Probabilistic-learning-based stochastic surrogate model from small incomplete datasets for nonlinear dynamical systems

Christian Soize, Roger Ghanem

Probabilistic-learning-based stochastic surrogate model from small incomplete datasets for nonlinear dynamical systems

Christian Soize[a,*], R. Ghanem[b]

[a] *Université Gustave Eiffel, MSME UMR 8208, 5 bd Descartes, 77454 Marne-la-Vallée, France*
[b] *University of Southern California, 210 KAP Hall, Los Angeles, CA 90089, United States*

Abstract

We consider a high-dimensional nonlinear computational model of a dynamical system, parameterized by a vector-valued control parameter, in the presence of uncertainties represented by an uncontrolled parameter modeled by a vector-valued random variable, and possibly with stochastic excitation. The objective is to construct a statistical surrogate model where the input is any deterministic value of the control parameter, and the output is a vector-valued observation of the computational model, which is a random vector whose probability measure is updated using a target dataset. To construct this statistical surrogate model, the stochastic response of the computational model must be built, which is a vector-valued time-discretized stochastic process in high dimension, depending on the control parameter. It is assumed that the computational cost of a single evaluation of the deterministic model is high. For the probabilistic updating, we consider a subset of the components of the observation of the computational model, defined as the "identification observation" of the computational model, for which a small target dataset is available. Therefore, the target dataset is associated with partial observability, corresponding to an incomplete data case. Given a prior probability model of the random control and uncontrolled parameters, a training dataset is constructed, consisting of realizations of the random triplet composed of the stochastic response, the random identification observation, and the random control parameter. Since the computational cost of a single evaluation of the deterministic model is assumed to be large, the training dataset is also of small size. The main challenges in this problem are the high dimensionality, partial observability leading to incomplete data in the target dataset for the identification observation of the computational model (which is not sufficient to identify the computational stochastic responses), and the availability of a small training dataset. To address these challenges, we propose a methodology based on statistical methods for constructing necessary reduced representations, direct probabilistic learning under constraints using probabilistic learning on manifolds (PLoM) constrained by the target dataset, and the use of a weak formulation of the Fourier transform of probability measures. Statistical conditioning is also employed to explore the learned dataset. The constructed predictive statistical surrogate model can be implemented in the context of online computation. We apply this approach to a problem of nonlinear stochastic dynamics in high dimensions within the framework of deformable solids mechanics.

Keywords: Probabilistic learning, realizations as targets, statistical inverse problem, Kullback-Leibler divergence, uncertainty quantification

1. Introduction

1.1. Context of the paper

The development of surrogate models for parameterized large-scale computational models is challenging and an extremely active research topic, resulting in a huge number of publications. The scientific community has proposed several classes of methods, including deterministic representations, probabilistic/statistical-based approaches, and more recently, Machine Learning tools, with or without probabilistic/statistical formulations. It is not possible here,

*Corresponding author: C. Soize, christian.soize@univ-eiffel.fr
Email addresses: christian.soize@univ-eiffel.fr (Christian Soize), ghanem@usc.edu (R. Ghanem)

Published in Computer Methods in Applied Mechanics and Engineering, 418,116498,2024 *October 13, 2023*

nor is it the purpose, to attempt to review this vast field. Instead, we will focus on aspects directly related to the method proposed in this paper. After referencing some general works on surrogate modeling, we will specifically discuss methods for handling incomplete data, approaches for model reduction in the presence of uncertainties, and probabilistic approaches developed for small data. These three components are utilized in the proposed probabilistic-learning-based stochastic surrogate model from small incomplete datasets.

(i) Concerning the general aspects of the surrogate models, their construction can be carried out using a parametric and nonparametric approaches in a deterministic or a probabilistic framework (see for instance [1, 2, 3]).

- Typically the parametric approaches can be based on deterministic polynomial regressions [4, 5, 6], kriging approaches [7, 8, 9, 10, 11], polynomial chaos representations (see for instance [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23], for general methods, [24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34] for integrating data using the maximum likelihood or the Bayesian method, [35, 36, 37, 38]) for sparse modeling, and [23, 39, 34, 40, 41] for polynomial chaos on manifolds.

- The nonparametric approaches can be based on kernel-based regression (see for instance [42]), and more generally on spectral approaches, including deterministic projection methods based on the choice or the construction of a deterministic reduced-order basis (ROB) (such as a Hilbert basis adapted to the operators of the problem under consideration; see for instance [43, 44]) or ROB constructed in the probabilistic context, such as principal component analysis (PCA), Proper Orthogonal Decomposition (POD), and Karhunen-Loéve expansion (these three methods belonging to the same mathematical class of the spectral methods). A direct construction of a random ROB has also been proposed (see below).

- We must obviously mention the most emerging approaches, those of Machine Learning (ML), both parametric and nonparametric, that is to say the deep learning for artificial neural networks, which are effective and validated only for "complex mapping" when big data are available. In this ML context, many efforts are performed to develop probabilistic/statistical surrogate models for the small dataset cases. This last approach is the one followed in the present work.

(ii) The developments of methods in Machine Learning for the case of incomplete datasets are not recent and have given rise to a huge number of publications with different fields of application. It is therefore not a question here of offering a review, even a very partial one, on this vast subject. We only give a few aspects and references to situate the method that is proposed to take into account incomplete data. Two main classes can be identified, statistical methods and methods based on reduced-order representation. For the statistical methods, we can refer, for instance, to methods based on the likelihood and Bayesian approaches [45], to regression analyses [46], to regression neural network ensembles for multiple imputation [47, 48], to Bayesian-based density-based clustering approaches [49], or to an adaptive Bayesian SLOPE with missing values [50]. Concerning the reduced-order representation methods, see, for instance, [51, 52] for PCA-based methods, [53] for decomposition methods, [54] for spectral methods (in fact, these three methods belong to the large class of the spectral methods). In this paper, the context of incomplete data differs from that which is generally considered in the very large number of published papers relating to this subject. Indeed, we consider dynamical systems that are under observed and we do not seek to complete the "incomplete experimental data", but to take these "experimental data" into account to update the probability measure of all the observations of the dynamical system using the Kullback-Leibler divergence minimum principle with respect to the prior probability measure constructed using only the points of the training dataset. To take into account this framework of incomplete data, the proposed method is based both on a projection derived from a PCA representation and on the use of probabilistic learning based on the Kullback minimum principle. The method is therefore, straddling statistical methods and reduced-order representation methods.

(iii) The projection-based reduced-order computational model (ROM) that has a very small dimension with respect to the one of the large-scale computational model is a very attractive and efficient method in nonlinear computational dynamics (see for instance [55, 56, 57, 58, 59, 60, 61, 62]). It should be noted that a hyperreduction method allows for achieving computational efficiency in arbitrarily nonlinear parameterized settings [63]. However, in order to obtain a robust ROM against uncertainties, which is required, for example, to transform such a ROM into a digital twin of the physical system [64, 65], it is necessary to take into account the model uncertainties induced by modeling errors so that the computational model becomes predictive. The nonparametric probabilistic method (NPM) developed in

[66, 67, 68, 69, 70, 71, 72] is a way for taking into account model-form uncertainties in nonlinear computational model. The NPM consists in substituting the reduced-order basis by a random basis whose hyperparameters of its prior probability model can be identified from targets, and therefore, allows the model to be enriched using data. We have explained above this approach that couples the parameterized reduced-order models with the NPM formulation of the model-form uncertainty, enriched by target datasets, because the method proposed is an alternative way which makes it possible to build a surrogate predictive statistical model that can also be used online. The proposed method will couple the reduced-order representation and the probabilistic learning under the constraints defined by the target datasets (see below).

(iv) The probabilistic learning is also a very active domain of research for constructing surrogate models (see for instance, [73, 74, 75, 76, 77, 78, 79, 80]). In this context, the probabilistic learning on manifolds (PLoM) method has specifically been developed for the small dataset cases [81, 82, 83, 84] for which several extensions have been proposed to take into account implicit constraints induced by physics, computational models, and measurements [85, 86, 87], to reduce the stochastic dimension using a statistical partition approach [88], and to update the prior probability measure by a target dataset whose points are, for instance, experimental realisations of the system observations [89]. This last capability of PLoM can also be viewed as an alternative method to the Bayesian inference for the high dimension [90, 91, 92, 93, 94, 95, 96, 97, 98, 99] and is a complementary approach to existing methods in machine learning for sampling distributions on manifolds under constraints. It allows for solving unsupervised and supervised problems under uncertainty for which the training sets are small. This situation is encountered in many problems of physics and engineering sciences with expensive function evaluations. The exploration of the admissible solution space in these situations is thus hampered by available computational resources. The PLoM was successfully adapted to tackle these challenges for several related problems including nonconvex optimization under uncertainty [100, 101, 102, 68, 103, 104, 105], fracture paths in random composites [106], updating digital twins under uncertainties [65], calculation of Sobol's indices [107], dynamic monitoring [108], surrogate modeling of structural seismic response [109], for the waterflooding in oil reservoir [110]. As we indicated previously, the third ingredient of the proposed method to build the predictive surrogate model is a probabilistic learning method based on the Kullback-Leibler divergence minimum principle that allows the target dataset to be integrated for updating the prior probability measure that is built with the training set. In this context, the PLoM algorithm is also used to avoid the scattering of learned realizations associated with the updated probability measure in order to preserve its concentration in the neighborhood of the random manifold defined by the stochastic computational model.

