian measure, the problem can be in high dimension, the number of given realizations in the initial dataset of the prior model is assumed to be small, which corresponds, for instance, to the use of an expensive computer code for generating the initial data set (training), the number of experimental realizations is also small, and the number of posterior realizations can be arbitrarily large. For solving these difficult problem, a novel methodology has been developed. The method and the associated algorithms have been adapted to take into account all the constraints induced by the given framework. Three applications have been presented for validating the approach proposed: two are relatively simple and can easily be reproduced, and the third one corresponds to a difficult statistical inverse problem. The method proposed will have to be tested for many other applications for confirming its robustness and its capability to treat problems in high dimension and in the framework of the small-data challenge.

The results obtained for the three applications presented are good. Nevertheless, the small differences obtained in the third application require further discussion. Below, we discuss from a general point of view the main steps of the proposed Bayesian identification methodology, highlighting their possible impact on the quality of the results.

The first step of the methodology is the use of the PLoM method, increasing the number of realizations of the initial dataset, which is not enough to obtain the convergence of the statistical estimators. This step is not an approximation but rather a modeling choice, which makes it possible to tackle a key obstacle in the context of small-data. Obviously, this enrichment must also be correct. This property has been shown for many published applications and recently, its properties have been proven mathematically.

The second step is the use of a PCA representation, on the one hand for the quantities of interest (the outputs) and on the other hand for the control parameters (the inputs). The approximation errors thus introduced can be, and are, reduced to a level as small as desired. There are therefore no errors introduced by this step.

The third step is that of the regularization of the covariance matrix of the vector of the concatenated coordinates of the two PCAs. This step is clearly an approximation, which can introduce an irreducible error and therefore, affects the quality of the results. Indeed although the error induced by the introduced regularization is minimized in the construction of the proposed method, this error cannot be estimated and controlled. This regularization is the price to pay to address the second obstacle induced by the high dimension in the Bayesian framework. We cannot introduce a single PCA, which would solve the difficulty of high dimension without having to introduce a regularization, but which would then not allow the calculation of the likelihood function. The choice of the introduction of two PCAs solves the problem of the high dimension and allows the calculation of the likelihood function. On the other hand, this method requires the introduction of a regularization whose model is arbitrary although the one proposed herein results from an in-depth analysis.

The fourth step is again the use of the PLoM method to generate the realizations of the posterior probability measure. An affine transformation of the posterior probability

measure is constructed to improve the exploration of the support of this measure, which can still live in a high-dimensional set (despite the PCA reductions introduced). This transformation does not introduce errors. It simply improves the exploration by the MCMC generator, which is based on the reduced-order Itô stochastic differential equation, obtained by projection on the diffusion-maps basis, in order to generate the realizations of the posterior model.

Finally, following this discussion, it appears that the only point that can generate an irreducible error is the regularization. This question on the choice of regularization remains an open problem.

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

References

1. Afshari, H.H., Gadsden, S.A., Habibi, S.: Gaussian filters for parameter and state estimation: A general review of theory and recent trends. Signal Processing **135**, 218–238 (2017)
2. Andrieu, C., Thoms, J.: A tutorial on adaptive mcmc. Statistics and computing **18**(4), 343–373 (2008)
3. Arnst, M., Abello Álvarez, B., Ponthot, J.P., Boman, R.: Itô-SDE MCMC method for Bayesian characterization of errors associated with data limitations in stochastic expansion methods for uncertainty quantification. Journal of Computational Physics **349**, 59–79 (2017). DOI 10.1016/j.jcp.2017.08.005
4. Arnst, M., Ghanem, R., Soize, C.: Identification of Bayesian posteriors for coefficients of chaos expansions. Journal of Computational Physics **229**(9), 3134–3154 (2010). DOI 10.1016/j.jcp.2009.12.033
5. Beskos, A., Girolami, M., Lan, S., Farrell, P.E., Stuart, A.M.: Geometric mcmc for infinite-dimensional inverse problems. Journal of Computational Physics **335**, 327–351 (2017). DOI 10.1016/j.jcp.2016.12.041
6. Bilionis, I., Zabaras, N.: Bayesian uncertainty propagation using Gaussian processes. In: R. Ghanem, D. Higdon, O. H. (eds.) Handbook of Uncertainty Quantification, chap. 15, pp. 555–600. Springer, Cham, Switzerland (2017)
7. Bowman, A., Azzalini, A.: Applied Smoothing Techniques for Data Analysis. Oxford University Press, Oxford, UK (1997)
8. Capiez-Lernout, E., Soize, C., Mignolet, M.P.: Post-buckling nonlinear static and dynamical analyses of uncertain cylindrical shells and experimental validation. Computer Methods in Applied Mechanics and Engineering **271**, 210–230 (2014)
9. Carlberg, K., Bou-Mosleh, C., Farhat, C.: Efficient non-linear model reduction via a least-squares petrov–galerkin projection and compressive tensor approximations. International Journal for Numerical Methods in Engineering **86**(2), 155–181 (2011)
10. Carlin, B.P., Louis, T.A.: Bayesian Methods for Data Analysis. Third Edition, Chapman & Hall / CRC Press, Boca Raton (2009)
11. Chaturantabut, S., Sorensen, D.C.: Nonlinear model reduction via discrete empirical interpolation. SIAM Journal on Scientific Computing **32**(5), 2737–2764 (2010)
12. Congdon, P.: Bayesian Statistical Modelling. Second Edition, John Wiley & Sons, Chichester (2007)
13. Cotter, S.L., Roberts, G.O., Stuart, A.M., White, D.: Mcmc methods for functions: modifying old algorithms to make them faster. Statistical Science pp. 424–446 (2013). DOI 10.1214/13-STS421
14. Dashti, M., Stuart, A.M.: The Bayesian approach to inverse problems. In: R. Ghanem, D. Higdon, O. H. (eds.) Handbook of Uncertainty Quantification, chap. 10, pp. 311–428. Springer, Cham, Switzerland (2017)
15. Desceliers, C., Soize, C., Grimal, Q., Haiat, G., Naili, S.: A time domain method to solve transient elastic wave propagation in a multilayer medium with a hybrid spectral-finite element space approximation. Journal of Wave Motion **45**(4), 383–399 (2008)
16. Desceliers, C., Soize, C., Naili, S., Haïat, G.: Probabilistic model of the human cortical bone with mechanical alterations in ultrasonic range. Mechanical Systems and Signal Processing **32**(-), 170–177 (2012). DOI 10.1016/j.ymssp.2012.03.008
17. Dolgov, S., Khoromskij, B.N., Litvinenko, A., Matthies, H.G.: Polynomial chaos expansion of random coefficients and the solution of stochastic partial differential equations in the tensor train format. SIAM/ASA Journal on Uncertainty Quantification **3**(1), 1109–1135 (2015)
18. Duong, T., Cowling, A., Koch, I., Wand, M.: Feature significance for multivariate kernel density estimation. Computational Statistics & Data Analysis **52**(9), 4225–4242 (2008). DOI 10.1016/j.csda.2008.02.035
19. Duong, T., Hazelton, M.L.: Cross-validation bandwidth matrices for multivariate kernel density estimation. Scandinavian Journal of Statistics **32**(3), 485–506 (2005)
20. Farhat, C., Chapman, T., Avery, P.: Structure-preserving, stability, and accuracy properties of the energy-conserving sampling and weighting method for the hyper reduction of nonlinear finite element dynamic models. International Journal for Numerical Methods in Engineering **102**(5), 1077–1110 (2015). DOI 10.1002/nme.4820
21. Filippone, M., Sanguinetti, G.: Approximate inference of the bandwidth in multivariate kernel density estimation. Computational Statistics & Data Analysis **55**(12), 3104–3122 (2011). DOI 10.1016/j.csda.2011.05.023
22. Flath, H.P., Wilcox, L.C., Akçelik, V., Hill, J., van Bloemen Waanders, B., Ghattas, O.: Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial hessian approximations. SIAM Journal on Scientific Computing **33**(1), 407–432 (2011). DOI 10.1137/090780717
23. Ghahramani, Z.: Probabilistic machine learning and artificial intelligence. Nature **521**(7553), 452 (2015)
24. Ghanem, R., Doostan, R.: Characterization of stochastic system parameters from experimental data: A bayesian inference approach. Journal of Computational Physics **217**(1), 63–81 (2006)
25. Ghanem, R., Higdon, D., Owhadi, H.: Handbook of Uncertainty Quantification, vol. 1 to 3. Springer, Cham, Switzerland (2017). DOI 10.1007/978-3-319-12385-1
26. Ghanem, R., Soize, C.: Probabilistic nonconvex constrained optimization with fixed number of function evaluations. International Journal for Numerical Methods in Engineering **113**(4), 719–741 (2018). DOI 10.1002/nme.5632
27. Ghanem, R., Soize, C., Thimmisetty, C.: Optimal well-placement using probabilistic learning. Data-Enabled Discovery and Applications **2**(1), 4,1–16 (2018). DOI 10.1007/s41688-017-0014-x
28. Ghanem, R.G., Soize, C., Safta, C., Huan, X., Lacaze, G., Oefelein, J.C., Najm, H.N.: Design optimization of a scramjet under uncertainty using probabilistic learning on manifolds. Journal of Computational Physics **399**, 108930, 1–14 (2019). DOI 10.1016/j.jcp.2019.108930
29. Giraldi, L., Le Maître, O.P., Mandli, K.T., Dawson, C.N., Hoteit, I., Knio, O.M.: Bayesian inference of earthquake parameters from buoy data using a polynomial chaos-based surrogate. Computational Geosciences **21**(4), 683–699 (2017)
30. Givens, G.H., Hoeting, J.A.: Computational statistics, 2nd edition. John Wiley & Sons (2013)
31. Gordon, N.J., Salmond, D.J., Smith, A.F.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings for radar and signal processing **140**(2), 107–113 (1993)

32. Grepl, M.A., Maday, Y., Nguyen, N.C., Patera, A.T.: Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. ESAIM: Mathematical Modelling and Numerical Analysis **41**(3), 575–605 (2007)
33. Isaac, T., Petra, N., Stadler, G., Ghattas, O.: Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the antarctic ice sheet. Journal of Computational Physics **296**, 348–368 (2015). DOI 10.1016/j.jcp.2015.04.047
34. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Springer-Verlag, New York (2005)
35. Kennedy, M.C., O'Hagan, A.: Predicting the output from a complex computer code when fast approximations are available. Biometrika **87**(1), 1–13 (2000)
36. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63**(3), 425–464 (2001). DOI 10.1111/1467-9868.00294
37. Korb, K.B., Nicholson, A.E.: Bayesian artificial intelligence. CRC press, Boca Raton (2010)
38. Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D.: Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. Statistics and computing **10**(4), 325–337 (2000)
39. Marzouk, Y., Najm, H.: Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. Journal of Computational Physics **228**(6), 1862–1902 (2009)
40. Marzouk, Y., Najm, H., Rahn, L.: Stochastic spectral methods for efficient Bayesian solution of inverse problems. Journal of Computational Physics **224**(2), 560–586 (2007)
41. Matthies, H.G., Zander, E., Rosić, B.V., Litvinenko, A., Pajonk, O.: Inverse problems in a Bayesian setting. In: Computational Methods for Solids and Fluids, pp. 245–286. Springer (2016)
42. Meyer, M., Matthies, H.G.: Efficient model reduction in non-linear dynamics using the karhunen-loeve expansion and dual-weighted-residual methods. Computational Mechanics **31**(1-2), 179–191 (2003)
43. Murphy, K.P.: Machine Learning: A Probabilistic Perspective. The MIT press (2012)
44. Nagel, J.B., Sudret, B.: Spectral likelihood expansions for Bayesian inference. Journal of Computational Physics **309**, 267–294 (2016)
45. Nouy, A.: Low-rank tensor methods for model order reduction. In: R. Ghanem, D. Higdon, O. H. (eds.) Handbook of Uncertainty Quantification, chap. 25, pp. 857–882. Springer, Cham, Switzerland (2017)
46. Pajonk, O., Rosić, B.V., Litvinenko, A., Matthies, H.G.: A deterministic filter for non-Gaussian Bayesian estimation - applications to dynamical system estimation with noisy measurements. Physica D: Nonlinear Phenomena **241**(7), 775–788 (2012)
47. Palacios, M.B., Steel, M.F.J.: Non-gaussian bayesian geostatistical modeling. Journal of the American Statistical Association **101**(474), 604–618 (2006)
48. Parussini, L., Venturi, D., Perdikaris, P., Karniadakis, G.E.: Multi-fidelity Gaussian process regression for prediction of random fields. Journal of Computational Physics **336**, 36–50 (2017)
49. Perrin, G., Soize, C., Ouhbi, N.: Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints. Computational Statistics & Data Analysis **119**, 139–154 (2018)
50. Pratola, M.T., Sain, S.R., Bingham, D., Wiltberger, M., Rigler, E.J.: Fast sequential computer model calibration of large nonstationary spatial-temporal processes. Technometrics **55**(2), 232–242 (2013)
51. Puntanen, S., Styan, G.: The schur complements in statistics and probability. In: F. Zhang (ed.) The Schur Complement and its Applications, chap. 6, pp. 163–226. Springer, Boston, MA (2005)
52. Robert, C., Casella, G.: Monte Carlo statistical methods. Springer Science & Business Media (2013)
53. Rosic, B.V., Litvinenko, A., Pajonk, O., Matthies, H.G.: Sampling-free linear Bayesian update of polynomial chaos representations. Journal of Computational Physics **231**(17), 5761–5787 (2012)
54. Ryckelynck, D.: A priori hyperreduction method: an adaptive approach. Journal of computational physics **202**(1), 346–366 (2005)
55. Santner, T., Williams, B., Notz, W.: The Design and Analysis of Computer Experiments. Springer-Verlag, Berlin, New York (2003)
56. Scott, D.: Multivariate Density Estimation: Theory, Practice, and Visualization, 2nd edn. John Wiley and Sons, New York (2015)
57. Scott, S.L., Blocker, A.W., Bonassi, F.V., Chipman, H.A., George, E.I., McCulloch, R.E.: Bayes and big data: The consensus Monte Carlo algorithm. International Journal of Management Science and Engineering Management **11**(2), 78–88 (2016)
58. Soize, C.: The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions. World Scientific Publishing Co Pte Ltd, Singapore (1994)
59. Soize, C.: Construction of probability distributions in high dimension using the maximum entropy principle. applications to stochastic processes, random fields and random matrices. International Journal for Numerical Methods in Engineering **76**(10), 1583–1611 (2008). DOI 10.1002/nme.2385
60. Soize, C.: Polynomial chaos expansion of a multimodal random vector. SIAM/ASA Journal on Uncertainty Quantification **3**(1), 34–60 (2015). DOI 10.1137/140968495
61. Soize, C.: Uncertainty Quantification. An Accelerated Course with Advanced Applications in Computational Engineering. Springer, New York (2017). DOI 10.1007/978-3-319-54339-0
62. Soize, C., Farhat, C.: Probabilistic learning for modeling and quantifying model-form uncertainties in nonlinear computational mechanics. International Journal for Numerical Methods in Engineering **117**, 819–843 (2019). DOI 10.1002/nme.5980
63. Soize, C., Ghanem, R.: Data-driven probability concentration and sampling on manifold. Journal of Computational Physics **321**, 242–258 (2016). DOI 10.1016/j.jcp.2016.05.044
64. Soize, C., Ghanem, R., Safta, C., Huan, X., Vane, Z., Oefelein, J., Lacaze, G., Najm, H., Tang, Q., Chen, X.: Entropy-based closure for probabilistic learning on manifolds. Journal of Computational Physics (2019). DOI 10.1016/j.jcp.2018.12.029
65. Soize, C., Ghanem, R.G.: Physics-constrained non-gaussian probabilistic learning on manifolds. International Journal for Numerical Methods in Engineering **121**(1), 110–145 (2020). DOI 10.1002/nme.6202
66. Soize, C., Ghanem, R.G.: Probabilistic learning on manifolds. arXiv preprint:2002.12653, 2020, math.ST, 28 February 2020 (2020)
67. Soize, C., Ghanem, R.G., Safta, C., Huan, X., Vane, Z.P., Oefelein, J.C., Lacaze, G., Najm, H.N.: Enhancing model predictability for a scramjet using probabilistic learning on manifolds. AIAA Journal **57**(1), 365–378 (2019). DOI 10.2514/1.J057069
68. Spall, J.C.: Introduction to stochastic search and optimization: estimation, simulation, and control, vol. 65. John Wiley & Sons (2005)
69. Spantini, A., Cui, T., Willcox, K., Tenorio, L., Marzouk, Y.: Goal-oriented optimal approximations of Bayesian linear inverse problems. SIAM Journal on Scientific Computing **39**(5), S167–S196 (2017)
70. Stuart, A.M.: Inverse problems: a Bayesian perspective. Acta Numerica **19**, 451–559 (2010). DOI 10.1017/S0962492910000061
71. Tarantola, A.: Inverse problem theory and methods for model parameter estimation, vol. 89. siam, Philadelphia (2005)
72. Tipireddy, R., Ghanem, R.: Basis adaptation in homogeneous chaos spaces. Journal of Computational Physics **259**, 304–317 (2014)

73. Tsilifis, P., Ghanem, R.: Bayesian adaptation of chaos representations using variational inference and sampling on geodesics. Proc. R. Soc. A **474**(2217), 20180285 (2018)
74. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: Data Mining: Practical machine learning tools and techniques, fourth edn. Morgan Kaufmann, imprint of Elsevier, Cambridge, MA, United States (2017)
75. Zhou, Q., Liu, W., Li, J., Marzouk, Y.: An approximate empirical Bayesian method for large-scale linear-Gaussian inverse problems. Inverse Problems (2018)
76. Zougab, N., Adjabi, S., Kokonendji, C.C.: Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. Computational Statistics & Data Analysis **75**, 28–38 (2014). DOI 10.1016/j.csda.2014.02.002

Noname manuscript No.
(will be inserted by the editor)

Supplementary material of the paper: Sampling of Bayesian posteriors with a non-Gaussian probabilistic learning on manifolds from a small dataset

Journal: Statistics and Computing

Christian Soize · Roger G. Ghanem · Christophe Desceliers

the date of receipt and acceptance should be inserted later

This supplementary material is made up

- Section 12 devoted to the presentation of Applications (AP1) and (AP2).
- Appendices A to I. Appendix A is a summary of the algorithm of the PLoM. In Appendix B, we give the proof of the mean-square convergence of the random sequence $\mathbf{X}^{(\nu_q,\nu_w)}$ related to the introduction of the reduced-order representations of $\mathbf{Q}$ and $\mathbf{W}$. Appendix C is devoted to the construction of a regularization model of the covariance matrix $[C_{\widehat{\mathbf{X}}}]$. The proof of the range of the values of the covariance matrix of $\widehat{\mathbf{X}} = (\widehat{\mathbf{Q}}, \widehat{\mathbf{W}})$ is detailed in Appendix D. In Appendix E, we give the proof of the consistency of the estimator of the regularized pdf of $\widehat{\mathbf{X}}$, for which an upper bound is constructed as a function of ε, ν_{ar}, and ν. Appendix F deals with the computational expression of the drift vector of the ISDE that is the MCMC generator of the posterior. Appendix G allows for verifying that the linearized ISDE is well adapted. The construction of the diffusion-maps basis for the posterior model is detailed in Appendix H. Finally, the Störmer-Verlet scheme for solving the reduced-order ISDE is given in Appendix I.

C. Soize
Université Paris-Est Marne-la-Vallée, Laboratoire Modélisation et Simulation Multi-Echelle, MSME UMR 8208, 5 bd Descartes, 77454 Marne-la-Vallée, France
Tel.: +33-1-60957661, Fax: ++33-1-60957799
E-mail: christian.soize@u-pem.fr

R.G. Ghanem
University of Southern California, 210 KAP Hall, Los Angeles, CA 90089, United States
E-mail: ghanem@usc.edu

C. Desceliers
Université Paris-Est Marne-la-Vallée, Laboratoire Modélisation et Simulation Multi-Echelle, MSME UMR 8208, 5 bd Descartes, 77454 Marne-la-Vallée, France
E-mail: christophe.desceliers@u-pem.fr

Concerning the numbering of the equations, the following convention is used. The label Eq. (n) refers to equation number n of the **Supplementary material**. On the other hand, the label Eq. (P-n) refers to the equation number n of the **paper**. Similarly, for the numbering of the sections, the label Section n refers to section number n of the **Supplementary material**. On the other hand, the label Section (P-n) refers to the section number n of the **paper**.

12 Applications (AP1) and (AP2)

In this section, two applications are presented and are used for performing the validation of the methodology and algorithms presented. All the random variables are defined on probability space $(\Theta, \mathcal{T}, \mathcal{P})$. These two applications will be referenced as (AP1) and (AP2) for application 1 and 2. These two applications are relatively simple and can be easily reproduced.

12.1 Stochastic model and simulated experiments for applications (AP1) and (AP2)

Stochastic model for (AP1) and (AP2) The stochastic model used for generating the initial dataset $\mathbb{D}_{N_d} = \{\mathbb{x}_d^j = (\mathbb{q}_d^j, \mathbb{w}_d^j), j = 1, \ldots, N_d\}$ (see Eq. (P-1)) relative to random variable $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$ in which $\mathbb{Q} = (\mathbb{Q}_1, \ldots, \mathbb{Q}_{n_q})$ and $\mathbb{W} = (\mathbb{W}_1, \ldots, \mathbb{W}_{n_w})$, is written, for (AP1) and (AP2), as

$$\mathbb{Q} = [B(\mathbb{U})]\,(\mathbb{W} + V\,\mathbf{b})\,,$$

in which $\mathbb{U}$, V, and $\mathbb{W}$ are independent random variables. The maximum value of N_d is 200 and $n_w = 20$. We have $n_q = 200$ for (AP1) and $n_q = 20\,000$ for (AP2). The deterministic vector $\mathbf{b}$ in $\mathbb{R}^{n_w}$ is written as $\mathbf{b} = 0.2\,\mathbf{u} + 0.9$ in which all the components of $\mathbf{u}$ belongs to $]0, 1[$ (generated

with the Matlab script: rng('default'); $\mathbf{u} = rand(n_w, 1)$). The real-valued random variable $V = 0.2\,\mathcal{U} + 0.9$ for (AP1) and $V = 0.2\,\mathcal{U} - 0.1$ for (AP2) in which $\mathcal{U}$ is a uniform random variable on $[0,1]$. The random vector $\mathbb{U} = (\mathbb{U}_1, \ldots, \mathbb{U}_{n_u})$ with $n_u = 6$ is written, for $\alpha = 1, \ldots, n_u$, as $\mathbb{U}_\alpha = 2\,u_\alpha\,\mathcal{U}_\alpha + 1 - u_\alpha$ in which $\mathcal{U}_1, \ldots, \mathcal{U}_{n_u}$ are n_u independent uniform random variables on $[0,1]$ and where, $u_\alpha = 0.2(\alpha-1)/(n_u-1)$ for (AP1) and $u_\alpha = 0.7(\alpha-1)/(n_u-1)$ for (AP2). The entries $[B(\mathbb{U})]_{kj}$ of the $(n_q \times n_w)$ random matrix are

$$[B(\mathbb{U})]_{kj} = \sum_{\alpha=1}^{n_u} \lambda_\alpha(\mathbb{U}_\alpha)\,\varphi_k^\alpha(\mathbb{U}_\alpha)\,\varphi_k^\alpha(\mathbb{U}_\alpha)\,\varphi_{j+n_q/2}^\alpha(\mathbb{U}_\alpha)\,.$$

For (AP1), $\varphi_k^\alpha(\mathbb{U}_\alpha) = \sin\{\alpha\, k\pi/(n_q+1)\}$ is independent of $\mathbb{U}_\alpha$ (deterministic) and $\lambda_\alpha(\mathbb{U}_\alpha) = 1/(\alpha\mathbb{U}_\alpha)^2$. For (AP2), $\varphi_k^\alpha(\mathbb{U}_\alpha) = \sin\{\alpha\,\mathbb{U}_\alpha\, k\pi/(n_q+1)\}$ and $\lambda_\alpha(\mathbb{U}_\alpha) = 5(1-\mathbb{U}_\alpha) + 1/(\alpha\mathbb{U}_\alpha)^2$. The random vector $\mathbb{W}$ is written as $\mathbb{W} = \sum_{\beta=1}^{3} \sqrt{\mu_\beta}\,\boldsymbol{\phi}^\beta\,\eta_\beta$, in which $\mu_\beta = 1/\beta^2$ and $\boldsymbol{\phi}^\beta = (\phi_1^\beta, \ldots, \phi_{n_w}^\beta)$ with $\phi_j^\beta = \sin\{\beta\pi j/(1+n_w)\}$. The non-Gaussian centered random vector $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)$ is written as $\boldsymbol{\eta} = \sum_{\gamma=1}^{27} \mathbf{y}^\gamma\,\psi_{\alpha_1^{(\gamma)}}(\Xi_1)\,\psi_{\alpha_2^{(\gamma)}}(\Xi_2)$ 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 Gaussian chaos. The matrix $[y] = [\mathbf{y}^1 \ldots \mathbf{y}^{27}]$ is such that $[y]\,[y]^T = [I_3]$ and is generated using the Matlab script: rng('default'); a_1 = $randn$(27,27); [a_2,] = $eig(a_1{*}(a_1)')$; $a_2(:, 4{:}27)$ = []; [y] = $(a_2)'$.

Simulated experiments for (AP1) and (AP2) The experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$ is generated with $n_r = 200$ independent experimental realizations $\{\mathbb{q}^{\text{exp},r}, r = 1, \ldots n_r\}$ of $\mathbb{Q}^{\text{exp}} = (\mathbb{Q}_1^{\text{exp}}, \ldots, \mathbb{Q}_{n_q}^{\text{exp}})$. As already explained, we also generate the independent experimental realizations $\{\mathbb{w}^{\text{exp},r}, r = 1, \ldots n_r\}$ of $\mathbb{W}^{\text{exp}} = (\mathbb{W}_1^{\text{exp}}, \ldots, \mathbb{W}_{n_w}^{\text{exp}})$ in order to validate the choice of the regularization parameter ε (see Section P-10). The experimental model is written, for (AP1) and (AP2), as

$$\mathbb{Q}^{\text{exp}} = [B(\mathbb{U}^{\text{exp}})]\,(\mathbb{W}^{\text{exp}} + V^{\text{exp}}\,\mathbf{b})\,,$$

in which $\mathbb{U}^{\text{exp}}$, V^{exp}, and $\mathbb{W}^{\text{exp}}$ are independent random variables that are also independent of $\mathbb{U}$, V, and $\mathbb{W}$. The real-valued random variable $V^{\text{exp}} = 0.2\mathcal{U}^{\text{exp}} + 0.9$ for (AP1) and $V^{\text{exp}} = 0.2\mathcal{U}^{\text{exp}} - 0.1$ for (AP2) in which $\mathcal{U}^{\text{exp}}$ is a uniform random variable on $[0,1]$ independent of $\mathcal{U}$. The random vector $\mathbb{U}^{\text{exp}} = (\mathbb{U}_1^{\text{exp}}, \ldots, \mathbb{U}_{n_u}^{\text{exp}})$ is written, for $\alpha = 1, \ldots, n_u$, as $\mathbb{U}_\alpha^{\text{exp}} = 2\,u_\alpha^{\text{exp}}\,\mathcal{U}_\alpha^{\text{exp}} + 1 - u_\alpha^{\text{exp}}$ in which $\mathcal{U}_1^{\text{exp}}, \ldots, \mathcal{U}_{n_u}^{\text{exp}}$ are n_u independent uniform random variables on $[0,1]$ and where, $u_\alpha^{\text{exp}} = 0.3(\alpha-1)/(n_u-1)$ for (AP1) and $u_\alpha = 0.7(\alpha-1)/(n_u-1)$ for (AP2). Note that for (AP1), the coefficient is 0.3 and not 0.2 as in the stochastic model. The mapping $\mathbb{u} \mapsto [B(\mathbb{u})]$ is the same as the one of the stochastic model. The random vector $\mathbb{W}^{\text{exp}}$ is written as $\mathbb{W}^{\text{exp}} = 0.2 \times \mathbf{1} + \widetilde{\mathbb{W}}^{\text{exp}}$ in which $\mathbf{1} \in \mathbb{R}^{n_w}$ is the vector whose components are equal to 1 and where $\widetilde{\mathbb{W}}^{\text{exp}}$ is an independent copy of the stochastic model of $\mathbb{W}$.

12.2 Values of the numerical parameters for the computation of (AP1) and (AP2)

Table 1 summarizes the values of all the numerical parameters introduced in the algorithms. Except for regularization parameter ε and for the convergence of the learning in probability distribution with respect to dimension N_d of initial dataset $\mathbb{D}_{N_d}$ (theses two parameters will be the subject of a particular analysis presented later), the other values of the numerical parameters have been obtained by using the existing criteria or by performing a local convergence analysis.

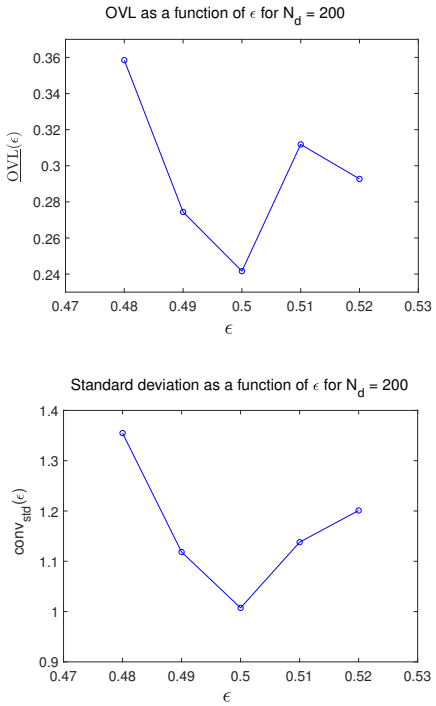

Fig. 1 Application AP1: validation of the choice $\varepsilon = 0.5$. For $N_d = 200$, graph of $\varepsilon \mapsto \underline{\text{OVL}}(\varepsilon)$ (up) and graph of $\varepsilon \mapsto \text{conv}_{\text{std}}(\varepsilon)$ (down).