1.2. Novelty of the paper

The novelty of this paper consists in the development of a methodology for constructing a predictive statistical surrogate model to represent any parameterized, uncertain, stochastic, nonlinear computational model in cases where the dimension is high, there is partial observability leading to incomplete data for the small target dataset of identification observations, the nonlinear mapping that computes the identification observations from the computational stochastic responses is not injective, and only a small training dataset is available. Additionally, the stochasticity of the random responses is caused by uncontrolled random parameters and stochastic excitations in the nonlinear computational model. Machine learning formulations based on artificial neural networks are not well-suited for such cases due to the high dimension, small training dataset, small target dataset, incomplete data, random uncontrolled parameters, and stochastic excitations in the nonlinear computational model. The proposed approach and algorithms are based on statistical methods, adapted reduced representations, direct probabilistic learning under constraints using PLoM constrained by the target dataset (utilizing a weak formulation of the Fourier transform of probability measures), and an effective description of a predictive statistical surrogate model using conditional statistics. These statistics explore the learned dataset and can be carried out online for any given value of the control parameter without invoking the stochastic computational model.

1.3. Organization of the paper

The framework of the problem under consideration and a summary of the methodology have been presented in Section 2. Additional developments primarily focused on convergence analyses and algorithms are provided in Section 3. An application is then carried out in Section 4 and involves the mechanical system described in [70], which

is used to validate the proposed methodology. This application pertains to a three-dimensional MEMS device in which the nonlinear stochastic dynamics are studied. It is particularly interesting and challenging for the proposed approach as the effects of the nonlinearities on the nonlinear stochastic responses are highly sensitive to the values of the control parameters.

1.4. Terminology and notations

(i) *Convention used for the variables, vectors, and matrices*.
A lower-case Latin or Greek letter, such as x or η, is a deterministic real variable.
A boldface lower-case Latin or Greek letter, such as $\mathbf{x}$ or $\boldsymbol{\eta}$, is a deterministic vector.
An upper-case Latin letter, such as X, is a real-valued random variable.
A boldface upper-case Latin letter, such as $\mathbf{X}$, is a vector-valued random variable.
A lower- or upper-case Latin letter between brackets, such as $[x]$ or $[X]$, is a deterministic matrix.
A boldface upper-case letter between brackets, such as $[\mathbf{X}]$, is a matrix-valued random variable.

(ii) *Probability space, random variable, probability measure, and probability density function.*
For any finite integer $m \geq 1$, the Euclidean space $\mathbb{R}^m$ is equipped with the σ-algebra $\mathcal{B}_{\mathbb{R}^m}$. If $\mathbf{Y}$ is a $\mathbb{R}^m$-valued random variable defined on the probability space $(\Theta, \mathcal{T}, \mathcal{P})$, $\mathbf{Y}$ is a mapping $\theta \mapsto \mathbf{Y}(\theta)$ from Θ into $\mathbb{R}^m$, measurable from $(\Theta, \mathcal{T})$ into $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$, and $\mathbf{Y}(\theta)$ is a realization (sample) of $\mathbf{Y}$ for $\theta \in \Theta$. The probability distribution of $\mathbf{Y}$ is the probability measure $P_{\mathbf{Y}}(d\mathbf{y})$ on the measurable set $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$ (we will simply say on $\mathbb{R}^m$). The Lebesgue measure on $\mathbb{R}^m$ is noted $d\mathbf{y}$ and when $P_{\mathbf{Y}}(d\mathbf{y})$ is written as $p_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}$, $p_{\mathbf{Y}}$ is the probability density function (pdf) on $\mathbb{R}^m$ of $P_{\mathbf{Y}}(d\mathbf{y})$ with respect to $d\mathbf{y}$. Finally, E denotes the mathematical expectation operator and *a.s.* means "almost surely".

(iii) *Algebraic notations*.
$\mathbb{N}$, $\mathbb{R}$: set of all the integers $\{0, 1, 2, \ldots\}$, set of all the real numbers.
$\mathbb{R}^n$: Euclidean vector space on $\mathbb{R}$ of dimension n.
$\mathbb{M}_{n,m}$: set of all the $(n \times m)$ real matrices.
$\mathbb{M}_n$: set of all the square $(n \times n)$ real matrices.
$\mathbb{M}_n^+$: set of all the positive-definite symmetric $(n \times n)$ real matrices.
$[I_n]$: identity matrix in $\mathbb{M}_n$.
$\mathbf{x} = (x_1, \ldots, x_n)$: point in $\mathbb{R}^n$.
$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \ldots + x_n y_n$: inner product in $\mathbb{R}^n$.
$\| \mathbf{x} \|$: norm in $\mathbb{R}^n$ such that $\| \mathbf{x} \| = \langle \mathbf{x}, \mathbf{x} \rangle$.
$\| [x] \| = \sup_{\| \mathbf{y} \| = 1} \| [x]\, \mathbf{y} \|$ the operator norm.
$[x]^T$: transpose of matrix $[x]$.
$\mathrm{tr}\{[x]\}$: trace of the square matrix $[x]$.
$\| [x] \|_F$: Frobenius norm of matrix $[x]$.
$\delta_{kk'}$: Kronecker's symbol.
$\delta_{\mathbf{x}_0}$: Dirac measure at point $\mathbf{x}_0$.

(iv) *List of symbols*.
$\mathbb{C}$: fourth-order tensor-valued random field.
$C_u \subset \mathbb{R}^{n_u}$: support of the probability distribution of $\mathbf{U}$.
$C_w \subset \mathbb{R}^{n_w}$: admissible set of the control parameter $\mathbf{w}$, which is the support of the probability distribution of $\mathbf{W}$.
$\mathcal{D}_{\text{learn}}(\boldsymbol{\eta}_{\text{ud}})$: learned dataset for the updated random vector $\mathbf{H}_{\text{ud}}$.
$\mathcal{D}_{\text{learn}}(\mathbf{o}_{\text{ud}}, \mathbf{w}_{\text{ud}})$: learned dataset for random vector $(\mathbf{O}_{\text{ud}}, \mathbf{W}_{\text{ud}})$.
$\mathcal{D}_{\text{targ}}(\boldsymbol{\eta}_{\text{targ}})$: set of the N_r vectors, $\boldsymbol{\eta}^r_{\text{targ}}$.
$\mathcal{D}_{\text{targ}}(\mathbf{o}_{\text{id}})$: target dataset constituted of the N_r vectors $\mathbf{o}^r_{\text{targ}}$, associated with the $\mathbf{O}_{\text{id}}$.
$\mathcal{D}_{\text{learn}}(\mathbf{x}_{\text{ud}})$: learned dataset for the updated random vector $\mathbf{X}_{\text{ud}}$.
$\mathcal{D}_{\text{train}}(\boldsymbol{\eta})$: training dataset for random vector $\mathbf{H}$.
$\mathcal{D}_{\text{train}}(\mathbf{x})$: training dataset constituted of the n_d realizations $\mathbf{x}^j$ of $\mathbf{X}$.
$\mathcal{D}_{\text{train}}(\mathbf{y}, \mathbf{o}_{\text{id}}, \mathbf{w})$: training dataset constituted of the n_d realizations $(\mathbf{y}^j, \mathbf{o}^j_{\text{id}}, \mathbf{w}^j)$.