12.3 Quantities used for validating the choice of ε and for studying the convergence in probability distribution of the probabilistic learning

Definition of the graphs that are plotted As already explained in Section P-10, we propose the value 0.5 for the regular-

Table 1 For applications (AP1) and (AP2), Table 1 defines the values of all the parameters introduced in the algorithm, and which are used in computations.

	Parameters	(AP1)	(AP2)
	n_q	200	20 000
	n_w	20	20
	n	220	20 020
	N_d	≤ 200	≤ 200
	$\max N_d$	200	200
	n_r	200	200
Learning (Appendix A)	ν_x	9	15
	$\varepsilon_{\text{diff}}$	48	120
	m	12	17
	f_0	1.5	1.5
	n_{MC}	150	150
	ν_{ar}	30 000	30 000
	Δt	0.1649	0.0903
	M_0	100	100
	ℓ_0	100	100
PCA of **Q** and **W**	ν_q	6	12
	$\text{err}_{\mathbf{Q}}(\nu_q)$	$4.3e{-5}$	$5.3e{-5}$
	ν_w	3	3
	$\text{err}_{\mathbf{W}}(\nu_w)$	$3.7e{-15}$	$5.3e{-15}$
Posterior	ν	9	15
	ν_1	6	12
	ε	0.5	0.5
	N_s	200	200
	ε_{N_s}	0.0422	0.0322
	f_0^{post}	10^{-5}	10^{-5}
	$\varepsilon_{\text{diff}}^{\text{post}}$	4 000	3 000
	m_{post}	9	11
	$n_{\text{MC}}^{\text{post}}$	200	200
	ν_{post}	40 000	40 000
	Δt	0.0277	0.03384
	M_0^{post}	100	100
	ℓ_0^{post}	10 000	10 000

ization parameter ε. In order to validate this choice, for the applications and for N_d fixed, we have plotted:

(i) the graph of function $\varepsilon \mapsto \underline{\text{OVL}}(\varepsilon)$ defined by Eq. (P-77), which has to be minimum in the neighborhood of $\varepsilon = 0.5$;

(ii) the graph of function $\varepsilon \mapsto \text{conv}_{\text{std}}(\varepsilon)$ that is defined hereinafter and whose value should be close to 1 in the neighborhood of $\varepsilon = 0.5$. Let $\boldsymbol{\sigma}_{\mathbb{W}}^{\text{post}}(\varepsilon) = (\sigma_{\mathbb{W}_1}^{\text{post}}(\varepsilon), \ldots, \sigma_{\mathbb{W}_{m_w}}^{\text{post}}(\varepsilon))$ be the vector of the standard deviations estimated with the ν_{post} realizations of the components of random vector $\mathbb{W}^{\text{post}}$ (estimated with the ν_{post} realizations) and let $\boldsymbol{\sigma}_{\mathbb{W}}^{\text{exp}} = (\sigma_{\mathbb{W}_1}^{\text{exp}}, \ldots, \sigma_{\mathbb{W}_{m_w}}^{\text{exp}})$ be the standard deviations of the components of random vector $\mathbb{W}$ (estimated with experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$). The function conv_{std} is $\text{conv}_{\text{std}}(\varepsilon) = \|\boldsymbol{\sigma}_{\mathbb{W}}^{\text{post}}(\varepsilon)\|/\|\boldsymbol{\sigma}_{\mathbb{W}}^{\text{exp}}\|$.

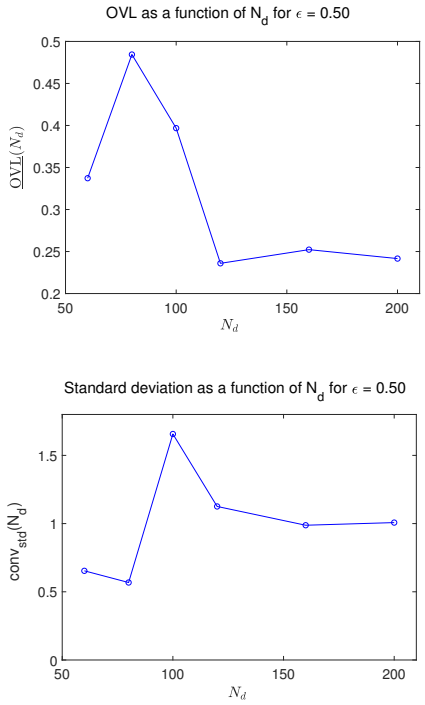

Fig. 2 Application AP1: convergence of the probabilistic learning with respect to N_d. For $\varepsilon = 0.5$, graph of $N_d \mapsto \underline{\text{OVL}}(N_d)$ (up) and graph of $N_d \mapsto \text{conv}_{\text{std}}(N_d)$ (down).

Studying the convergence of the probabilistic learning for the posterior model For ε fixed at 0.5, the convergence in probability distribution of the probabilistic learning is analyzed with respect to N_d by studying the function $N_d \mapsto \underline{\text{OVL}}(N_d)$ defined by Eq. (P-77) (replacing ε by N_d) and the function $N_d \mapsto \text{conv}_{\text{std}}(N_d)$ that is defined by $\text{conv}_{\text{std}}(N_d) = \|\boldsymbol{\sigma}_{\mathbb{W}}^{\text{post}}(N_d)\|/\|\boldsymbol{\sigma}_{\mathbb{W}}^{\text{exp}}\|$.

Validating the proposed method In addition to the quantities just described and for several components of index k, we will compare the graph of the pdf $w \mapsto p_{\mathbb{W}_k}^{d}(w)$ of $\mathbb{W}_k$ estimated with the initial dataset $\mathbb{D}_{N_d}$ of $N_d = 200$ realizations, with the graph of the pdf $w \mapsto p_{\mathbb{W}_k}^{\text{exp}}(w)$ of $\mathbb{W}_k$ estimated with the experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$ realizations, and with the graph of the pdf $w \mapsto p_{\mathbb{W}_k}^{\text{post}}(w)$ of $\mathbb{W}_k^{\text{post}}$ estimated for $\varepsilon = 0.5$ and $N_d = 200$, and $\nu_{\text{post}} = 40\,000$ realizations. The pdf $w \mapsto p_{\mathbb{W}_k}^{\text{post}}(w)$ should be close to $w \mapsto p_{\mathbb{W}_k}^{\text{exp}}(w)$ (the reference).

12.4 Results and comments for applications (AP1) and (AP2)

The results are presented in Figs. 1 to 3 for application (AP1) and in Figs. 4 to 6 for application (AP2).
(i) Concerning the validation of the choice $\varepsilon = 0.5$ of the regularization parameter, Figs. 1 and 4 show that function

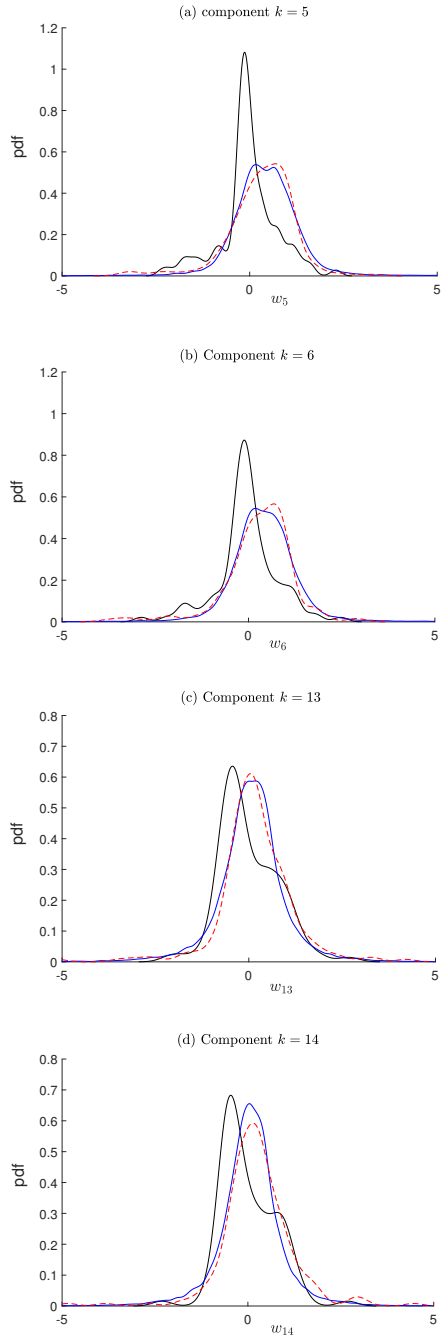

Fig. 3 Application AP1: pdf $w \mapsto p^d_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the initial dataset $\mathbb{D}_{N_d}$ of $N_d = 200$ realizations (thin black line), pdf $w \mapsto p^{\text{exp}}_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the experimental dataset $\mathbb{D}^{\text{exp}}_{n_r}$ of $n_r = 200$ realizations (thick red dashed line), pdf $w \mapsto p^{\text{post}}_{\mathbb{W}_k}(w)$ of $\mathbb{W}^{\text{post}}_k$ estimated with $\varepsilon = 0.5$, $N_d = 200$, and $\nu_{\text{post}} = 40\,000$ realizations (thick blue line), for $k = 1$ (a), $k = 37$ (b), $k = 145$ (c), and $k = 613$ (d).

$\varepsilon \mapsto \underline{\text{OVL}}(\varepsilon)$ has effectively a minimum in the neighborhood of $\varepsilon = 0.5$ for these two applications.

(ii) For the two applications with $\varepsilon = 0.5$, the convergence in probability distribution of the probabilistic learning with

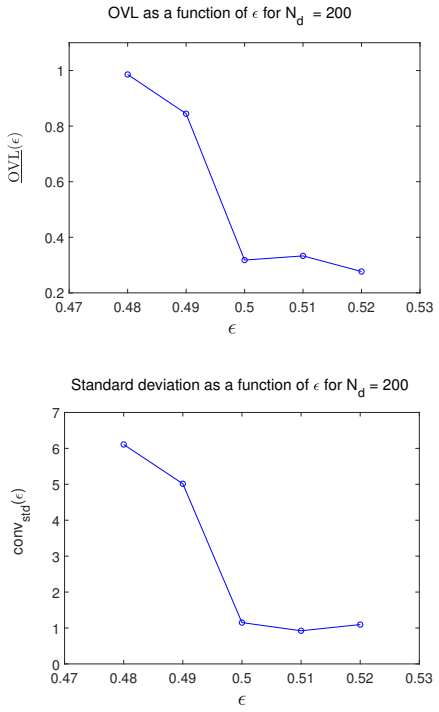

Fig. 4 Application AP2: validation of the choice $\varepsilon = 0.5$. For $N_d = 200$, graph of $\varepsilon \mapsto \underline{\text{OVL}}(\varepsilon)$ (up) and graph of $\varepsilon \mapsto \text{conv}_{\text{std}}(\varepsilon)$ (down).

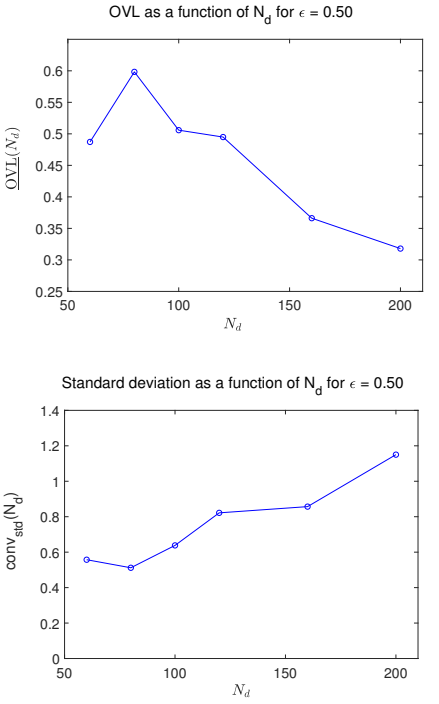

Fig. 5 Application AP2: convergence of the probabilistic learning with respect to N_d. For $\varepsilon = 0.5$, graph of $N_d \mapsto \underline{\text{OVL}}(N_d)$ (up) and graph of $N_d \mapsto \text{conv}_{\text{std}}(N_d)$ (down).

respect to the size N_d of the initial dataset that is used in all

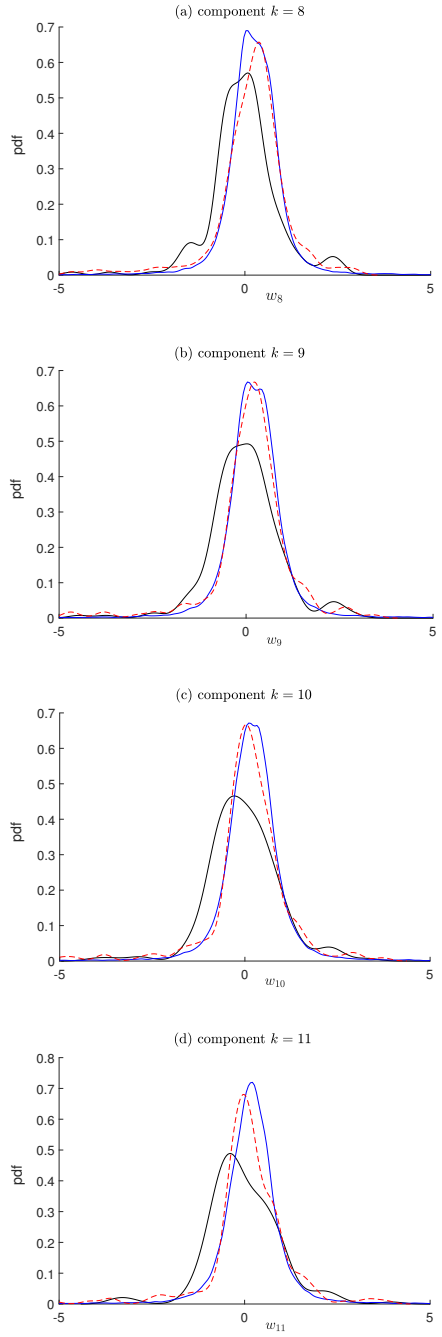

Fig. 6 Application AP2: pdf $w \mapsto p^d_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the initial dataset $\mathbb{D}_{N_d}$ of $N_d = 200$ realizations (thin black line), pdf $w \mapsto p^{\text{exp}}_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the experimental dataset $\mathbb{D}^{\text{exp}}_{n_r}$ of $n_r = 200$ realizations (thick red dashed line), pdf $w \mapsto p^{\text{post}}_{\mathbb{W}_k}(w)$ of $\mathbb{W}^{\text{post}}_k$ estimated with $\varepsilon = 0.5$, $N_d = 200$, and $\nu_{\text{post}} = 40\,000$ realizations (thick blue line), for $k = 5$ (a), $k = 6$ (b), $k = 13$ (c), and $k = 14$ (d).

the calculations detailed in Sections P-4 to P-9, Figs. 2 and 5 show the results obtained for the functions $N_d \mapsto \underline{\text{OVL}}(N_d)$ and $N_d \mapsto \text{conv}_{\text{std}}(N_d)$. For application (AP2), the convergence of the learning is slower and a best convergence could certainly be obtained by increasing the maximum value of N_d that should be considered. Nevertheless, this slower convergence of the learning with respect to N_d does not interfere with the validation of the proposed methodology, because, for a fixed value of N_d, the results obtained show that the posterior model that is estimated allows the prior model to be significantly improved; a better convergence of the learning with respect to N_d would lead to even greater improvement of the posterior model.