$\mathbf{f}^j$: j-th realization of $\mathbf{F}$ in the training dataset.
$\mathbf{F}$: time-discretized stochastic process indexed by $\mathcal{J}$ representing the excitation of the dynamical system.
$\boldsymbol{\eta}^j$: j-th realization of $\mathbf{H}_{\text{ud}}$.
$\boldsymbol{\eta}^r_{\text{targ}}$: projection on the reduced model in $\mathbf{H}$ of the target dataset.
$\boldsymbol{\eta}^\ell_{\text{ud}}$: ℓ-th learned realization of $\mathbf{H}_{\text{ud}}$.
$[\eta_d]$: matrix in $\mathbb{M}_{\nu,n_d}$, whose columns are $\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^{n_d}$.
$\mathbf{H}$: random vector resulting from the principal component analysis of $\mathbf{X}$.
$\mathbf{H}_{\text{ud}}$: random vector corresponding to the updating of $\mathbf{H}$ under the constraint.
$\mathcal{J}$: set of the sampling time of $[t_0\,, T]$ in n_{time} points.
$\mathcal{J}_{\text{snps}}$: set of the snapshots.
n_d: number of independent realizations of $(\mathbf{W}, \mathbf{U}, \mathbf{F})$.
n_{pMC}: number of learned realizations used by the PLoM algorithm.
n_{o}: number of components of $\mathbf{O}_{\text{id}}(\mathbf{w})$.
n_q: number of components of $\mathbf{Q}(t; \mathbf{w})$.
n_{snps}: number of points in $\mathcal{J}_{\text{snps}}$.
n_{time}: number of points in $\mathcal{J}$ (sampling time).
n_x: number of components of $\mathbf{X}$ such that $n_x = n_{\text{time}} \times n_q + n_{\text{o}} + n_w$.
n_y: number of components of $\mathbf{Y}(t; \mathbf{w})$.
ν: number of components of random vector $\mathbf{H}$ and $\mathbf{H}_{\text{ud}}$.
N_d: is $n_d \times n_{\text{snps}}$.
N_{o}: number of components of $\mathbf{O}(\mathbf{w})$.
N_r: number of points in $\mathcal{W}_{\text{targ}}$.
N_{ud}: number of learned realizations for the updated random quantities.
$\mathcal{N}$: nonlinear operator defining the stochastic equation of the computational model.
$\mathbf{o}^r_{\text{targ}}$: r-th target for $\mathbf{O}_{\text{id}}(\mathbf{w})$ ("experimental measures").
$\mathbf{o}^j$: j-th realization $O(\mathbf{y}^j)$ of $\mathbf{O}(\mathbf{w}^j)$.
$\mathbf{o}^j_{\text{id}}$: j-th realization $O_{\text{id}}(\mathbf{y}^j)$ of $\mathbf{O}_{\text{id}}(\mathbf{w}^j)$, used for the identification.
$\mathbf{o}^\ell_{\text{ud}}$: ℓ-th learned realization of $\mathbf{O}_{\text{ud}}$.
O: operator acting on $\mathbf{Y}$, defining the observations of the computational model.
O_{id}: operator acting on $\mathbf{Y}$, defining the observations of the computational model, used for the identification.
$\mathbf{O}(\mathbf{w})$: random observation $O(\mathbf{Y}(\cdot; \mathbf{w}))$ of the computational model.
$\mathbf{O}_{\text{id}}(\mathbf{w})$: random identification observation $O_{\text{id}}(\mathbf{Y}(\cdot; \mathbf{w}))$.
$\mathbf{O}_{\text{ud}}$: updated random observation of the computational model for random control parameter.
$\mathbf{O}_{\text{ud}}(\mathbf{w})$: updating of $\mathbf{O}(\mathbf{w})$ (output of the statistical surrogate model for given $\mathbf{w}$).
$\mathbf{q}^j(t)$: j-th realization of random vector $\mathbf{Q}(t; \mathbf{W})$ for $t \in \mathcal{J}$.
$\mathbb{q}^j$: vector $(\mathbf{q}^j(t_1), \ldots, \mathbf{q}^j(t_{n_{\text{time}}}))$ as the j-th realization of random vector $\mathbb{Q}$.
$\mathbf{Q}(t, \mathbf{w})$: random vector such that $\mathbf{Y}^{(n_q)}(t; \mathbf{w}) = [V]\, \mathbf{Q}(t; \mathbf{w})$.
$\mathbb{Q}$: random vector $(\mathbf{Q}(t_1; \mathbf{W}), \ldots, \mathbf{Q}(t_{n_{\text{time}}}; \mathbf{W}))$.
$\mathbf{U}$: random uncontrolled parameter of the dynamical system.
$\mathbf{u}^j$: j-th realization of $\mathbf{U}$.
$[V]$: matrix in $\mathbb{M}_{n_y,n_q}$ representing the ROB such that $\mathbf{Y}^{(n_q)}(t; \mathbf{w}) = [V]\, \mathbf{Q}(t; \mathbf{w})$.
$\mathbf{w}$: control parameters of the dynamical system.
$\mathbf{w}^j$: j-th realization of $\mathbf{W}$.
$\mathbf{w}^j_{\text{ref}}$ specific value of the control parameter $\mathbf{w}$ for the validation of the SSM.
$\mathbf{w}^\ell_{\text{ud}}$: ℓ-th learned realization of $\mathbf{W}_{\text{ud}}$.
$\mathbf{w}^r_{\text{targ}}$: r-th value of $\mathbf{w}$ for which the corresponding observation of the dynamical system is "measured".
$\mathbf{W}$: random variable modeling $\mathbf{w}$.
$\mathbf{W}_{\text{ud}}$: updated random control parameter $\mathbf{W}$.
$\mathcal{W}_{\text{targ}}$: set of the N_r values $\mathbf{w}^r_{\text{targ}}$ of $\mathbf{w}$, for which $\mathbf{o}^r_{\text{targ}}$ is given as a target.
$\mathcal{W}_{\text{train}}$: set of the n_d realizations $\mathbf{w}^j$ of $\mathbf{W}$.
$\mathbf{x}^j$: vector $(\mathbb{q}^j, \mathbf{o}^j_{\text{id}}, \mathbf{w}^j)$ of the j-th realization of $\mathbf{X} = (\mathbb{Q}, \mathbf{O}_{\text{id}}, \mathbf{W})$.

$\mathbf{x}_{\text{ud}}^{\ell}$: ℓ-th learned realization of $\mathbf{X}_{\text{ud}}$.
$[x_d]$: matrix in $\mathbb{M}_{n_x,n_d}$, whose columns are $\mathbf{x}^1,\ldots,\mathbf{x}^{n_d}$.
$\mathbf{X}$: is the random vector $(\mathbb{Q},\mathbf{O}_{\text{id}},\mathbf{W})$.
$\mathbf{X}_{\text{ud}}$: random vector corresponding to the updating of $\mathbf{X}$ under the constraint.
$\mathbf{y}^j$: j-th realization of $\{\mathbf{Y}(t;\mathbf{W}), t\in\mathcal{J}\}$ with $\mathbf{y}^j=\mathbf{y}(.\,;\mathbf{w}^j,\mathbf{u}^j,\mathbf{f}^j)$ the solution of the deterministic computational model.
$\mathbf{Y}(t;\mathbf{w})$: vector-valued stochastic response of the computational model.
$\mathbf{Y}^{(n_q)}(t;\mathbf{w})$: n_q-order approximation of $\mathbf{Y}(t;\mathbf{w})$.

(v) *List of acronyms*.
dof: degree of freedom.
ISDE: Itô stochastic differential equation.
KDE: kernel density estimation.
KLDMP: Kullback-Leibler divergence minimum principle.
PCA: principal component analysis.
pdf: probability density function.
POD: proper orthogonal decomposition.
ROB: reduced order basis.
SCM: stochastic computational model.
SSM: statistical surrogate model.

(vi) *Terminology related to the dynamical systems and various datasets*.

Control parameter: the control parameter, $\mathbf{w}$, of the computational model corresponds to the "input" of the "statistical surrogate model" (SSM). It is a deterministic parameter that belongs to an admissible set $C_w\subset\mathbb{R}^{n_w}$. For the construction of the SSM, the control parameter $\mathbf{w}$ is modeled by a random variable $\mathbf{W}$ whose realizations are generated with its given prior probability measure for which its support is the admissible set C_w. These realizations allow the training dataset to be constructed.
Uncontrolled parameter: there is an internal vector-valued parameter $\mathbf{u}$ in the computational model that is not used as an input or an output of the statistical surrogate model. For the construction of the SSM, the uncontrolled control parameter $\mathbf{u}$ is modeled by a random variable $\mathbf{U}$ whose realizations are generated with its given prior probability measure. These realizations allow the training dataset to be constructed.
Observation: the vector-valued "observation" of the computational model corresponds to the "output" of the statistical surrogate model (the quantity of interest). Concerning the prediction performed with the SSM, for a deterministic value $\mathbf{w}$ of the control parameter given in its admissible set, the SSM estimates the probability measure of the "updated random observation" (the conditional random output given $\mathbf{w}$).
Identification observation and target dataset: For the construction of the SSM, a specific vector-valued observation of the computational model is introduced and is defined as the "identification observation". The "target dataset" is constituted of a set of values of the identification observation, which are either experimental measures or other simulated data.
Training dataset: given a prior probability model of the random control and uncontrolled parameters, a training dataset is constructed, consisting of realizations of the random triplet composed of the stochastic response of the computational model, the random identification observation, and the random control parameter.
Learned datasets: there are several learned datasets, which are related to the different random variables introduced in the construction of the SSM. Each "learned dataset" is built from the corresponding "training dataset" associated with each random variable. In particular, the learned dataset for updating the random observation of the computational model, is obtained by constraining the "identification observation" by the "target dataset".
Reference dataset for validation: To validate the SSM, a specific set of values of the control parameter $\mathbf{w}$, denoted $\mathbf{w}_{\text{ref}}$, is introduced. This set is defined as the "reference dataset" and is made up of different points from those in the training dataset.

2. Summary of the proposed methodology

In this section, in order to facilitate the reading of this paper, we present a summary of the proposed method, indicating the framework of the developments and giving the main hypotheses used and the reasons for which they are introduced.

2.1. Parameterized nonlinear uncertain stochastic computational model and its observation

We consider a high-dimension nonlinear uncertain stochastic computational model of a dynamical system, parameterized by a vector-valued control parameter $\mathbf{w}$, in presence of uncertainties represented by an uncontrolled parameter that is a $\mathbb{R}^{n_u}$-valued random variable $\mathbf{U} = (U_1, \ldots, U_{n_u})$ whose support of its probability distribution is $C_u \subset \mathbb{R}^{n_u}$, and possibly with a stochastic excitation. The control parameter $\mathbf{w} = (w_1, \ldots, w_{n_w})$ is with values in an admissible subset C_w of $\mathbb{R}^{n_w}$. The time-discretized nonlinear stochastic computational model, simply called stochastic computational model (SCM), corresponds to the time-discretized nonlinear differential equation of order m_{diff} in t, which is written as,

$$\mathcal{N}\left(\mathbf{Y}(t;\mathbf{w}), \mathbf{Y}^{(1)}(t;\mathbf{w}), \ldots, \mathbf{Y}^{(m_{\text{diff}})}(t;\mathbf{w}); t, \mathbf{w}, \mathbf{U}\right) = \mathbf{F}(t;\mathbf{w}) \quad , \quad \forall\, t \in \mathcal{J} = \{t_1, \ldots, t_{n_{\text{time}}}\} \subset [t_0\,, T]\,, \tag{2.1}$$

in which $\mathcal{N}$ is a nonlinear operator and where $\mathbf{F} = \{\mathbf{F}(t;\mathbf{w}), t \in \mathcal{J}\}$ with values in $\mathbb{R}^{n_y}$ is a given time-discretized stochastic process (time series) that depends on $\mathbf{w}$. A deterministic initial condition is given at $t = t_0$, where $t_n = t_0 + n\,\Delta t$ for $n = 1, \ldots, n_{\text{time}}$. The random response is the time-discretized stochastic process $\mathbf{Y}(\cdot\,;\mathbf{w}) = \{\mathbf{Y}(t;\mathbf{w}), t \in \mathcal{J}\}$ with values in $\mathbb{R}^{n_y}$, which depends on the control parameter $\mathbf{w}$ but also on $\mathbf{U}$ and $\mathbf{F}$, and can be written as $\mathbf{Y}(t;\mathbf{w}) = \mathbf{y}(t;\mathbf{w},\mathbf{U},\mathbf{F})$ in which $\mathbf{y}$ is a deterministic vector-valued function, $(t,\mathbf{w},\mathbf{u},\mathbf{f}) \mapsto \mathbf{y}(t,\mathbf{w},\mathbf{u},\mathbf{f})$. Finally, $\{\mathbf{Y}^{(m)}(t;\mathbf{w}) = d^m\mathbf{Y}(t;\mathbf{w})/dt^m, 1 \le m \le m_{\text{diff}}\}$. The SCM is assumed to be in high dimension, that is to say, n_y is large. For all $\mathbf{w} \in C_w$ and for a given prior probability measure of $\mathbf{U}$ and $\mathbf{F}$, it is assumed that $\mathbf{Y}(\cdot\,;\mathbf{w})$ is the unique solution of the nonlinear SCM, and that $\mathbf{Y}(\cdot\,;\mathbf{w}), \mathbf{Y}^{(1)}(\cdot\,;\mathbf{w}), \ldots, \mathbf{Y}^{(m_{\text{diff}})}(\cdot\,;\mathbf{w})$ are second-order stochastic processes indexed by $\mathcal{J}$ with values in $\mathbb{R}^{n_y}$. We are interested in the prediction of an observation that is represented by a $\mathbf{w}$-dependent random variable $\mathbf{O}(\mathbf{w})$ with values in $\mathbb{R}^{N_o}$, defined by a non-injective nonlinear mapping O of $\mathbf{Y}(.\,;\mathbf{w})$ such that, $\mathbf{O}(\mathbf{w}) = O(\mathbf{Y}(.\,;\mathbf{w}))$. For example, the random vector $\mathbf{O}(\mathbf{w})$ can be related to the logarithm of the modulus of the frequency-sampled Fourier transform in time of a subset of components $\{Y_k(t;\mathbf{w}), t \in \mathcal{J}\}$ of $\{\mathbf{Y}(t;\mathbf{w}), t \in \mathcal{J}\}$. The objective of the predictions is not that of the prediction of $\mathbf{Y}(\cdot\,;\mathbf{w})$ but is that of the prediction of $\mathbf{O}(\mathbf{w})$. This remark is important within the framework of the proposed methodology.

2.2. Identification observation and small target dataset

For the probabilistic updating using a target dataset, a subset of components of $\mathbf{O}(\mathbf{w})$ is considered, which is represented by the random variable $\mathbf{O}_{\text{id}}(\mathbf{w})$ with values in $\mathbb{R}^{n_o}$ with $n_o \ll N_o$, and which will be called the "identification observation". The corresponding restriction of mapping O is denoted by O_{id}. The random identification observation is then written as $\mathbf{O}_{\text{id}}(\mathbf{w}) = O_{\text{id}}(\mathbf{Y}(.\,;\mathbf{w}))$. Relatively to $\mathbf{O}_{\text{id}}(\mathbf{w})$, a deterministic target dataset $\mathcal{D}_{\text{targ}}(\mathbf{o}_{\text{id}}) = \{\mathbf{o}^1_{\text{targ}}, \ldots, \mathbf{o}^{N_r}_{\text{targ}}\}$ is given. For each r in $\{1, \ldots, N_r\}$, the vector $\mathbf{o}^r_{\text{targ}} \in \mathbb{R}^{n_o}$ corresponds to a "measurement" performed on the dynamical system for a given value $\mathbf{w}^r_{\text{targ}} \in C_w$ of the control parameter $\mathbf{w}$. Let $\mathcal{W}_{\text{targ}} = \{\mathbf{w}^1_{\text{targ}}, \ldots, \mathbf{w}^{N_r}_{\text{targ}}\}$ be the set of these control parameter values. The target dataset is used to update the probability measure of the random surrogate model whose statistical fluctuations are induced by the modeling errors existing in the SCM and that are simulated thanks to the presence of random vector $\mathbf{U}$ that models uncertainties. It should be noted that the target dataset is associated with a partial observability, which thus corresponds to an incomplete data case, that is an important difficulty.