(iii) Concerning the validation of the proposed method, Figs. 1 and 4 show that, for $N_d = 200$ and $\varepsilon = 0.5$, the norm $\text{conv}_{\text{std}}(\varepsilon)$ of the vector of the standard deviations, normalized by its counterpart for the experiments, is close to 1. For these two applications, Figs. 3 and 6 show, for selected components $\mathbb{W}_k$ of random vector $\mathbb{W}$, the comparison of three probability density functions: the pdf $w \mapsto p^d_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the initial dataset $\mathbb{D}_{N_d}$ with $N_d = 200$, the pdf $w \mapsto p^{\text{exp}}_{\mathbb{W}_k}(w)$ of $\mathbb{W}_k$ estimated with the experimental dataset $\mathbb{D}^{\text{exp}}_{n_r}$ with $n_r = 200$, and the pdf $w \mapsto p^{\text{post}}_{\mathbb{W}_k}(w)$ of the posterior $\mathbb{W}^{\text{post}}_k$ estimated with $\varepsilon = 0.5$, $N_d = 200$, and $\nu_{\text{post}} = 40\,000$. For each value of k that is considered, the comparison between $p^d_{\mathbb{W}_k}$ and $p^{\text{exp}}_{\mathbb{W}_k}$ shows that there are significant differences (mean value, standard deviation, non-Gaussianity) between these two pdf's, which justify the use of the Bayesian approach for improving $p^d_{\mathbb{W}_k}$ with $p^{\text{post}}_{\mathbb{W}_k}$. The values of k selected for plotting, for each application, correspond to the greatest difference between these two pdfs. An important element for the validation is the comparison between $p^{\text{post}}_{\mathbb{W}_k}$ and $p^{\text{exp}}_{\mathbb{W}_k}$. It can be seen that the results are quite good for these two applications.

Appendix A. Summary of the algorithm of the probabilistic learning on manifolds

In this Appendix, we summarize the algorithm of the probabilistic learning on manifolds (PLoM) that is used in Section P-5. This algorithm has been introduced in [16]. Complementary developments can be found in [3,17–19]. Applications and validations can be found in [4,5,20,15]. In addition, we give the formula for estimating the values of the two hyperparameters that control the algorithm of the PLoM.

Let $\{\mathbf{x}^j_d = (\mathbf{q}^j_d, \mathbf{w}^j_d), j = 1, \ldots, N_d\}$ be the set of the N_d independent realizations given in $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ with $n = n_q + n_w$, which constitute the *initial data set* D_{N_d}. Let $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ be the random variable with values in $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ for which $\{\mathbf{x}^j_d, j = 1, \ldots, N_d\}$ constitutes N_d independent realizations. The objective of the PLoM is to generate $\nu_{\text{ar}} \gg N_d$ additional realizations $\{\mathbf{x}^\ell_{\text{ar}}, \ell = 1, \ldots, \nu_{\text{ar}}\}$ of random vector $\mathbf{X}$. As soon as the set $\{\mathbf{x}^\ell_{\text{ar}}, \ell = 1, \ldots, \nu_{\text{ar}}\}$ has been constructed, the additional realizations for $\mathbf{Q}$ and

$\mathbf{W}$ can be extracted as $(\mathbf{q}_{\text{ar}}^{\ell}, \mathbf{w}_{\text{ar}}^{\ell}) = \mathbf{x}_{\text{ar}}^{\ell}$ for $\ell = 1, \ldots, \nu_{\text{ar}}$, which constitute the *learned dataset* $D_{\nu_{\text{ar}}}$.

A.1. Normalization of the initial dataset. The N_d independent realizations $\{\mathbf{x}_d^j, j = 1, \ldots, N_d\}$ of $\mathbf{X}$ with values in $\mathbb{R}^n$ can be represented by the matrix $[x_d] = [\mathbf{x}_d^1 \ldots \mathbf{x}_d^{N_d}]$ in $\mathbb{M}_{n,N_d}$. Let $[\mathbf{X}] = [\mathbf{X}^1, \ldots, \mathbf{X}^{N_d}]$ be the random matrix with values in $\mathbb{M}_{n,N_d}$, whose columns are N_d independent copies of random vector $\mathbf{X}$. Therefore, $[x_d]$ is one realization of random matrix $[\mathbf{X}]$. The normalization of random matrix $[\mathbf{X}]$ is attained with the random matrix $[\mathbf{H}] = [\mathbf{H}^1, \ldots, \mathbf{H}^{N_d}]$ with values in $\mathbb{M}_{\nu_x,N_d}$ with $\nu_x \leq n$, obtained by using the principal component analysis of random vector $\mathbf{X}$. Consequently, random matrix $[\mathbf{X}]$ is written as,

$$[\mathbf{X}] = [\underline{x}] + [\varphi]\,[\lambda]^{1/2}\,[\mathbf{H}]\,,$$

in which $[\lambda]$ is the $(\nu_x \times \nu_x)$ diagonal matrix of the ν_x positive eigenvalues of the empirical estimate of the covariance matrix of $\mathbf{X}$ (computed using $\mathbf{x}_d^1, \ldots, \mathbf{x}_d^{N_d}$), where $[\varphi]$ is the $(n \times \nu_x)$ matrix of the associated eigenvectors such $[\varphi]^T\,[\varphi] = [I_{\nu_x}]$, and where $[\underline{x}]$ is the matrix in $\mathbb{M}_{n,N_d}$ with identical columns, each one being equal to the empirical estimate $\underline{\mathbf{x}} \in \mathbb{R}^n$ of the mean value of random vector $\mathbf{X}$ (computed using $\mathbf{x}_d^1, \ldots, \mathbf{x}_d^{N_d}$). The columns of $[\mathbf{H}]$ are N_d independent copies of a random vector $\mathbf{H}$ wit values in $\mathbb{R}^{\nu_x}$. The realization $[\eta_d] = [\boldsymbol{\eta}^1 \ldots \boldsymbol{\eta}^{N_d}] \in \mathbb{M}_{\nu_x,N_d}$ of $[\mathbf{H}]$ (associated with the realization $[x_d]$ of $[\mathbf{X}]$) is computed by $[\eta_d] = [\lambda]^{-1/2}[\varphi]^T\,([x_d] - [\underline{x}])$. When n is small, ν_x can be chosen as n. If some eigenvalues are zero, they must be eliminated and then $\nu_x < n$. When n is high, a statistical reduction can be done as usual and therefore $\nu_x < n$ in such a case.

A.2. Diffusion-maps basis. This is an algebraic basis of vector space $\mathbb{R}^{N_d}$, which is constructed using the diffusion maps proposed in [2]. Let $[\mathbb{b}]$ be the positive-definite diagonal real matrix in $\mathbb{M}_{N_d}$ such that $[\mathbb{b}]_{ij} = \delta_{ij}\,\sum_{j'=1}^{N_d}[K]_{jj'}$ in which $[K]_{jj'} = \exp(-\frac{1}{4\,\varepsilon_{\text{diff}}}\|\boldsymbol{\eta}^j - \boldsymbol{\eta}^{j'}\|^2)$, depending on a real smoothing parameter $\varepsilon_{\text{diff}} > 0$. Let $[\mathbb{P}]$ be the transition matrix in $\mathbb{M}_{N_d}$ of a Markov chain such that $[\mathbb{P}] = [\mathbb{b}]^{-1}\,[K]$. For $1 < m \leq N_d$, let $\mathbf{g}^1, \ldots, \mathbf{g}^m$ be the right eigenvectors in $\mathbb{R}^{N_d}$ of matrix $[\mathbb{P}]$ such that $[\mathbb{P}]\,\mathbf{g}^{\alpha} = \Lambda_{\alpha}\,\mathbf{g}^{\alpha}$, whose eigenvalues are real and such that $1 = \Lambda_1 > \Lambda_2 > \ldots > \Lambda_m$. The normalization condition of these eigenvectors is $[g]^T\,[\mathbb{b}]\,[g] = [I_m]$, in which $[g] = [\mathbf{g}^1 \ldots \mathbf{g}^m] \in \mathbb{M}_{N_d,m}$ is the *diffusion-maps basis*. The eigenvector $\mathbf{g}^1$ associated with the largest eigenvalue $\Lambda_1 = 1$ is a constant vector. For $m = N_d$, the diffusion-maps basis is an algebraic basis of $\mathbb{R}^{N_d}$. The right-eigenvalue problem of the nonsymmetric matrix $[\mathbb{P}]$ can be performed solving the eigenvalue problem $[\mathbb{b}]^{-1/2}\,[K]\,[\mathbb{b}]^{-1/2}\,\boldsymbol{\Phi}^{\alpha} = \Lambda_{\alpha}\,\boldsymbol{\Phi}^{\alpha}$ related to a positive-definite symmetric real matrix, and with the normalization $[\Phi]^T\,[\Phi] = [I_m]$, in which $[\Phi] = [\boldsymbol{\Phi}^1 \ldots \boldsymbol{\Phi}^m]$. Therefore, $\mathbf{g}^{\alpha}$ can be deduced from $\boldsymbol{\Phi}^{\alpha}$ by $\mathbf{g}^{\alpha} = [\mathbb{b}]^{-1/2}\,\boldsymbol{\Phi}^{\alpha}$. The construction introduces two hyperparameters: the dimension $m \leq N_d$ and the smoothing parameter $\varepsilon_{\text{diff}} > 0$. An algorithm is proposed in [17] for estimating their values. Most of the time, m and $\varepsilon_{\text{diff}}$ can be chosen as follows. Let $\varepsilon_{\text{diff}} \mapsto \widehat{m}(\varepsilon_{\text{diff}})$ be the function from $\mathbb{R}^{+*} =]0\,, +\infty[$ into $\mathbb{N}$ such that

$$\widehat{m}(\varepsilon_{\text{diff}}) = \arg \min_{\alpha\,|\,\alpha \geq 3} \left\{ \frac{\Lambda_{\alpha}(\varepsilon_{\text{diff}})}{\Lambda_2(\varepsilon_{\text{diff}})} < 0.1 \right\}. \tag{1}$$

If function $\widehat{m}$ is a decreasing function of $\varepsilon_{\text{diff}}$ in the broad sense (if not, see [17]), then the optimal value $\varepsilon_{\text{diff}}^{\text{opt}}$ of $\varepsilon_{\text{diff}}$ can be chosen as the smallest value of the integer $\widehat{m}(\varepsilon_{\text{diff}}^{\text{opt}})$ such that

$$\{\widehat{m}(\varepsilon_{\text{diff}}^{\text{opt}}) < \widehat{m}(\varepsilon_{\text{diff}})\,, \forall \varepsilon_{\text{diff}} \in]0, \varepsilon_{\text{diff}}^{\text{opt}}[\,\} \cap \{\widehat{m}(\varepsilon_{\text{diff}}^{\text{opt}}) = \widehat{m}(\varepsilon_{\text{diff}})\,, \forall \varepsilon_{\text{diff}} \in]\varepsilon_{\text{diff}}^{\text{opt}}, 1.5\,\varepsilon_{\text{diff}}^{\text{opt}}[\,\}\,. \tag{2}$$

The corresponding optimal value m^{opt} of m is then given by $m^{\text{opt}} = \widehat{m}(\varepsilon_{\text{diff}}^{\text{opt}})$.

A.3. Reduced-order representation of random matrices $[\boldsymbol{H}]$ and $[\boldsymbol{X}]$. The diffusion-maps vectors $\mathbf{g}^1, \ldots, \mathbf{g}^m \in \mathbb{R}^{N_d}$ span a subspace of $\mathbb{R}^{N_d}$ that characterizes, for the optimal values m^{opt} and $\varepsilon_{\text{diff}}^{\text{opt}}$ of m and $\varepsilon_{\text{diff}}$, the local geometry structure of the dataset $\{\boldsymbol{\eta}^j, j = 1, \ldots, N_d\}$. The reduced-order representation is obtained by projecting each column of the $\mathbb{M}_{N_d,\nu_x}$-valued random matrix $[\mathbf{H}]^T$ on the subspace of $\mathbb{R}^{N_d}$, spanned by $\{\mathbf{g}^1, \ldots, \mathbf{g}^m\}$. Let $[\mathbf{Z}]$ be the random matrix with values in $\mathbb{M}_{\nu_x,m}$ such that

$$[\mathbf{H}] = [\mathbf{Z}]\,[g]^T\,. \tag{3}$$

Since the eigenvector $\mathbf{g}^1$ is a constant vector and since random matrix $[\mathbf{H}]$ is centered, this eigenvector can be removed from the basis. As the matrix $[g]^T\,[g] \in \mathbb{M}_m$ is invertible, the least-squares approximation of $[\mathbf{Z}]$ is written as $[\mathbf{Z}] = [\mathbf{H}]\,[a]$ in which

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

and the realization $[z_d] \in \mathbb{M}_{\nu_x,m}$ of $[\mathbf{Z}]$ is written as

$$[z_d] = [\eta_d]\,[a] \in \mathbb{M}_{\nu_x,m}\,.$$

A.4. Generation of additional realizations $\{\boldsymbol{\eta}_{\text{ar}}^{\ell}, \ell = 1, \ldots, \nu_{\text{ar}}\}$ of random vector $\boldsymbol{H}$.

An MCMC generator for random matrix $[\mathbf{H}]$ is constructed using the approach proposed in [13, 14] belonging to the class of Hamiltonian Monte Carlo methods [13, 6, 7]. The generation of additional realizations $[z_{\text{ar}}^1], \ldots, [z_{\text{ar}}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{Z}]$ is carried out by using an unusual MCMC method based on a reduced-order Itô stochastic differential equation (ISDE) that is constructed as the projection on the diffusion-maps basis of the ISDE related to a dissipative

Hamiltonian dynamical system for which the invariant measure is the pdf of random matrix $[\mathbf{H}]$ constructed with the Gaussian kernel-density estimation method and $[\eta_d]$. This method preserves the concentration of the probability measure and avoids the scatter phenomenon. Let $\{([\boldsymbol{\mathcal{Z}}(t)], [\boldsymbol{\mathcal{Y}}(t)]), t \in \mathbb{R}^+\}$ be the unique asymptotic (for $t \to +\infty$) stationary diffusion stochastic process with values in $\mathbb{M}_{\nu_x,m} \times \mathbb{M}_{\nu_x,m}$, of the following reduced-order ISDE (stochastic nonlinear second-order dissipative Hamiltonian dynamical system), for $t > 0$,

$$d[\boldsymbol{\mathcal{Z}}(t)] = [\boldsymbol{\mathcal{Y}}(t)]\, dt\,,$$

$$d[\boldsymbol{\mathcal{Y}}(t)] = [\mathcal{L}([\boldsymbol{\mathcal{Z}}(t)])]\, dr - \frac{1}{2} f_0\, [\boldsymbol{\mathcal{Y}}(t)]\, dt + \sqrt{f_0}\, [d\boldsymbol{\mathcal{W}}^{\text{wien}}(t)]\,,$$

with the initial condition $[\boldsymbol{\mathcal{Z}}(0)] = [z_d]$ and $[\boldsymbol{\mathcal{Y}}(0)] = [\boldsymbol{\mathcal{N}}]\,[a]$ almost surely.