2.3. Small training dataset

For constructing the training dataset, a prior probability model is introduced for the random variables $\mathbf{W}$, $\mathbf{U}$, and for the time series $\mathbf{F}$ that are assumed to be statistically independent. Let $\{(\mathbf{w}^j, \mathbf{u}^j, \mathbf{f}^j), j = 1, \ldots, n_d\}$ be n_d independent realizations of $(\mathbf{W}, \mathbf{U}, \mathbf{F})$ and let $\mathcal{W}_{\text{train}} = \{\mathbf{w}^1, \ldots, \mathbf{w}^{n_d}\}$ be the set of sampled values of $\mathbf{W}$. The training dataset, $\mathcal{D}_{\text{train}}(\mathbf{y}, \mathbf{o}_{\text{id}}, \mathbf{w})$, is constituted of n_d points constructed using the SCM. For every j in $\{1, \ldots, n_d\}$, let $\mathbf{y}^j = \mathbf{y}(.\,;\mathbf{w}^j, \mathbf{u}^j, \mathbf{f}^j)$ be the solution of the deterministic computational model. Once $\mathbf{y}^j$ is known, the corresponding realization $\mathbf{o}^j_{\text{id}} = O_{\text{id}}(\mathbf{y}^j)$ of $\mathbf{O}_{\text{id}}(\mathbf{w}^j)$ is computed. The training dataset is thus defined by

$$\mathcal{D}_{\text{train}}(\mathbf{y}, \mathbf{o}_{\text{id}}, \mathbf{w}) = \{(\mathbf{y}^j, \mathbf{o}^j_{\text{id}}, \mathbf{w}^j), j = 1, \ldots, n_d\} \quad \text{with} \quad \mathbf{y}^j = \mathbf{y}(.\,;\mathbf{w}^j, \mathbf{u}^j, \mathbf{f}^j) \quad \text{and} \quad \mathbf{o}^j_{\text{id}} = O_{\text{id}}(\mathbf{y}^j)\,. \tag{2.2}$$

It should be noted that $\mathbf{y}^j = \{\mathbf{y}^j(t), t \in \mathcal{J}\}$ is a function defined on $\mathcal{J}$ with values in $\mathbb{R}^{n_y}$ while $(\mathbf{o}_{\text{id}}^j, \mathbf{w}^j)$ is in $\mathbb{R}^{n_o} \times \mathbb{R}^{n_w}$. It is assumed that the numerical cost of a single evaluation of the deterministic computational model is large. Therefore the number n_d is assumed to be small, a few tens or a few hundreds. Under these conditions the number of points in the training dataset is small and we are in the case of a small training dataset (as opposed to big data). Under these conditions, machine learning formulations based on the learning of artificial neural networks are not suitable for the considered framework of the high dimension and of small training and target datasets. For this reason, we propose a methodology based on statistical methods for constructing reduced representation and also on a direct probabilistic learning under constraints of the probability measures.

2.4. Predictive statistical surrogate model

The objective of this work is the construction of a predictive statistical surrogate model defined by the family $\{\mathbf{O}_{\text{ud}}(\mathbf{w}), \mathbf{w} \in C_w\}$ of random variables with values in $\mathbb{R}^{N_o}$, which is the probabilistic updating of the family $\{\mathbf{O}(\mathbf{w}), \mathbf{w} \in C_w\}$. This means that we have to estimate the $\mathbf{w}$-dependent probability measure $P_{\mathbf{O}_{\text{ud}}(\mathbf{w})}(d\mathbf{o}; \mathbf{w})$ of the $\mathbb{R}^{N_o}$-valued random variable $\mathbf{O}_{\text{ud}}(\mathbf{w})$ and to develop a generator of independent realizations. This problem is difficult due to the high dimension, to the partial observability inducing incomplete data for the target dataset of the identification observations, to the non-injectivity of nonlinear mapping O, and to small training and target datasets.

2.5. Large learned dataset from the small training dataset, built with a probabilistic learning constrained by the small target dataset

Below, we summarize the main steps of the proposed methodology that is developed for circumventing the identified difficulties and we explain the reasons for the choices made.

(i) Reduced representation of $\mathbf{Y}$ *with a* $(t, \mathbf{w})$*-independent ROB in* $\mathbb{R}^{n_y}$ *represented by a matrix* $[V]$ *in* $\mathbb{M}_{n_y,n_q}$. The $\mathbb{R}^{N_o}$-valued random observation $\mathbf{O}(\mathbf{w})$ is expressed as a nonlinear mapping of the finite family $\{\mathbf{Y}(t; \mathbf{w}), t \in \mathcal{J}\}$ of $\mathbb{R}^{n_y}$-valued random variables. The probabilistic learning constrained by the target dataset required to perform the learning for a random vector of length $n_y \times n_{\text{time}}$, for which n_d realizations are available, and for which $n_d \times n_{\text{MC}}$ learned realizations should be generated. Therefore, learned data would be represented by a matrix containing $(n_y \times n_{\text{time}}) \times (n_d \times n_{\text{MC}})$ 64-bit words. For instance, for $n_y = 10^5$, $n_{\text{time}} = 10^4$, $n_d = 10^2$, and $n_{\text{MC}} = 10^4$, this matrix requires 10^{15} 64-bit words! It is then necessary to construct a reduced representation of $\{\mathbf{Y}(t; \mathbf{w}), t \in \mathcal{J}\}$. However, we want to perform a global reduced representation that yields a good approximation for all t in $\mathcal{J}$ and for all $\mathbf{w}$ in C_w. We then use the existing POD methodology for constructing the reduced representation $\mathbf{Y}^{(n_q)}(\cdot\,; \mathbf{w})$ of $\mathbf{Y}(\cdot\,; \mathbf{w})$, which is written as,

$$\mathbf{Y}^{(n_q)}(t; \mathbf{w}) = [V]\, \mathbf{Q}(t; \mathbf{w}) \quad , \quad \forall t \in \mathcal{J} \ , \ \forall \mathbf{w} \in C_w \, . \tag{2.3}$$

The matrix $[V] \in \mathbb{M}_{n_y,n_q}$ is independent of t and $\mathbf{w}$, whose columns constitute an orthonormal reduced-order basis (ROB) in $\mathbb{R}^{n_y}$ (for the usual Euclidean inner product), and where $n_q \ll n_y$ is estimated by a convergence analysis. Let $\mathcal{J}_{\text{snps}} = \{\tau_1, \ldots, \tau_{n_{\text{snps}}}\} \subset \mathcal{J}$ be the subset of snapshots time with $n_{\text{snps}} < n_{\text{time}}$ (possibly with $n_{\text{snps}} = n_{\text{time}}$) and let $[\mathbb{y}^j] = [\mathbf{y}^j(\tau_1) \ldots \mathbf{y}^j(\tau_{n_{\text{snps}}})]$ be the matrix in $\mathbb{M}_{n_y,n_{\text{snps}}}$. Then matrix $[V]$ is constructed by compression of the matrix

$$[\mathbb{y}] = [\mathbb{y}^1 \ldots \mathbb{y}^{n_d}] \in \mathbb{M}_{n_y,N_d} \quad , \quad N_d = n_d \times n_{\text{snps}} \quad , \quad [\mathbb{y}^j] = [\mathbf{y}^j(\tau_1) \ldots \mathbf{y}^j(\tau_{n_{\text{snps}}})] \in \mathbb{M}_{n_y,n_{\text{snps}}} \, , \tag{2.4}$$

and is such that $[V]^T [V] = [I_{n_q}]$. It should be noted that $[\mathbb{y}]$ is the collection of matrices $[\mathbb{y}^1]$, $[\mathbb{y}^2]$, …, and $[\mathbb{y}^{n_d}]$. For all $\mathbf{w}$ in C_w, $\{\mathbf{Q}(t; \mathbf{w}), t \in \mathcal{J}\}$ is a finite family of $\mathbb{R}^{n_q}$-valued random variables (sometimes called the random vector of the generalized coordinates). The finite family of functions $\{\mathbf{y}^1, \ldots, \mathbf{y}^{n_d}\}$ with $\mathbf{y}^j : \mathcal{J} \mapsto \mathbb{R}^{n_y}$ is replaced by the finite family $\{\mathbf{q}^1, \ldots, \mathbf{q}^{n_d}\}$ with $\mathbf{q}^j : \mathcal{J} \mapsto \mathbb{R}^{n_q}$ such that, for all t in $\mathcal{J}$,

$$\mathbf{q}^j(t) = [V]^T \mathbf{y}^j(t) \quad , \quad \forall t \in \mathcal{J} \ , \ \forall j \in \{1, \ldots, n_d\} \, . \tag{2.5}$$

This means that $\mathbf{q}^j(t)$ is the projection of $\mathbf{y}^j(t)$ on the subspace of $\mathbb{R}^{n_y}$ spanned by the ROB represented by $[V]$.