(i) The random matrix $[\mathcal{L}([\boldsymbol{\mathcal{Z}}(t)])]$ with values in $\mathbb{M}_{\nu_x,m}$ is such that $[\mathcal{L}([\boldsymbol{\mathcal{Z}}(t)])] = [L([\boldsymbol{\mathcal{Z}}(t)]\,[g]^T)]\,[a]$. For all $[u] = [\mathbf{u}^1 \ldots \mathbf{u}^{N_d}]$ in $\mathbb{M}_{\nu_x,N_d}$ with $\mathbf{u}^j = (u_1^j, \ldots, u_\nu^j)$ in $\mathbb{R}^{\nu_x}$, the matrix $[L([u])]$ in $\mathbb{M}_{\nu_x,N_d}$ is defined, for all $k = 1, \ldots, \nu_x$ and for all $j = 1, \ldots, N_d$, by

$$[L([u])]_{kj} = \frac{1}{p(\mathbf{u}^j)}\, \{\boldsymbol{\nabla}_{\mathbf{u}^j}\, p(\mathbf{u}^j)\}_k\,,$$

$$p(\mathbf{u}^j) = \frac{1}{N_d} \sum_{j'=1}^{N_d} \exp\{-\frac{1}{2\widehat{s}_{\nu_x}^2} \| \frac{\widehat{s}_{\nu_x}}{s_{\nu_x}} \boldsymbol{\eta}^{j'} - \mathbf{u}^j \|^2\}\,,$$

$$\boldsymbol{\nabla}_{\mathbf{u}^j}\, p(\mathbf{u}^j) = \frac{1}{\widehat{s}_{\nu_x}^2\, N_d} \sum_{j'=1}^{N_d} (\frac{\widehat{s}_{\nu_x}}{s_{\nu_x}} \boldsymbol{\eta}^{j'} - \mathbf{u}^j) \exp\{-\frac{1}{2\widehat{s}_{\nu_x}^2} \| \frac{\widehat{s}_{\nu_x}}{s_{\nu_x}} \boldsymbol{\eta}^{j'} - \mathbf{u}^j \|^2\}\,,$$

in which $\widehat{s}_{\nu_x}$ is the modified Silverman bandwidth s_{ν_x}, which has been introduced in [14],

$$\widehat{s}_{\nu_x} = \frac{s_{\nu_x}}{\sqrt{s_{\nu_x}^2 + \frac{N_d - 1}{N_d}}} \quad , \quad s_{\nu_x} = \left\{ \frac{4}{N_d(2 + \nu_x)} \right\}^{1/(\nu_x + 4)}\,.$$

(ii) $[\boldsymbol{\mathcal{W}}^{\text{wien}}(t)] = [\mathbb{W}^{\text{wien}}(t)]\,[a]$ where $\{[\mathbb{W}^{\text{wien}}(t)], t \in \mathbb{R}^+\}$ is the $\mathbb{M}_{\nu_x,N_d}$-valued normalized Wiener stochastic process.
(iii) $[\boldsymbol{\mathcal{N}}]$ is the $\mathbb{M}_{\nu_x,N_d}$-valued normalized Gaussian random matrix that is independent of stochastic process $[\mathbb{W}^{\text{wien}}]$.
(iv) The free parameter f_0, such that $0 < f_0 < 4$, allows the dissipation term of the nonlinear second-order dynamical system (dissipative Hamiltonian system) to be controlled in order to kill the transient part induced by the initial conditions. A common value is $f_0 = 1.5$.
(v) We then have $[\mathbf{Z}] = \lim_{t \to +\infty} [\boldsymbol{\mathcal{Z}}(t)]$ in probability distribution, which allows for generating the additional realizations, $[z_{\text{ar}}^1], \ldots, [z_{\text{ar}}^{n_{\text{MC}}}]$, and then, generating the additional realizations $[\eta_{\text{ar}}^1], \ldots, [\eta_{\text{ar}}^{n_{\text{MC}}}]$ by using Eq. (3), such that $[\eta_{\text{ar}}^\ell] = [z_{\text{ar}}^\ell]\,[g]^T$ (see Section A.6).

A.5. Algorithm for solving the reduced-order ISDE. Let $M = n_{\text{MC}} \times M_0$ be the positive integer in which n_{MC} and M_0 are integers. The reduced-order ISDE is solved on the finite interval $\mathcal{R} = [0\,, M\,\Delta t]$, in which Δt is the sampling step of the continuous index parameter t. The integration scheme is based on the use of the $M + 1$ sampling points $t_{\ell'}$ such that $t_{\ell'} = \ell'\,\Delta t$ for $\ell' = 0, \ldots, M$ for which $[\boldsymbol{\mathcal{Z}}_{\ell'}] = [\boldsymbol{\mathcal{Z}}(t_{\ell'})]$, $[\boldsymbol{\mathcal{Y}}_{\ell'}] = [\boldsymbol{\mathcal{Y}}(t_{\ell'})]$, and $[\boldsymbol{\mathcal{W}}_{\ell'}^{\text{wien}}] = [\boldsymbol{\mathcal{W}}^{\text{wien}}(t_{\ell'})]$, with $[\boldsymbol{\mathcal{Z}}_0] = [z_d]$, $[\boldsymbol{\mathcal{Y}}_0] = [\boldsymbol{\mathcal{N}}]\,[a]$, and $[\boldsymbol{\mathcal{W}}_0^{\text{wien}}] = [0_{\nu_x,m}]$. For $\ell' = 0, \ldots, M - 1$, let $[\Delta\boldsymbol{\mathcal{W}}_{\ell'+1}^{\text{wien}}] = [\Delta\mathbb{W}_{\ell'+1}^{\text{wien}}]\,[a]$ be the sequence of random matrices with values in $\mathbb{M}_{\nu_x,m}$, in which the increments $[\Delta\mathbb{W}_1^{\text{wien}}], \ldots, [\Delta\mathbb{W}_M^{\text{wien}}]$ are M independent random matrices with values in $\mathbb{M}_{\nu_x,N_d}$. For all $k = 1, \ldots, \nu_x$ and for all $j = 1, \ldots, N_d$, the real-valued random variables $\{[\Delta\mathbb{W}_{\ell'+1}^{\text{wien}}]_{kj}\}_{kj}$ are independent, Gaussian, second-order, and centered random variables such that

$$E\{[\Delta\mathbb{W}_{\ell'+1}^{\text{wien}}]_{kj} [\Delta\mathbb{W}_{\ell'+1}^{\text{wien}}]_{k'j'}\} = \Delta t\, \delta_{kk'}\, \delta_{jj'}\,.$$

For $\ell' = 0, \ldots, M - 1$, the Störmer-Verlet scheme applied to the reduced-order ISDE yields

$$[\boldsymbol{\mathcal{Z}}_{\ell'+\frac{1}{2}}] = [\boldsymbol{\mathcal{Z}}_{\ell'}] + \frac{\Delta t}{2}\, [\boldsymbol{\mathcal{Y}}_{\ell'}]\,,$$

$$[\boldsymbol{\mathcal{Y}}_{\ell'+1}] = \frac{1 - b}{1 + b}\, [\boldsymbol{\mathcal{Y}}_{\ell'}] + \frac{\Delta t}{1 + b}\, [\boldsymbol{\mathcal{L}}_{\ell'+\frac{1}{2}}] + \frac{\sqrt{f_0}}{1 + b}\, [\Delta\boldsymbol{\mathcal{W}}_{\ell'+1}^{\text{wien}}]\,,$$

$$[\boldsymbol{\mathcal{Z}}_{\ell'+1}] = [\boldsymbol{\mathcal{Z}}_{\ell'+\frac{1}{2}}] + \frac{\Delta t}{2}\, [\boldsymbol{\mathcal{Y}}_{\ell'+1}]\,,$$

with the initial condition defined before, where $b = f_0\, \Delta t\, /4$, and where $[\boldsymbol{\mathcal{L}}_{\ell'+\frac{1}{2}}]$ is the $\mathbb{M}_{\nu_x,m}$-valued random variable such that

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

A.6. Additional realizations $\{\boldsymbol{x}_{\text{ar}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$ of random vector $\boldsymbol{X}$. The reduced-order ISDE is then used for generating n_{MC} additional realizations, $[z_{\text{ar}}^1], \ldots, [z_{\text{ar}}^{n_{\text{MC}}}]$ in $\mathbb{M}_{\nu_x,m}$, of random matrix $[\mathbf{Z}]$, and therefore, for deducing the n_{MC} additional realizations, $[\eta_{\text{ar}}^1], \ldots, [\eta_{\text{ar}}^{n_{\text{MC}}}]$ in $\mathbb{M}_{\nu_x,N_d}$ of random matrix $[\mathbf{H}]$, such that $[\eta_{\text{ar}}^\ell] = [z_{\text{ar}}^\ell]\,[g]^T$ for $\ell = 1, \ldots, n_{\text{MC}}$. The computation is performed as follows. Let $\nu_{\text{ar}} = n_{\text{MC}} \times N_d$, in which n_{MC} is a any given integer. Let $[\mathbb{W}^{\text{wien}}(\cdot;\theta)]$ with $\theta \in \Theta$ be a realization of the Wiener stochastic process $[\mathbb{W}^{\text{wien}}]$ defined in Section A.4-(ii). Let $\{([\boldsymbol{\mathcal{Z}}(t;\theta)], [\boldsymbol{\mathcal{Y}}(t;\theta)]), t \in \mathbb{R}^+\}$ be one realization of the $(\mathbb{M}_{\nu_x,m} \times \mathbb{M}_{\nu_x,m})$-valued stochastic process $\{([\boldsymbol{\mathcal{Z}}(t)], [\boldsymbol{\mathcal{Y}}(t)]), t \in \mathbb{R}^+\}$, for which its time-sampling is computed using the algorithm presented in Section A.5. Let ℓ_0 be the integer such that, for $t \geq \ell_0\, \Delta t$, the solution is asymptotic to the stationary solution. Therefore, the independent realizations $\{\boldsymbol{\eta}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$ of $\mathbf{H}$ are generated as follows. Let M_0 be a given positive integer.

For $\kappa = 1, \ldots, n_{\mathrm{MC}}$ and for $t_{\ell'} = \ell' \, \Delta t$ with $\ell' = \ell_0 + \kappa \, M_0$, we have, for $j = 1, \ldots, N_d$ and for $k = 1, \ldots, \nu_x$, $\eta_k^\ell = \{[\boldsymbol{\mathcal{Z}}(t_{\ell'}, \theta)] \, [g]^T\}_{kj}$ with $\ell = j + (\ell' - 1) \, N_d$. In this method of generation, only one realization θ is used and M_0 is chosen sufficiently large in order that $[\boldsymbol{\mathcal{Z}}(t_{\ell'})]$ and $[\boldsymbol{\mathcal{Z}}(t_{(\ell'+M_0)})]$ be two random matrices that are approximately independent. The realizations $\{\mathbf{x}_{\mathrm{ar}}^\ell, \ell = 1, \ldots, \nu_{\mathrm{ar}}\}$ of random vector $\mathbf{X}$ are then calculated by $\mathbf{x}_{\mathrm{ar}}^\ell = \underline{\mathbf{x}} + [\varphi] \, [\lambda]^{1/2} \, \boldsymbol{\eta}^\ell$ with $\boldsymbol{\eta}^\ell = (\eta_1^\ell, \ldots, \eta_\nu^\ell)$.

Appendix B. Proof of the mean-square convergence of the random sequence $\mathbf{X}^{(\nu_q, \nu_w)}$.

Since $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$, $\underline{\mathbf{x}}_{\mathrm{ar}} = (\underline{\mathbf{q}}_{\mathrm{ar}}, \underline{\mathbf{w}}_{\mathrm{ar}})$, and $\mathbf{X}^{(\nu_q, \nu_w)} = (\mathbf{Q}^{(\nu_q)}, \mathbf{W}^{(\nu_w)})$, we have

$$E\{\|\mathbf{X} - \underline{\mathbf{x}}_{\mathrm{ar}}\|^2\} = E\{\|\mathbf{Q} - \underline{\mathbf{q}}_{\mathrm{ar}}\|^2\} + E\{\|\mathbf{W} - \underline{\mathbf{w}}_{\mathrm{ar}}\|^2\} \, ,$$

that is equal to $\mathrm{tr}\,[C_{\mathbf{Q}}] + \mathrm{tr}\,[C_{\mathbf{W}}]$, in which $\mathbf{X}$, $\mathbf{Q}$, and $\mathbf{W}$ stand for $\mathbf{X}^{(n_q, n_w)}$, $\mathbf{Q}^{(n_q)}$, and $\mathbf{W}^{(n_w)}$. We also have

$$E\{\|\mathbf{X} - \mathbf{X}^{(\nu_q, \nu_w)}\|^2\} = E\{\|\mathbf{Q} - \mathbf{Q}^{(\nu_q)}\|^2\} + E\{\|\mathbf{W} - \mathbf{W}^{(\nu_w)}\|^2\} \, ,$$

that can be rewritten, using Eqs. (P-8) and (P-14), as

$$E\{\|\mathbf{X} - \mathbf{X}^{(\nu_q, \nu_w)}\|^2\} = \mathrm{err}_{\mathbf{Q}}(\nu_q) \, E\{\|\mathbf{Q} - \underline{\mathbf{q}}_{\mathrm{ar}}\|^2\} + \mathrm{err}_{\mathbf{W}}(\nu_w) \, E\{\|\mathbf{W} - \underline{\mathbf{w}}_{\mathrm{ar}}\|^2\} \, .$$

Since $\mathrm{tr}\,[C_{\mathbf{Q}}] > 0$ and $\mathrm{tr}\,[C_{\mathbf{W}}] > 0$, it can then be deduced that

$$\mathrm{err}_{\mathbf{X}}(\nu_q, \nu_w) = \frac{\mathrm{err}_{\mathbf{Q}}(\nu_q)}{1 + \mathrm{tr}\,[C_{\mathbf{W}}]/\mathrm{tr}\,[C_{\mathbf{Q}}]} + \frac{\mathrm{err}_{\mathbf{W}}(\nu_w)}{1 + \mathrm{tr}\,[C_{\mathbf{Q}}]/\mathrm{tr}\,[C_{\mathbf{W}}]} \, .$$

Defining $\zeta = \max\{(1 + \mathrm{tr}\,[C_{\mathbf{W}}]/\mathrm{tr}\,[C_{\mathbf{Q}}])^{-1}, (1 + \mathrm{tr}\,[C_{\mathbf{Q}}]/\mathrm{tr}\,[C_{\mathbf{W}}])^{-1}\} > 0$ yields $\mathrm{err}_{\mathbf{X}}(\nu_q, \nu_w) \leq \zeta \, (\mathrm{err}_{\mathbf{Q}}(\nu_q) + \mathrm{err}_{\mathbf{W}}(\nu_w))$. Since $\zeta < 1$, we then obtain

$$\mathrm{err}_{\mathbf{X}}(\nu_q, \nu_w) \leq \mathrm{err}_{\mathbf{Q}}(\nu_q) + \mathrm{err}_{\mathbf{W}}(\nu_w) \, . \tag{4}$$