(ii) Definition of the training dataset $\mathcal{D}_{\text{train}}(\mathbf{x})$. For all $j \in \{1, \ldots, n_d\}$, we define $\mathbf{x}^j \in \mathbb{R}^{n_x}$ such that

$$\mathbf{x}^j = (\mathbb{q}^j, \mathbf{o}_{\text{id}}^j, \mathbf{w}^j) \quad , \quad \mathbb{q}^j = (\mathbf{q}^j(t_1), \ldots, \mathbf{q}^j(t_{n_{\text{time}}})) \in \mathbb{R}^{n_{\text{time}} \times n_q} \quad , \quad n_x = n_{\text{time}} \times n_q + n_o + n_w \, . \tag{2.6}$$

The training dataset relative to the n_d points $\mathbf{x}^j$ is thus $\mathcal{D}_{\text{train}}(\mathbf{x}) = \{\mathbf{x}^1, \dots, \mathbf{x}^{n_d}\}$ and is represented by the matrix $[x_d] \in \mathbb{M}_{n_x,n_d}$ defined by

$$[x_d] = [\mathbf{x}^1 \dots \mathbf{x}^{n_d}] \in \mathbb{M}_{n_x,n_d} \,. \tag{2.7}$$

There are three main difficulties for using a probabilistic learning from the training dataset $\mathcal{D}_{\text{train}}(\mathbf{x})$ under the constraint defined by the target dataset $\mathcal{D}_{\text{targ}}(\mathbf{o}_{\text{id}})$. The first difficulty, which is the most important, is that we find ourselves in a case of incomplete data induced by a partial observability. Indeed, the target dataset $\mathcal{D}_{\text{targ}}(\mathbf{o}_{\text{id}})$ concerns only the part $\mathbf{o}_{\text{id}}^j \in \mathbb{R}^{n_{\text{o}}}$ of the vector $\mathbf{x}^j \in \mathbb{R}^{n_x}$. There is not a target value for the $\mathbb{q}^j \in \mathbb{R}^{n_{\text{time}} \times n_q}$ part. The second difficulty is due to the fact that there does not exist a mapping $\mathbf{o}_{\text{id}}^j \mapsto \mathbb{q}^j$ that is the inverse of the non-injective nonlinear mapping $\mathbb{q}^j \mapsto \mathbf{o}_{\text{id}}^j$. Therefore, we cannot complete the data by this way. Finally, the third difficulty is related to the fact that n_d is not equal to N_r, and therefore, the possibility to replace $\mathbf{x}^j = (\mathbb{q}^j, \mathbf{o}_{\text{id}}^j, \mathbf{w}^j)$ by a vector such as $(\mathbb{q}^j, \mathbf{o}_{\text{id}}^j, \mathbf{w}^j, \mathbf{o}_{\text{targ}}^r)$ cannot be done. We thus propose the following approach to bypass these three difficulties.

(iii) Reduced representation of $\mathbf{X}$ *using a* PCA. Let $\mathbf{X} = (\mathbb{Q}, \mathbf{O}_{\text{id}}, \mathbf{W})$ be the $\mathbb{R}^{n_x}$-valued random variable for which $\mathbf{x}^1, \dots, \mathbf{x}^{n_d}$, grouped in matrix $[x_d]$, are n_d independent realizations. A principal component analysis (PCA) of $\mathbf{X}$ is performed using $[x_d]$. The reduced representation $\mathbf{X}^{(\nu)}$ of $\mathbf{X}$ is thus obtained for which the mean-square convergence (in the vector space of all second-order $\mathbb{R}^{n_x}$-valued random variables) is controlled with respect to the dimension ν of the reduction, which is written as

$$\mathbf{X}^{(\nu)} = \underline{\mathbf{x}} + [\Phi]\,[\lambda]^{1/2}\,\mathbf{H} \quad , \quad \nu < n_x \,, \tag{2.8}$$

in which $\mathbf{H}$ is a centered $\mathbb{R}^{\nu}$-valued random variable with covariance matrix equal to $[I_\nu]$. In general, n_{o} is large and consequently, the reduction will be important yielding $\nu \ll n_{\text{o}}$. The vector $\underline{\mathbf{x}}$ is the mean value of $\mathbf{X}$ estimated with $\mathbf{x}^1, \dots, \mathbf{x}^{n_d}$. The matrix $[\Phi] \in \mathbb{M}_{n_x,\nu}$ and the positive-definite diagonal matrix $[\lambda] \in \mathbb{M}_{\nu}$ are the eigenvectors and the eigenvalues of the covariance matrix of $\mathbf{X}$, which is estimated with $\mathbf{x}^1, \dots, \mathbf{x}^{n_d}$. Matrix $[\Phi]$ is such that $[\Phi]^T[\Phi] = [I_\nu]$. By construction, we have $[x_d] = [\underline{x}] + [\Phi]\,[\lambda]^{1/2}\,[\eta_d]$ in which $[\underline{x}] = [\underline{\mathbf{x}} \dots \underline{\mathbf{x}}] \in \mathbb{M}_{n_x,n_d}$. The columns of the matrix $[\eta_d] \in \mathbb{M}_{\nu,n_d}$ are the n_d realizations $\boldsymbol{\eta}^1, \dots, \boldsymbol{\eta}^{n_d}$ of random vector $\mathbf{H}$. At mean-square convergence, matrix $[\eta_d]$ is computed by

$$[\eta_d] = [\lambda]^{-1/2}[\Phi]^T\,([x_d] - [\underline{x}]) \,. \tag{2.9}$$

The training dataset for the $\mathbb{R}^{\nu}$-valued random variable $\mathbf{H}$ is then defined as $\mathcal{D}_{\text{train}}(\boldsymbol{\eta}) = \{\boldsymbol{\eta}^1, \dots, \boldsymbol{\eta}^{n_d}\}$.

(iv) "Projection" of the target on the model in the context of incomplete data due to the partial observability. Now we have to associate a vector $\boldsymbol{\eta}_{\text{targ}}^r \in \mathbb{R}^{n_{\text{o}}}$ to $\mathbf{o}_{\text{targ}}^r$ for $r = 1, \dots, N_r$. For that we need to build a mapping that associates a vector $\boldsymbol{\eta} \in \mathbb{R}^{\nu}$ to each $\mathbf{o}_{\text{id}} \in \mathbb{R}^{n_{\text{o}}}$. We now remove superscript (ν) for simplifying the writing. For any realization $\boldsymbol{\eta}$ in $\mathbb{R}^{\nu}$ of $\mathbf{H}$, the corresponding realization $\mathbf{x}$ of $\mathbf{X}$ is given by Eq. (2.8), $\mathbf{x} = \underline{\mathbf{x}} + [\Phi]\,[\lambda]^{1/2}\,\boldsymbol{\eta}$. Since $\mathbf{x} = (\mathbb{q}, \mathbf{o}_{\text{id}}, \mathbf{w})$, the extraction of $\mathbf{o}_{\text{id}} \in \mathbb{R}^{n_{\text{o}}}$ from $\mathbf{x} \in \mathbb{R}^{n_x}$ yields

$$\mathbf{o}_{\text{id}} = \underline{\mathbf{o}}_{\text{id}} + [A_{\text{o}}]\,\boldsymbol{\eta} \quad , \quad \underline{\mathbf{o}}_{\text{id}} \in \mathbb{R}^{n_{\text{o}}} \quad , \quad [A_{\text{o}}] = [\Phi_{\text{o}}]\,[\lambda]^{1/2} \in M_{n_{\text{o}},\nu} \quad , \quad [\Phi_{\text{o}}] \in \mathbb{M}_{n_{\text{o}},\nu} \,. \tag{2.10}$$

The matrix $[A_{\text{o}}] \in M_{n_{\text{o}},\nu}$ admits a unique left pseudo-inverse $[A_{\text{o}}^{\text{inv}}] \in \mathbb{M}_{\nu,n_{\text{o}}}$ The desired mapping is constructed by solving the equation $[A_{\text{o}}]\,\boldsymbol{\eta} = \mathbf{o}_{\text{id}} - \underline{\mathbf{o}}_{\text{id}}$ in the linear least-squares sense, which admits the unique solution $\boldsymbol{\eta} = [A_{\text{o}}^{\text{inv}}]\,(\mathbf{o}_{\text{id}} - \underline{\mathbf{o}}_{\text{id}})$. We then have

$$\boldsymbol{\eta}_{\text{targ}}^r = [A_{\text{o}}^{\text{inv}}]\,(\mathbf{o}_{\text{targ}}^r - \underline{\mathbf{o}}_{\text{id}}) \quad , \quad r \in \{1, \dots, N_r\} \,. \tag{2.11}$$

Note that this construction can be viewed as a "projection" of the target onto the model in the context of incomplete data due to partial observability. In order to simplify the following explanation, let us consider that $\mathbf{F} = \mathbf{f}$ is a deterministic time function. Therefore, if the trace of the covariance matrix of the random uncontrolled parameter $\mathbf{U}$ tends to zero, then for each given value $\mathbf{w}$ of the control parameter, the level of statistical fluctuations of the stochastic response $\mathbf{Y}(\cdot, ; \mathbf{w}) = \mathbf{y}(\cdot, ; \mathbf{w}, \mathbf{U}, \mathbf{f})$ will also tend to zero (assuming continuity with respect to $\mathbf{U}$). This means that for small statistical fluctuations of $\mathbf{U}$, if the "distance" (between two clusters) from the training dataset $\mathcal{D}_{\text{train}}(\boldsymbol{\eta})$ to the target dataset $\mathcal{D}_{\text{targ}}(\boldsymbol{\eta}_{\text{targ}})$ is "significant", then the projection of the target dataset onto the model we have presented will only give a weak "contribution," and the constraint will have little effect in the probabilistic updating. To remedy this, the prior probability model of $\mathbf{U}$ has to be carefully defined in order to generate enough statistical fluctuations in the stochastic response $\mathbf{y}(\cdot, ; \mathbf{w}, \mathbf{U}, \mathbf{F})$. Thus, the level of statistical fluctuations of $\mathbf{U}$ makes it possible to control

the "diameter" of the generated family of computational models. This situation is similar (although very different methodologically and also in terms of assumptions and objectives) to that of Gaussian Bayesian inference: for a given level of Gaussian noise, the support of the prior probability measure must be adapted to the domain where the Gaussian likelihood function has significant contributions (which can be a difficult problem in high dimensions).