Appendix C. Construction of the regularization model of covariance matrix $[C_{\widehat{\mathbf{X}}}]$ defined by Eq. (P-24)

Let $[\widehat{C}_\varepsilon]$ be a regularization model in $\mathbb{M}_\nu^+$ of $[C_{\widehat{\mathbf{X}}}]$ such that its condition number is of order 1. Therefore, $[\widehat{C}_\varepsilon]^{-1}$ is in $\mathbb{M}_\nu^+$ and its condition number is also of order 1. The proposed regularization is constructed as follows. Let us consider the following classical spectral representation of matrix $[C_{\widehat{\mathbf{X}}}]$,

$$[C_{\widehat{\mathbf{X}}}] = [\Phi] \, [\lambda] \, [\Phi]^T \, , \tag{5}$$

in which the real eigenvalues are in decreasing order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_\nu \geq 0$ and where $[\Phi]$ is the matrix in $\mathbb{M}_\nu$ of the corresponding eigenvectors. Due to Eqs. (P-24) and (P-25), it is proven in Appendix D that these eigenvalues are such that

$$0 \leq \lambda_j \leq 2 \quad , \quad j \in \{1, \ldots, \nu\} \, . \tag{6}$$

If $[C_{qw}]$ was the zero matrix in $\mathbb{M}_{\nu_q, \nu_w}$, then matrix $[C_{\widehat{\mathbf{X}}}]$ would be the identity matrix and therefore, all the eigenvalues would be equal to 1. Since $[C_{qw}]$ is not the zero matrix and taking into account Eq. (6), there exists and we define (by construction of the regularization model) the integer ν_1, such that,

$$\lambda_{\nu_1} \geq 1 \quad , \quad \lambda_{\nu_1 + 1} < 1 \quad , \quad \nu_1 + 1 \leq \nu \, . \tag{7}$$

The regularization, $[\widehat{C}_\varepsilon]$ of $[C_{\widehat{\mathbf{X}}}]$ is defined by

$$[\widehat{C}_\varepsilon] = [\Phi] \, [\Lambda_\varepsilon] \, [\Phi]^T \, , \tag{8}$$

in which the diagonal matrix $[\Lambda_\varepsilon]$ is such that

$$[\Lambda_\varepsilon]_{jj} = \lambda_j, \; 1 \leq j \leq \nu_1 \, ; \; [\Lambda_\varepsilon]_{jj} = \varepsilon^2 \, \lambda_{\nu_1}, \; \nu_1 + 1 \leq j \leq \nu, \tag{9}$$

in which $\varepsilon \in [\varepsilon_{\min}, 1[$ were $\varepsilon_{\min} > 0$ is a hyperparameter that controls the regularization and whose value will be of close to 0.5. The methodology for choosing the value of ε will be presented in Section P-10. The following properties can then easily be deduced:

$$[\widehat{C}_\varepsilon] \in \mathbb{M}_\nu^+ \quad , \quad [\widehat{C}_\varepsilon]^{-1} = [\Phi] \, [\Lambda_\varepsilon]^{-1} \, [\Phi]^T \in \mathbb{M}_\nu^+ \, . \tag{10}$$

The condition numbers of $[\widehat{C}_\varepsilon]$ and $[\widehat{C}_\varepsilon]^{-1}$ are thus equal to $\lambda_1/(\varepsilon^2 \lambda_{\nu_1})$, and satisfy the following equation,

$$\mathrm{cond}([\widehat{C}_\varepsilon]) = \mathrm{cond}([\widehat{C}_\varepsilon]^{-1}) \leq \frac{2}{\varepsilon^2} \, .$$

For ε close to 0.5, the condition number is less that 8. We next make four observations relevant to the proposed regularization.

(i) Remark concerning the Tikhonov regularization The Tikhonov regularization $[\widehat{C}_\gamma]$ of $[C_{\widehat{\mathbf{X}}}]$ with respect to its inverse (see for instance [22]), would be such that $\widehat{\mathbf{y}}_\gamma = [\widehat{C}_\gamma]^{-1} \, \widehat{\mathbf{x}}$, in which $\widehat{\mathbf{y}}_\gamma$ is the unique solution in $\mathbb{R}^\nu$ of the optimization problem,

$$\widehat{\mathbf{y}}_\gamma = \min_{\widehat{\mathbf{y}} \in \mathbb{R}^\nu} \{\|[C_{\widehat{\mathbf{X}}}] \, \widehat{\mathbf{y}} - \widehat{\mathbf{x}}\|^2 + \gamma^2 \|\widehat{\mathbf{x}}\|^2\} \, , \tag{11}$$

for any given $\widehat{\mathbf{x}}$ in $\mathbb{R}^\nu$, where $\gamma > 0$ is the regularization parameter. The unique solution is such that $([C_{\widehat{\mathbf{X}}}]^2 + \gamma^2 \, [I_\nu]) \, \widehat{\mathbf{y}}_\gamma = [C_{\widehat{\mathbf{X}}}] \, \widehat{\mathbf{x}}$, which yields $[\widehat{C}_\gamma]^{-1} = ([C_{\widehat{\mathbf{X}}}]^2 + \gamma^2 \, [I_\nu])^{-1} \, [C_{\widehat{\mathbf{X}}}]$. Therefore, for $j = 1, \ldots, \nu$, the eigenvalues of $[\widehat{C}_\gamma]^{-1}$ are $\lambda_j/(\lambda_j^2 + \gamma^2)$ while those of $[\widehat{C}_\gamma]$ are $\lambda_j + \gamma^2/\lambda_j$. This regularization shows that $[\widehat{C}_\gamma]^{-1}$ is not positive definite if the rank of $[C_{\widehat{\mathbf{X}}}]$ is less that ν, and that, if the rank of $[C_{\widehat{\mathbf{X}}}]$ were ν, then the condition number $\mathrm{cond}([\widehat{C}_\gamma]^{-1})$ of $[\widehat{C}_\gamma]^{-1}$, which is equal to $\{\lambda_1/\lambda_\nu\} \times \{(\lambda_\nu + \gamma^2)/\lambda_1 + \gamma^2\}$, goes to infinity as λ_ν goes to zero, which is antinomic with the property sought. Consequently, the regularization that is constructed with Eq. (11) cannot be used.

(ii) Interpretation of the proposed regularization model as a Tikhonov regularization Let us assume that the eigenvalues $\lambda_1, \ldots, \lambda_\nu$ of $[C_{\widehat{\mathbf{X}}}]$ are such that, for $\varepsilon \in [\varepsilon_{\min}, 1[$ with $\varepsilon_{\min} > 0$, and for ν_1 defined by Eq. (7), we have $\varepsilon^2 \lambda_j \lambda_{\nu_1} - \lambda_j^2 \geq 0$ for all $j \geq \nu_1 + 1$. Let $\gamma_1, \ldots, \gamma_\nu$ be the real numbers defined by $\gamma_j = 0$ for $j = 1, \ldots, \nu_1$ and by $\gamma_j = (\varepsilon^2 \lambda_j \lambda_{\nu_1} - \lambda_j^2)^{1/2}$ for $j = \nu_1 + 1, \ldots, \nu$. Let $[\Gamma]$ be the matrix in $\mathbb{M}_\nu^{+0}$ defined by $[\Gamma] = [\Phi]\,[\gamma]\,[\Phi]^T$ in which $[\gamma]$ is the diagonal matrix such that $[\gamma]_{jk} = \gamma_j \delta_{jk}$. It can then be seen that the regularization $[\widehat{C}_\varepsilon]$ defined by Eq. (8) is such that, for all $\widehat{\mathbf{x}}$ in $\mathbb{R}^\nu$, $\widehat{\mathbf{y}}_\varepsilon = [\widehat{C}_\varepsilon]\,\widehat{\mathbf{x}}$ in which $\widehat{\mathbf{y}}_\varepsilon$ is given by

$$\widehat{\mathbf{y}}_\varepsilon = \min_{\widehat{\mathbf{y}} \in \mathbb{R}^\nu} \{ \| \, [C_{\widehat{\mathbf{X}}}]\,\widehat{\mathbf{y}} - \widehat{\mathbf{x}} \, \|^2 + \| \, [\Gamma]\,\widehat{\mathbf{x}} \, \|^2 \} \, .$$

(iii) Choice of the value of hyperparameter ε that controls the regularization The choice of the value of hyperparameter ε is presented in Section P-10.

(iv) Other remarks concerning the possible regularization models Other types of regularization models could *a priori* be used.

(1) If the rank of $[C_{\widehat{\mathbf{X}}}]$ is less than ν, the generalized inverse (or pseudo-inverse) of $[C_{\widehat{\mathbf{X}}}]$ (see for instance Chapter 6, pp. 163-226 in [10]) could be used. Such an approach would lead us to introduce a new parameterization of a submanifold for $\widehat{\mathbf{X}}$ whose dimension would be the rank of $[C_{\widehat{\mathbf{X}}}]$. The estimation $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}$ of pdf $p_{\widehat{\mathbf{X}}}$ could then be constructed by using, for instance, the approach proposed in [8]. Nevertheless, not only the construction of the pdf $p_{\widehat{\mathbf{W}}}^{(\nu_{\text{ar}})}$ of $\mathbf{W}$ would require an integration on the submanifold, which would induce difficulties, but above all, the "separation" of the representations of $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{W}}$ would be lost, and such a "separation" is necessary for our purpose. Moreover, this approach would be equivalent to doing a PCA of random vector $\mathbf{X}$ instead of two PCAs, one on $\mathbf{Q}$ and the other one on $\mathbf{W}$, a method that cannot be done as we have explained in Section P-6.

(2) A more classical regularization of $[C_{\widehat{\mathbf{X}}}]$ would consist in taking $[\widehat{C}_\eta] = [C_{\widehat{\mathbf{X}}}] + [C_\eta]$ with $[C_\eta]$ a covariance matrix in $\mathbb{M}_\nu^+$. A choice could be $[C_\eta] = \eta^2\,[I_\nu]$. Such a model corresponds to the introduction of an additional Gaussian noise represented by the random vector $\widehat{\mathbf{B}}_\eta$ independent of $\widehat{\mathbf{X}}$, such that $\widehat{\mathbf{X}}_\eta = \widehat{\mathbf{X}} + \widehat{\mathbf{B}}_\eta$ (taking into account the Gaussian kernel-density estimation used for the estimate $p_{\widehat{\mathbf{W}}}^{(\nu_{\text{ar}})}$ of $p_{\widehat{\mathbf{X}}}$ defined by Eq. (P-26). The numerical evaluation of such a regularization has been used for the applications presented in Sections 12 and P-13, and has demonstrated a lack of robustness when used with the MCMC generator of $p_{\widehat{\mathbf{W}}}^{\text{post}}$.

(3) A regularization of the probability measure $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}(\widehat{\mathbf{x}})\, d\widehat{\mathbf{x}}$ could also be constructed using the Rao metric between two probability distributions [11], which involves the Fisher information matrix. Nevertheless, the algebraic expression of $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}$ given by Eq. (P-26) is not easy due to the presence of the summation over the ν_{ar} realizations. In a similar framework, another way would have been to use the Riemann metric related to the geodesic distance on the manifold related to the positive-definite matrices [1], which is particularly well adapted to the Gaussian case as proposed, for instance in [21], but which induces difficulties for the non-Gaussian probability measure $p_{\widehat{\mathbf{X}}}^{(\nu_{\text{ar}})}(\widehat{\mathbf{x}})\, d\widehat{\mathbf{x}}$.

Appendix D. Proof of the range of the values of covariance matrix $[C_{\widehat{X}}]$ defined by Eq. (P-24)

Since matrix $[C_{\widehat{X}}]$ is positive or positive definite, we have $<[C_{\widehat{X}}]\,\widehat{\mathbf{x}}, \widehat{\mathbf{x}}> \geq 0$ for all $\widehat{\mathbf{x}} = (\widehat{\mathbf{q}}, \widehat{\mathbf{w}})$ in $\mathbb{R}^\nu = \mathbb{R}^{\nu_q} \times \mathbb{R}^{\nu_w}$. Using Eq. (P-24) yields

$$2 <[C_{qw}]^T\,\widehat{\mathbf{q}}\,, \widehat{\mathbf{w}}> + \|\widehat{\mathbf{q}}\|^2 + \|\widehat{\mathbf{w}}\|^2 \geq 0\,. \tag{12}$$

For all $\widehat{\mathbf{q}}$ in $\mathbb{R}^{\nu_q}$, we can choose $\widehat{\mathbf{w}} = -[C_{qw}]^T\,\widehat{\mathbf{q}}$ in Eq. (12), which yields

$$\|[C_{qw}]^T\,\widehat{\mathbf{q}}\|^2 \leq \|\widehat{\mathbf{q}}\|^2\,,$$

which can be rewritten as

$$<[C_{qw}]\,[C_{qw}]^T\,\widehat{\mathbf{q}}\,, \widehat{\mathbf{q}}> \;\leq\; \|\widehat{\mathbf{q}}\|^2 \quad , \quad \forall\, \widehat{\mathbf{q}} \in \mathbb{R}^{\nu_q}\,. \tag{13}$$

Let $\Lambda_1 \geq \ldots \geq \Lambda_{\nu_q} \geq 0$ be the eigenvalues of the positive matrix $[C_{qw}]\,[C_{qw}]^T$. Equation (13) shows that

$$1 \geq \Lambda_1 \geq \ldots \geq \Lambda_{\nu_q} \geq 0\,. \tag{14}$$

Let us consider the eigenvalue problem $[C_{\widehat{\mathbf{X}}}]\,\widehat{\varphi} = \lambda\,\widehat{\varphi}$ (see Eq. (5) in which the columns of matrix $[\Phi]$ are the eigenvectors $\widehat{\varphi}$). Using the block decomposition defined by Eq. (P-24) and $\widehat{\varphi} = (\widehat{\varphi}_q, \widehat{\varphi}_w)$ yield

$$\widehat{\varphi}_q + [C_{qw}]\,\widehat{\varphi}_w = \lambda\,\widehat{\varphi}_q\,, \tag{15}$$

$$[C_{qw}]^T\,\widehat{\varphi}_q + \widehat{\varphi}_w = \lambda\,\widehat{\varphi}_w\,. \tag{16}$$

Eliminating $\widehat{\varphi}_w$ between Eqs. (15) and (16) yields

$$[C_{qw}]\,[C_{qw}]^T\,\widehat{\varphi}_q = (1-\lambda)^2\,\widehat{\varphi}_q\,.$$

Consequently, $(1-\lambda)^2$ appears as the eigenvalue Λ of $[C_{qw}]\,[C_{qw}]^T$. Taking into account Eq. (14), it can be deduced that $-1 \leq 1-\lambda \leq 1$, which proves that any eigenvalue λ of matrix $[C_{\widehat{X}}]$ is such that $0 \leq \lambda \leq 2$.