(v) Probabilistic learning on manifolds (PLoM) constrained by the target dataset. Using the small training dataset $\mathcal{D}_{\text{train}}(\boldsymbol{\eta}) = \{\boldsymbol{\eta}^j, j = 1, \ldots, n_d\}$ relative to $\mathbf{H}$ and using the associated small target dataset $\mathcal{D}_{\text{targ}}(\boldsymbol{\eta}_{\text{targ}}) = \{\boldsymbol{\eta}^r_{\text{targ}}, r = 1, \ldots, N_r\}$, the probabilistic learning on manifolds (PLoM) is carried out under the constraints defined by the target dataset and based on the use of a weak formulation of Fourier transform of probability measures. This step allows for generating a large learned dataset $\mathcal{D}_{\text{learn}}(\boldsymbol{\eta}_{\text{ud}}) = \{\boldsymbol{\eta}^\ell_{\text{ud}}, \ell = 1, \ldots, N_{\text{ud}}\}$ of the $\mathbb{R}^\nu$-valued random variable $\mathbf{H}_{\text{ud}}$ that is the updating of $\mathbf{H}$ under the constraint defined by $\mathcal{D}_{\text{targ}}(\boldsymbol{\eta}_{\text{targ}})$. The number of points in the learned dataset is $N_{\text{ud}} = n_d \times n_{\text{MC}}$ in which $n_{\text{MC}} \ll 1$ is given.

(vi) Construction of the large learned dataset $\mathcal{D}_{\text{learn}}(\mathbf{o}_{\text{ud}}, \mathbf{w}_{\text{ud}})$. Eq. (2.8) allows for constructing the learned dataset $\mathcal{D}_{\text{learn}}(\mathbf{x}_{\text{ud}}) = \{\mathbf{x}^\ell_{\text{ud}}, \ell = 1, \ldots, N_{\text{ud}}\}$. From Eqs. (2.3) and (2.6) and for all $\ell \in \{1, \ldots, N_{\text{ud}}\}$, we deduce the learned realizations $\mathbb{y}^\ell_{\text{ud}} = (\mathbf{y}^\ell_{\text{ud}}(t_1), \ldots, \mathbf{y}^\ell_{\text{ud}}(t_{n_{\text{time}}})) \in \mathbb{R}^{n_{\text{time}} \times n_y}$ and $\mathbf{w}^\ell_{\text{ud}} \in \mathbb{R}^{n_w}$. The learned realizations $\mathbf{o}^\ell_{\text{ud}} \in \mathbb{R}^{N_{\text{o}}}$ of the updated $\mathbf{w}$-independent random observation $\mathbf{O}_{\text{ud}}$ are computed using nonlinear operator O such that $\mathbf{o}^\ell_{\text{ud}} = O(\mathbb{y}^\ell_{\text{ud}})$. Finally, we obtain the learned dataset

$$\mathcal{D}_{\text{learn}}(\mathbf{o}_{\text{ud}}, \mathbf{w}_{\text{ud}}) = \{(\mathbf{o}^\ell_{\text{ud}}, \mathbf{w}^\ell_{\text{ud}}) \in \mathbb{R}^{N_{\text{o}}} \times \mathbb{R}^{n_w} \ , \ \ell = 1, \ldots, N_{\text{ud}}\} \, . \tag{2.12}$$

2.6. Predictive statistical surrogate model available in an online computational context

The updated statistical surrogate model is defined by the $\mathbf{w}$-dependent $\mathbb{R}^{N_{\text{o}}}$-valued random variable $\mathbf{O}_{\text{ud}}(\mathbf{w})$. Let us assume that all the considered probability measures involving in the conditional statistical estimations admit densities with respect to the Lebesgue measures $d\mathbf{o}$ and $d\mathbf{w}$ on $\mathbb{R}^{N_{\text{o}}}$ and $\mathbb{R}^{n_w}$. For any $\mathbf{w}$ given in C_w, the $\mathbf{w}$-dependent probability density function on $\mathbb{R}^{N_{\text{o}}}$ of the random variable $\mathbf{O}_{\text{ud}}(\mathbf{w})$ is given by

$$p_{\mathbf{O}_{\text{ud}}(\mathbf{w})}(\mathbf{o}; \mathbf{w}) = p_{\mathbf{O}_{\text{ud}} \,|\, \mathbf{W}_{\text{ud}}}(\mathbf{o}|\mathbf{w}) \quad , \quad \mathbf{o} \in \mathbb{R}^{N_{\text{o}}} \quad , \quad \mathbf{w} \in C_w \subset \mathbb{R}^{n_w} \, , \tag{2.13}$$

in which the conditional pdf $p_{\mathbf{O}_{\text{ud}} \,|\, \mathbf{W}_{\text{ud}}}(\mathbf{o}|\mathbf{w})$ of $\mathbf{O}_{\text{ud}}$ given $\mathbf{W}_{\text{ud}} = \mathbf{w}$ is written as $p_{\mathbf{O}_{\text{ud}} \,|\, \mathbf{W}_{\text{ud}}}(\mathbf{o}|\mathbf{w}) = p_{\mathbf{O}_{\text{ud}}, \mathbf{W}_{\text{ud}}}(\mathbf{o}, \mathbf{w}) / p_{\mathbf{W}_{\text{ud}}}(\mathbf{w})$. The joint pdf $p_{\mathbf{O}_{\text{ud}}, \mathbf{W}_{\text{ud}}}(\mathbf{o}, \mathbf{w})$ is estimated using the Gaussian KDE method with the learned dataset $\mathcal{D}_{\text{learn}}(\mathbf{o}_{\text{ud}}, \mathbf{w}_{\text{ud}})$ defined by Eq. (2.12). The estimation of the pdf $p_{\mathbf{W}_{\text{ud}}}(\mathbf{w})$ is deduced by an explicit integration with respect to $\mathbf{o}$ of the KDE estimate of $p_{\mathbf{O}_{\text{ud}}, \mathbf{W}_{\text{ud}}}(\mathbf{o}, \mathbf{w})$. Once steps (v-1) to (v-4) have been performed (offline computation), step (v-5) corresponds to the effective description of the predictive statistical surrogate model for which the output statistics can rapidly be computed in an online context using only conditional statistics on the learned dataset. For any value of the control parameter $\mathbf{w}$ given in C_w, this predictive statistical surrogate model yields the statistics of the random observations, such as mean values, variances, probability density functions, confidence domains, etc.

3. Algorithmic complements and convergence analysis

In this section we present additional developments that allow us to specify the convergence criteria and the algorithms to implement the proposed methodology for building the predictive statistical surrogate model. All notations, assumptions, developments introduced in Section 2 are used without repeating them and without systematically referring to this section.