Appendix E. Proof of Eq. (P-38) for the consistency of the estimator defined by Eq. (P-37) corresponding to the estimation defined by Eq. (P-35)

The proof is inspired of [9], is slightly different, is adapted to the Gaussian kernel-density, and the upper bound defined by Eq. (P-38) is not the same. The Silverman bandwidth $s_{\rm ar}$ is defined by Eq. (P-27) and $\widehat{\mathbf{x}}$ is a point fixed in $\mathbb{R}^\nu$. Let $\widehat{\mathbf{x}} \mapsto \kappa(\widehat{\mathbf{x}})$ be the Gaussian pdf, centered, with invertible covariance matrix $[\widehat{C}_\varepsilon]$ defined by Eq. (8), such that $[G] = [\widehat{C}_\varepsilon]^{-1} \in \mathbb{M}_\nu^+$, and let $\widehat{\mathbf{x}} \mapsto \kappa_{\nu_{\rm ar}}(\widehat{\mathbf{x}})$ be the function on $\mathbb{R}^\nu$, such that,

$$\kappa(\widehat{\mathbf{x}}) = \frac{\sqrt{\det[G]}}{(2\pi)^{\nu/2}}\, \exp\{-\frac{1}{2}\, <[G]\,\widehat{\mathbf{x}}\,, \widehat{\mathbf{x}}>\}\,,$$

$$\kappa_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) = \frac{1}{s_{\rm ar}^\nu}\, \kappa(\frac{\widehat{\mathbf{x}}}{s_{\rm ar}})\,. \tag{17}$$

Using the change of variable $\widehat{\mathbf{x}} = [\Phi]\,\widehat{\boldsymbol{\eta}}$ with $[G] = [\Phi]\,[\Lambda_\varepsilon]^{-1}\,[\Phi]^T$ (see Eq. (10)) and since $s_{\rm ar} \to 0$ when $\nu_{\rm ar} \to +\infty$, it can be seen that we have the following limit in the space of measures on $\mathbb{R}^\nu$,

$$\lim_{\nu_{\rm ar}\to+\infty} \kappa_{\nu_{\rm ar}}(\widehat{\mathbf{x}})\, d\widehat{\mathbf{x}} = \delta_0(\widehat{\mathbf{x}})\,, \tag{18}$$

in which $d\widehat{\mathbf{x}}$ is the Lebesgue measure on $\mathbb{R}^\nu$ and $\delta_0(\widehat{\mathbf{x}})$ is the Dirac measure on $\mathbb{R}^\nu$ at point $\widehat{\mathbf{x}} = \mathbf{0}$.

(i) Sequence of estimators of $p_{\widehat{\mathbf{X}}}$. Let $\widehat{\mathbf{X}}^1, \ldots, \widehat{\mathbf{X}}^{\nu_{\rm ar}}$ be $\nu_{\rm ar}$ independent copies of random variable $\widehat{\mathbf{X}}$ whose pdf is $p_{\widehat{\mathbf{X}}}$. Therefore, $\widehat{\mathbf{x}}^\ell$ is a realization of $\widehat{\mathbf{X}}^\ell$. For $\widehat{\mathbf{x}}$ fixed in $\mathbb{R}^\nu$, the sequence of estimators of $p_{\widehat{\mathbf{X}}}(\widehat{\mathbf{x}})$, whose an estimation is $p_{\widehat{\mathbf{X}}}^{(\nu_{\rm ar})}(\widehat{\mathbf{x}})$ defined by Eq. (P-35), is the sequence $\{P_{\nu_{\rm ar}}(\widehat{\mathbf{x}})\}_{\nu_{\rm ar}}$ of positive-valued random variables defined by

$$P_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) = \frac{1}{\nu_{\rm ar}} \sum_{\ell=1}^{\nu_{\rm ar}} \kappa_{\nu_{\rm ar}}(\widehat{\mathbf{X}}^\ell - \widehat{\mathbf{x}})\,.$$

(ii) Mean value $\underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}})$ *of* $P_{\nu_{\rm ar}}(\widehat{\mathbf{x}})$. The mean value of random variable $P_{\nu_{\rm ar}}(\widehat{\mathbf{x}})$ is written as $E\{P_{\nu_{\rm ar}}(\widehat{\mathbf{x}})\} = \frac{1}{\nu_{\rm ar}} \sum_{\ell=1}^{\nu_{\rm ar}} E\{\kappa_{\nu_{\rm ar}}(\widehat{\mathbf{X}}^\ell - \widehat{\mathbf{x}})\}$, which yields

$$\underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) = \int_{\mathbb{R}^\nu} \kappa_{\nu_{\rm ar}}(\widetilde{\mathbf{x}} - \widehat{\mathbf{x}})\, p_{\widehat{\mathbf{X}}}(\widetilde{\mathbf{x}})\, d\widetilde{\mathbf{x}}\,. \tag{19}$$

Assuming that $p_{\widehat{\mathbf{X}}}$ is a continuous function in $\widehat{\mathbf{x}} \in \mathbb{R}^\nu$, using Eq. (18) yields

$$\lim_{\nu_{\rm ar}\to+\infty} \underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) = p_{\widehat{\mathbf{X}}}(\widehat{\mathbf{x}})\,. \tag{20}$$

(iii) Variance of $P_{\nu_{\rm ar}}(\widehat{\mathbf{x}})$. Since the random variables $\widehat{\mathbf{X}}^1, \ldots, \widehat{\mathbf{X}}^{\nu_{\rm ar}}$ are independent copies of $\widehat{\mathbf{X}}$, and using Eq. (17), the variance of $P_{\nu_{\rm ar}}(\widehat{\mathbf{x}})$ is such that

$$\begin{aligned} E\{(P_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) - \underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}}))^2\} &= \frac{1}{\nu_{\rm ar}} E\{(\kappa_{\nu_{\rm ar}}(\widehat{\mathbf{X}} - \widehat{\mathbf{x}}))^2\} - \frac{1}{\nu_{\rm ar}} (\underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}}))^2 \\ &\le \frac{1}{\nu_{\rm ar}} E\{(\kappa_{\nu_{\rm ar}}(\widehat{\mathbf{X}} - \widehat{\mathbf{x}}))^2\} \\ &= \frac{1}{\nu_{\rm ar}} \int_{\mathbb{R}^\nu} (\kappa_{\nu_{\rm ar}}(\widetilde{\mathbf{x}} - \widehat{\mathbf{x}}))^2\, p_{\widehat{\mathbf{X}}}(\widetilde{\mathbf{x}})\, d\widetilde{\mathbf{x}}\,. \end{aligned}$$

Since $\forall\, \widehat{\mathbf{x}},\ \sup_{\widetilde{\mathbf{x}}} \kappa_{\nu_{\rm ar}}(\widetilde{\mathbf{x}} - \widehat{\mathbf{x}}) = \frac{1}{s_{\rm ar}^\nu} \frac{\sqrt{\det[G]}}{(2\pi)^{\nu/2}}$ and using Eqs. (17) and (19), we have

$$E\{(P_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) - \underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}}))^2\} \le \frac{1}{\nu_{\rm ar} s_{\rm ar}^\nu} \frac{\sqrt{\det[G]}}{(2\pi)^{\nu/2}}\, \underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}})\,. \tag{21}$$

Substituting $s_{\rm ar}$ given by Eq. (P-27) into the right-hand side of Eq. (21) yields

$$\begin{aligned} &E\{(P_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) - \underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}}))^2\} \\ &\quad \le \left\{\frac{1}{\nu_{\rm ar}}\right\}^{4/(\nu+4)} \left\{\frac{\nu+2}{4}\right\}^{\nu/(\nu+4)} \frac{\sqrt{\det[G]}}{(2\pi)^{\nu/2}}\, \underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}})\,. \end{aligned} \tag{22}$$

(iv) Properties of the sequence of estimators. It can be seen that

$$\begin{aligned} &E\{(P_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) - p_{\widehat{\mathbf{X}}}(\widehat{\mathbf{x}}))^2\} \\ &= E\{(P_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) - \underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}}))^2\} + (\underline{P}_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) - p_{\widehat{\mathbf{X}}}(\widehat{\mathbf{x}}))^2\,. \end{aligned} \tag{23}$$

Using Eqs. (20), (22), and (23) for $\nu_{\rm ar} \to +\infty$, it can be seen that the estimator $P_{\nu_{\rm ar}}(\widehat{\mathbf{x}})$ is asymptotically unbiased and is consistent because

$$\lim_{\nu_{\rm ar}\to+\infty} E\{(P_{\nu_{\rm ar}}(\widehat{\mathbf{x}}) - p_{\widehat{\mathbf{X}}}(\widehat{\mathbf{x}}))^2\} = 0\,. \tag{24}$$

The mean-square convergence corresponding to Eq. (24) implies the convergence in probability.

Appendix F. Expression of the mapping $[L]$ defined by Eq. (P-53) adapted to computation

An explicit algebraic expression is constructed for the mapping $[u] \mapsto [L([u])]$ defined by Eq. (P-53), using Eqs. (P-50) for p, Eqs. (P-41) and (P-42) for $p_{\widehat{\mathbf{Q}},\widehat{\mathbf{W}}}$, and Eqs. (P-44) and (P-45) for $p_{\widehat{\mathbf{W}}}$. These equations show the presence of a summation of exponential terms (summation over the number $\nu_{\rm ar}$ of realizations $\widehat{\mathbf{q}}^\ell$ and $\widehat{\mathbf{w}}^\ell$ of $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{W}}$). Consequently, an adapted algebraic representation must be developed in order to minimize the numerical cost for each evaluation of $[L([u])]$ and to avoid numerical noise, overflow, and underflow during the computation. Several expressions have been developed and evaluated. We present the most efficient one

with respect to the above criteria. For $k = 1, \ldots, \nu_w$, for $j = 1, \ldots, N_s$, and for $[u] = [\mathbf{u}^1, \ldots, \mathbf{u}^{N_s}]$ in $\mathbb{M}_{\nu_w, N_s}$,

$$[L([u])]_{kj} = \frac{1}{s_{\text{ar}}^2}\{-[G_{0w}]\,\mathbf{u}^j - \mathbf{b}^{\text{exp}} + (1-n_r)\,\mathbf{a}_0(\mathbf{u}^j) + \sum_{r=1}^{n_r} \mathbf{a}_1^r(\mathbf{u}^j)\}_k\,, \quad (25)$$

where $\mathbf{a}_0(\mathbf{u}^j) = (a_{0,1}(\mathbf{u}^j), \ldots, a_{0,\nu_w}(\mathbf{u}^j))$ and $\mathbf{a}_1^r(\mathbf{u}^j) = (a_{1,1}^r(\mathbf{u}^j), \ldots, a_{1,\nu_w}^r(\mathbf{u}^j))$ are vectors in $\mathbb{R}^{\nu_w}$ such that

$$\mathbf{a}_0(\mathbf{u}^j) = (\sum_{\ell=1}^{\nu_{\text{ar}}} \widetilde{\mathbf{w}}_0^\ell\, \zeta_0^\ell(\mathbf{u}^j)) \,/\, \sum_{\ell=1}^{\nu_{\text{ar}}} \zeta_0^\ell(\mathbf{u}^j)\,, \quad (26)$$

and for $r \in \{1, \ldots, n_r\}$,

$$\mathbf{a}_1^r(\mathbf{u}^j) = (\sum_{\ell=1}^{\nu_{\text{ar}}} \widetilde{\mathbf{w}}_1^\ell\, \zeta_1^{r\ell}(\mathbf{u}^j)) \,/\, \sum_{\ell=1}^{\nu_{\text{ar}}} \zeta_1^{r\ell}(\mathbf{u}^j)\,. \quad (27)$$

In Eq. (25), the symmetric $(\nu_w \times \nu_w)$ real matrix $[G_{0w}]$ is given by

$$[G_{0w}] = (1-n_r)\,[G_0] + n_r\,[G_w]\,. \quad (28)$$

From Eqs. (P-46) and (P-51), it can be deduced that $[G_{0w}] \in \mathbb{M}_{\nu_w}^+$ for $n_r \geq 2$. The vector $\mathbf{b}^{\text{exp}} \in \mathbb{R}^{\nu_w}$ is given by

$$\mathbf{b}^{\text{exp}} = [G_{qw}]^T \sum_{r=1}^{n_r} \widehat{\mathbf{q}}^{\text{exp},r}\,.$$

In Eq. (26), for all $\ell \in \{1, \ldots, \nu_{\text{ar}}\}$, we have

$$\widetilde{\mathbf{w}}_0^\ell = [G_0]\,\widehat{\mathbf{w}}^\ell \in \mathbb{R}^{\nu_w}\,,$$

$$\zeta_0^\ell(\mathbf{u}^j) = \exp\{-\frac{1}{2s_{\text{ar}}^2}\|\,[\mathcal{L}_0]\,(\mathbf{u}^j - \widehat{\mathbf{w}}^\ell)\|^2\} \in \mathbb{R}^+\,, \quad (29)$$

in which the upper triangular $(\nu_w \times \nu_w)$ real matrix $[\mathcal{L}_0]$ follows from the Cholesky factorization, $[G_0] = [\mathcal{L}_0]^T\,[\mathcal{L}_0]$. In Eq. (27), for all $\ell \in \{1, \ldots, \nu_{\text{ar}}\}$, we have

$$\widetilde{\mathbf{w}}_1^\ell = [G_w]\,\widehat{\mathbf{w}}^\ell + [G_{qw}]^T\,\widehat{\mathbf{q}}^\ell \in \mathbb{R}^{n_w}\,,$$

and for $r \in \{1, \ldots, n_r\}$,

$$\zeta_1^{r\ell}(\mathbf{u}^j) = \exp\{-\frac{1}{2s_{\text{ar}}^2}(p_0^{r\ell} + p_1^{r\ell}(\mathbf{u}^j))\} \in \mathbb{R}^+\,, \quad (30)$$

in which the positive real number $p_0^{r\ell}$ is expressed as

$$p_0^{r\ell} = \|\,[\mathcal{L}_q]\,(\widehat{\mathbf{q}}^{\text{exp},r} - \widehat{\mathbf{q}}^\ell)\|^2\,,$$

with $[\mathcal{L}_q]$ the real upper triangular $(\nu_q \times \nu_q)$ Cholesky factor, $[G_q] = [\mathcal{L}_q]^T\,[\mathcal{L}_q]$, and where

$$p_1^{r\ell}(\mathbf{u}^j) = \|\,[\mathcal{L}_w]\,(\mathbf{u}^j - \widehat{\mathbf{w}}^\ell)\|^2 + 2 < [G_{qw}]^T(\widehat{\mathbf{q}}^{\text{exp},r} - \widehat{\mathbf{q}}^\ell\}\,, \mathbf{u}^j - \widehat{\mathbf{w}}^\ell >\,,$$

where the upper triangular $(\nu_w \times \nu_w)$ real matrix $[\mathcal{L}_w]$ is obtained from the Cholesky factorization, $[G_w] = [\mathcal{L}_w]^T\,[\mathcal{L}_w]$. The numerical experiments that have been carried out have shown that, for computation of $\zeta_1^{r\ell}(\mathbf{u}^j)$ defined by Eq. (30), the term in the exponential must be computed before exponentiation in order to avoid underflow and numerical noise.

Appendix G. Verifying that the linearized ISDE is well adapted for stochastic process $\{([\mathbf{S}(t)], [\mathbf{R}(t)]), t \in \mathbb{R}^+\}$

Using Eqs. (P-58) and (P-59), the linearization $[s] \mapsto [\widetilde{L}^{\text{lin}}([s])]$ of function $[s] \mapsto [\widetilde{L}([s])]$ defined by Eq. (P-68) is such that $[\widetilde{L}^{\text{lin}}([s])] = -[s]$. From Eqs. (P-64) and (P-65), it can be deduced that the linearized ISDE is written as

$$d[\mathbf{S}^{\text{lin}}(t)] = [\mathbf{R}^{\text{lin}}(t)]\,dt\,,$$

$$d[\mathbf{R}^{\text{lin}}(t)] = -[\mathbf{S}^{\text{lin}}(t)])]\,dt - \frac{1}{2}\,f_0^{\text{post}}[\mathbf{R}^{\text{lin}}(t)]\,dt + \sqrt{f_0^{\text{post}}}\,d[\mathbf{W}^{\text{wien}}(t)]\,.$$

Let us write $[\mathbf{S}^{\text{lin}}(t)] = [\mathbf{S}^{\text{lin},1}(t) \ldots \mathbf{S}^{\text{lin},N_s}(t)]$ whose columns are statistically dependent for $t > 0$ due to the coupling by the initial conditions defined by Eqs. (P-66) and (P-67). Nevertheless, for the asymptotic solution ($t \to +\infty$), denoted as $\{([\mathbf{S}^{\text{lin}}(t_{\text{st}})],\, [\mathbf{R}^{\text{lin}}(t_{\text{st}})]), t_{\text{st}} \in \mathbb{R}^+\}$, these are statistically independent and it is known (see for instance Page 241 of [12]) that each column $\{\mathbf{S}^{\text{lin},j}(t_{\text{st}}), t_{\text{st}} \in \mathbb{R}^+\}$ of $\{[\mathbf{S}^{\text{lin}}(t_{\text{st}})], t_{\text{st}} \in \mathbb{R}^+\}$ is a Gaussian, stationary, centered stochastic process whose covariance matrix $[C^{\text{lin}}] = E\{\mathbf{S}^{\text{lin},j}(t_{\text{st}})\,\mathbf{S}^{\text{lin},j}(t_{\text{st}})^T\}$ is independent of j and such that $[C^{\text{lin}}] = [I_{\nu_w}]$. This result shows that the nonlinear ISDE defined by Eqs. (P-64) to (P-68) is well adapted to the covariance matrix of the asymptotic stochastic process $\{[\mathbf{S}(t_{\text{st}})], t_{\text{st}} \in \mathbb{R}^+\}$ and therefore, to the covariance of $[\mathbf{W}^{\text{post}}]$ via the transformation defined by Eqs. (P-60) and (P-61).

Appendix H. Construction of the diffusion-maps basis for the posterior model

The construction, based on [2], is the one presented in [16] and is summarized in Appendix A.2, using the N_s independent realizations $\{\mathbf{s}^j, j = 1, \ldots, N_s\}$ defined by Eq. (P-72). Let $[\mathbb{P}_s] = [\mathbb{b}_s]^{-1}\,[K_s]$ be the matrix in $\mathbb{M}_{N_s}$ such that, for all i, j and j' in $\{1, \ldots, N_s\}$, $[K_s]_{jj'} = \exp(-\frac{1}{4\,\varepsilon_{\text{diff}}^{\text{post}}}\|\mathbf{s}^j - \mathbf{s}^{j'}\|^2)$ and $[\mathbb{b}_s]_{ij} = \delta_{ij}\,\sum_{j'=1}^{N_s}[K_s]_{jj'}$ depending on a positive parameter $\varepsilon_{\text{diff}}^{\text{post}}$ whose value depends on dataset $\{\mathbf{s}^j, j = 1, \ldots, N_s\}$. Therefore, $[\mathbb{P}_s]$ is a transition matrix of a Markov chain. For $1 < m_{\text{post}} \leq N_s$, let $\mathbf{g}_s^1, \ldots, \mathbf{g}_s^{m_{\text{post}}}$ be the right eigenvectors in $\mathbb{R}^{N_s}$, of the eigenvalue problem $[\mathbb{P}_s]\,\mathbf{g}_s^\alpha = \Lambda_{s,\alpha}\,\mathbf{g}_s^\alpha$ with the normalization condition $[g_s]^T\,[\mathbb{b}_s]\,[g_s] = [I_{m_{\text{post}}}]$ and where the associated $m_{\text{post}} \leq N_s$ positive eigenvalues are such that $1 = \Lambda_{s,1} > \Lambda_{s,2} > \ldots > \Lambda_{s,m_{\text{post}}}$. The diffusion-maps basis is represented by the matrix $[g_s] = [\mathbf{g}_s^1 \ldots \mathbf{g}_s^{m_{\text{post}}}] \in \mathbb{M}_{N_s, m_{\text{post}}}$. The eigenvector $\mathbf{g}_s^1$ associated with the largest eigenvalue $\Lambda_{s,1} = 1$ is a constant vector that has to be kept because stochastic process $[\mathbf{S}]$ is not centered (it is $[\mathbf{S}^{\text{lin}}]$ that is a centered stochastic process). For $m_{\text{post}} = N_s$, the diffusion-maps basis is an algebraic basis of $\mathbb{R}^{N_s}$. The right-eigenvalue problem of the nonsymmetric matrix $[\mathbb{P}_s]$ can be performed solving the eigenvalue

problem $[\mathbb{b}_s]^{-1/2}\,[K_s]\,[\mathbb{b}_s]^{-1/2}\,\Phi_s^\alpha = \Lambda_{s,\alpha}\,\Phi_s^\alpha$ related to a positive-definite symmetric real matrix with the normalization $\|\Phi_s^\alpha\| = 1$. Therefore, $\mathbf{g}_s^\alpha$ can be deduced from Φ_s^α by $\mathbf{g}_s^\alpha = [\mathbb{b}_s]^{-1/2}\,\Phi_s^\alpha$. The construction introduces two hyperparameters: the dimension $m_{\text{post}} \leq N_s$ and the smoothing parameter $\varepsilon_{\text{diff}}^{\text{post}} > 0$. The algorithm for estimating the optimal values of $\varepsilon_{\text{diff}}^{\text{post}}$ and m_{post} is detailed in [17]. Most of the time, these optimal values can be calculated using Eqs. (1) and (2) in which $\varepsilon_{\text{diff}}$ has to be replaced by $\varepsilon_{\text{diff}}^{\text{post}}$.

Appendix I. Störmer-Verlet scheme for solving the reduced-order ISDE

Let $n_{\text{MC}}^{\text{post}}$ and M_0^{post} be the given integers defined in Section P-9.5. The reduced-order ISDE defined by Eqs. (P-74) to (P-76) is solved for $t \in [0, t_{\text{max}}]$ with $t_{\text{max}} = (\ell_0 + n_{\text{MC}}^{\text{post}}\, M_0^{\text{post}})\,\Delta t$ in which Δt is the sampling step and where ℓ_0 is chosen in order that the solution of Eqs. (P-74) to (P-76) has reached the stationary regime. For $\ell = 0, 1, \ldots, n_{\text{MC}}^{\text{post}} \times M_0^{\text{post}}$, we consider the sampling points $t_\ell = \ell\,\Delta t$ and the following notations: $[\boldsymbol{\mathcal{Z}}_\ell] = [\boldsymbol{\mathcal{Z}}(t_\ell)]$, $[\boldsymbol{\mathcal{Y}}_\ell] = [\boldsymbol{\mathcal{Y}}(t_\ell)]$, and $[\boldsymbol{\mathcal{W}}_\ell^{\text{wien}}] = [\boldsymbol{\mathcal{W}}^{\text{wien}}(t_\ell)]$. The Störmer-Verlet scheme is used for solving the reduced-order ISDE, which is written, for $\ell = 0, 1, \ldots, n_{\text{MC}}^{\text{post}}\, M_0^{\text{post}}$, as

$$[\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}] = [\boldsymbol{\mathcal{Z}}_\ell] + \frac{\Delta t}{2}\,[\boldsymbol{\mathcal{Y}}_\ell]\,,$$

$$[\boldsymbol{\mathcal{Y}}_{\ell+1}] = \frac{1-\beta}{1+\beta}\,[\boldsymbol{\mathcal{Y}}_\ell] + \frac{\Delta t}{1+\beta}\,[\widetilde{\boldsymbol{\mathcal{L}}}_{\ell+\frac{1}{2}}] + \frac{\sqrt{f_0^{\text{post}}}}{1+\beta}\,[\Delta\boldsymbol{\mathcal{W}}_{\ell+1}^{\text{wien}}]\,,$$

$$[\boldsymbol{\mathcal{Z}}_{\ell+1}] = [\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}] + \frac{\Delta t}{2}\,[\boldsymbol{\mathcal{Y}}_{\ell+1}]\,,$$

with the initial condition defined by Eq. (P-76), where $\beta = f_0^{\text{post}}\,\Delta t\,/4$, and where $[\widetilde{\boldsymbol{\mathcal{L}}}_{\ell+\frac{1}{2}}]$ is the $\mathbb{M}_{\nu_w, m_{\text{post}}}$-valued random variable such that

$$[\widetilde{\boldsymbol{\mathcal{L}}}_{\ell+\frac{1}{2}}] = [\widetilde{\mathcal{L}}([\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}])] = [\widetilde{L}([\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}]\,[g_s]^T)]\,[a_s]\,.$$

In the above equation, $[\Delta\boldsymbol{\mathcal{W}}_{\ell+1}^{\text{wien}}] = [\Delta\mathbf{W}_{\ell+1}^{\text{wien}}]\,[a_s]$ is a random variable with values in $\mathbb{M}_{\nu_w, m_{\text{post}}}$, in which the increment $[\Delta\mathbf{W}_{\ell+1}^{\text{wien}}] = [\mathbf{W}_{\ell+1}^{\text{wien}}] - [\mathbf{W}_\ell^{\text{wien}}]$. The increments are statistically independent. For all $k = 1, \ldots, \nu_w$ and for all $j = 1, \ldots, N_s$, the real-valued random variables $\{[\Delta\mathbf{W}_{\ell+1}^{\text{wien}}]_{kj}\}_{kj}$ are independent, Gaussian, second-order, and centered random variables such that

$$E\{[\Delta\mathbf{W}_{\ell+1}^{\text{wien}}]_{kj}[\Delta\mathbf{W}_{\ell+1}^{\text{wien}}]_{k'j'}\} = \Delta t\,\delta_{kk'}\,\delta_{jj'}\,.$$

References

1. Bhatia, R.: Positive definite matrices, vol. 24. Princeton university press (2009)
2. Coifman, R., Lafon, S., Lee, A., Maggioni, M., Nadler, B., Warner, F., Zucker, S.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. PNAS **102**(21), 7426–7431 (2005)
3. Ghanem, R., Soize, C.: Probabilistic nonconvex constrained optimization with fixed number of function evaluations. International Journal for Numerical Methods in Engineering **113**(4), 719–741 (2018). DOI 10.1002/nme.5632
4. Ghanem, R., Soize, C., Thimmisetty, C.: Optimal well-placement using probabilistic learning. Data-Enabled Discovery and Applications **2**(1), 4,1–16 (2018). DOI 10.1007/s41688-017-0014-x
5. Ghanem, R.G., Soize, C., Safta, C., Huan, X., Lacaze, G., Oefelein, J.C., Najm, H.N.: Design optimization of a scramjet under uncertainty using probabilistic learning on manifolds. Journal of Computational Physics **399**, 108930, 1–14 (2019). DOI 10.1016/j.jcp.2019.108930
6. Girolami, M., Calderhead, B.: Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistics Society **73**(2), 123–214 (2011). DOI 10.1111/j.1467-9868.2010.00765.x
7. Neal, R.: MCMC using hamiltonian dynamics. In: S. Brooks, A. Gelman, G. Jones, X.L. Meng (eds.) Handbook of Markov Chain Monte Carlo, chap. 5. Chapman and Hall-CRC Press, Boca Raton (2011). DOI 10.1201/b10905-6
8. Ozakin, A., Gray, A.: Submanifold density estimation. In: Proceedings of Neural Information Processing Systems Conference, NIPS 2009, pp. 1–8. Vancouver, Canada (2009)
9. Parzen, E.: On estimation ofa probability density function and mode. Annals of Mathematical Statistics **33**(3), 1065–1076 (1962)
10. Puntanen, S., Styan, G.: The schur complements in statistics and probability. In: F. Zhang (ed.) The Schur Complement and its Applications, chap. 6, pp. 163–226. Springer, Boston, MA (2005)
11. Rao, C.R.: Appendix 3. on the distance between two populations. Anthropometric Survey of the United Provinces, 1941: A Statistical Study **9**(2-3), 246–247 (1949)
12. Soize, C.: The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions. World Scientific Publishing Co Pte Ltd, Singapore (1994)
13. Soize, C.: Construction of probability distributions in high dimension using the maximum entropy principle. applications to stochastic processes, random fields and random matrices. International Journal for Numerical Methods in Engineering **76**(10), 1583–1611 (2008). DOI 10.1002/nme.2385
14. Soize, C.: Polynomial chaos expansion of a multimodal random vector. SIAM/ASA Journal on Uncertainty Quantification **3**(1), 34–60 (2015). DOI 10.1137/140968495
15. Soize, C., Farhat, C.: Probabilistic learning for modeling and quantifying model-form uncertainties in nonlinear computational mechanics. International Journal for Numerical Methods in Engineering **117**, 819–843 (2019). DOI 10.1002/nme.5980
16. Soize, C., Ghanem, R.: Data-driven probability concentration and sampling on manifold. Journal of Computational Physics **321**, 242–258 (2016). DOI 10.1016/j.jcp.2016.05.044
17. Soize, C., Ghanem, R., Safta, C., Huan, X., Vane, Z., Oefelein, J., Lacaze, G., Najm, H., Tang, Q., Chen, X.: Entropy-based closure for probabilistic learning on manifolds. Journal of Computational Physics (2019). DOI 10.1016/j.jcp.2018.12.029
18. Soize, C., Ghanem, R.G.: Physics-constrained non-gaussian probabilistic learning on manifolds. International Journal for Numerical Methods in Engineering **121**(1), 110–145 (2020). DOI 10.1002/nme.6202
19. Soize, C., Ghanem, R.G.: Probabilistic learning on manifolds. arXiv preprint:2002.12653, 2020, math.ST, 28 February 2020 (2020)

20. Soize, C., Ghanem, R.G., Safta, C., Huan, X., Vane, Z.P., Oefelein, J.C., Lacaze, G., Najm, H.N.: Enhancing model predictability for a scramjet using probabilistic learning on manifolds. AIAA Journal **57**(1), 365–378 (2019). DOI 10.2514/1.J057069
21. Spantini, A., Cui, T., Willcox, K., Tenorio, L., Marzouk, Y.: Goal-oriented optimal approximations of Bayesian linear inverse problems. SIAM Journal on Scientific Computing **39**(5), S167–S196 (2017)
22. Tikhonov, A., Goncharsky, A., Stepanov, V., Yagola, A.G.: Numerical methods for the solution of ill-posed problems (Mathematics and its Applications). Springer, Berlin (1995)