3.1. Illustration of a parameterized nonlinear uncertain computational model

As an illustration of the dynamic computational model whose time discretization yields Eq. (2.1), we consider the $\mathbf{w}$-parametric, nonlinear, uncertain, computational model

$$[M(\mathbf{w}, \mathbf{U})] \, \ddot{\mathbf{Y}}(t; \mathbf{w}) + \mathbf{g}\left(\mathbf{Y}(t, \mathbf{w}), \dot{\mathbf{Y}}(t; \mathbf{w}); \mathbf{w}, \mathbf{U}\right) = \mathbf{F}(t; \mathbf{w}) \quad , \quad t \in]t_0, T] \, , \tag{3.1}$$

arising from a large finite element semi-discretization of a boundary value problem governing the dynamic equilibrium of a mechanical structure, where $\mathbf{Y}(t;\mathbf{w}) = \mathbf{y}(t;\mathbf{w},\mathbf{U},\mathbf{F})$ represents the n_y displacement dofs at time t, $\dot{\mathbf{Y}}(t;\mathbf{w}) = d\mathbf{Y}(t;\mathbf{w})/dt$ is the velocity vector and $\ddot{\mathbf{Y}}(t;\mathbf{w}) = d^2\mathbf{Y}(t;\mathbf{w})/dt^2$ the acceleration vector at time t. The initial conditions associated with Eq. (3.1) are written as

$$\mathbf{Y}(t_0) = \mathbf{y}_0 \quad , \quad \dot{\mathbf{Y}}(t_0) = \mathbf{y}_1 \, ,$$

where $\mathbf{y}_0$ and $\mathbf{y}_1$ are two given vectors in $\mathbb{R}^{n_y}$. For all $\mathbf{w} \in C_w$ and $\mathbf{u} \in C_u$, $[M(\mathbf{w},\mathbf{u})]$ is the mass matrix belonging to $\mathbb{M}^+_{n_y}$, $\mathbf{g}\,(\mathbf{y}(t;\mathbf{w},\mathbf{u}), d\mathbf{y}(t;\mathbf{w},\mathbf{u})/dt;\mathbf{w},\mathbf{u})$ is the $\mathbb{R}^{n_y}$ vector representing the internal nonlinear forces at time t, and $\mathbf{F}(\cdot\,;\mathbf{w})$ is a given $\mathbb{R}^{n_y}$-valued stochastic process modeling the external forces.

3.2. Construction of the ROB represented by matrix [V] and convergence criterion

Let $\varepsilon_{\text{comp}}$ be the given relative tolerance (for instance 10^{-5}) to perform the data compression of $[\mathbb{y}] \in \mathbb{M}_{n_y,N_d}$ defined by Eq. (2.4). Let $n_q^{\max}$ be a given integer whose value is of the order of the rank of matrix $[\mathbb{y}] \in \mathbb{M}_{n_y,N_d}$ with $N_d = n_d \times n_{\text{snps}}$ and such that $n_q^{\max} < n_y$. The singular value decomposition (SVD) of $[\mathbb{y}]$, restricted to the only calculation of the $n_q^{\max}$ largest singular values $\sigma_1 \geq \ldots, \geq \sigma_{n_q^{\max}}$ (ordered by descending values), allows for computing the $n_q^{\max}$ associated left-singular vectors, represented by the matrix $[V^{\max}] \in \mathbb{M}_{n_y,n_q^{\max}}$ such that $[V^{\max}]^T[V^{\max}] = [I_{n_q^{\max}}]$. If $(\sigma_{n_q^{\max}}/\sigma_1)^2 > \varepsilon_{\text{comp}}$, then $n_q^{\max}$ has been chosen too small and the computation has to be restarted with a larger value of $n_q^{\max}$. The optimal value of $n_q \leq n_q^{\max}$ is then calculated such that

$$\text{err}_{\text{comp}}(n_q;N_d) \leq \varepsilon_{\text{comp}} < \text{err}_{\text{comp}}(n_q+1;N_d) \quad \text{with} \quad \text{err}_{\text{comp}}(n_q;N_d) = 1 - \frac{\sum_{\alpha=1}^{n_q}\sigma_\alpha^2}{\sum_{\alpha=1}^{n_q^{\max}}\sigma_\alpha^2} \, . \tag{3.2}$$

Finally, matrix $[V] \in \mathbb{M}_{n_y,n_q}$ is made up of the first n_q columns of $[V^{\max}]$ and is such that $[V]^T[V] = [I_{n_q}]$. For this fixed value of n_q, matrix $[V] \in \mathbb{M}_{n_y,n_q}$ is such that

$$[V] = \arg \min_{[v]\in\mathbb{M}_{n_y,n_q},[v]^T[v]=[I_{n_q}]} \| \, [\mathbb{y}] - [v]\,[v]^T\,[\mathbb{y}] \, \|_F^2 \, ,$$

in which $\|\cdot\|_F$ is the Frobenius norm.

3.3. Reduced representation of X using PCA and convergence criterion

The reduced representation $\mathbf{X}^{(\nu)}$ defined by Eq. (2.8) of the $\mathbb{R}^{n_x}$-valued random variable $\mathbf{X} = (\mathbb{Q}, \mathbf{O}_{\text{id}}, \mathbf{W})$ is constructed as follows, using the training dataset $\mathcal{D}_{\text{train}}(\mathbf{x}) = \{\mathbf{x}^1, \ldots, \mathbf{x}^{n_d}\}$. Beforehand the PCA computation, the training dataset $\mathcal{D}_{\text{train}}(\mathbf{x})$ is scaled using the formulation presented in [81] (at the end of the numerical procedure, a back scaling must be carried out). Let $\mathbf{x}_c^j = \mathbf{x}_d^j - \underline{\mathbf{x}}$ be the realization of $\mathbf{X}$ with $\underline{\mathbf{x}} = (1/n_d)\sum_{j=1}^{n_d}\mathbf{x}_d^j \in \mathbb{R}^{n_x}$. Let $[x_c] = [\mathbf{x}_c^1 \ldots \mathbf{x}_c^{n_d}]$ be the matrix in $\mathbb{M}_{n_x,n_d}$. Since $n_x \gg n_d$, the economy size SVD (thin SVD [111] of matrix $[x_c]$ is carried out, which allows for obtaining the left-singular vectors represented by the matrix $[\Phi_c] \in \mathbb{M}_{n_x,n_d}$ such that $[\Phi_c]^T[\Phi_c] = [I_{n_d}]$. The corresponding singular values $S_1 \geq \ldots \geq S_{n_d-1} > S_{n_d} = 0$ are in decreasing order. For $\nu \leq n_d - 1$, the reduced representation $\mathbf{X}^{(\nu)}$ of $\mathbf{X}$ is given by Eq. (2.8), in which $[\Phi] \in \mathbb{M}_{n_x,\nu}$ is made up of the first ν columns of matrix $[\Phi_c]$ and where $[\lambda]$ is the diagonal matrix in $\mathbb{M}^+_\nu$ such that $[\lambda]_{\alpha\alpha} = \lambda_\alpha = S_\alpha^2/(n_d-1)$. Note that the positive real numbers $\lambda_1 \geq \ldots \geq \lambda_\nu > 0$ are the ν largest positive eigenvalues of the estimated covariance matrix $[\widehat{C}_{\mathbf{X}}]$ of the covariance matrix $[C_{\mathbf{X}}]$ of $\mathbf{X}$, performed using the training dataset. Therefore, $[\lambda]$ and $[\Phi]$ depend on n_d. As it can be seen, these eigenvalues and the associated eigenvectors are computed without computing $[\widehat{C}_{\mathbf{X}}]$ because n_x can be very large. It should also be noted that, if $n_d = n_x$ and if $\nu < n_d - 1$, then the sequence of random variables $\mathbf{X}^{(\nu)}$ is mean-square convergent to $\mathbf{X}$ when ν goes to $n_d - 1$, and if $\nu = n_d - 1 = n_x - 1$, then Eq. (2.8) is not an approximation and corresponds to a change of basis. In general, for the high-dimension problems and small training datasets, n_x is large and $n_d \ll n_x$. Therefore, Eq. (2.8) corresponds to a reduced representation, which is an approximation whose accuracy depends on ν and n_d and which is classically controlled as follows. For $n_d \ll n_x$ and for $\nu < n_d - 1$, parameter ν is chosen such that

$$\text{err}_{\text{PCA}}(\nu\,;n_d) = \frac{E\{\|\,\mathbf{X} - \mathbf{X}^{(\nu)}\,\|^2\}}{E\{\|\,\mathbf{X}\,\|^2\}} \simeq 1 - \frac{\sum_{\alpha=1}^{\nu}\lambda_\alpha}{\text{tr}[\widehat{C}_{\mathbf{X}}]} \leq \varepsilon_{\text{PCA}} \quad , \quad \nu < n_d - 1 \, , \tag{3.3}$$

in which ε_{PCA} is a given positive real number sufficiently small, where $\|\cdot\|$ is the usual Euclidean norm, and where E is the mathematical expectation operator. The trace, $\text{tr}[\widehat{C}_{\mathbf{X}}]$, of matrix $[\widehat{C}_{\mathbf{X}}]$ is calculated by estimating the diagonal entries of $[\widehat{C}_{\mathbf{X}}]$ using the training dataset. Note that $\nu \mapsto \text{err}_{\text{PCA}}(\nu\,; n_d)$ defined by Eq. (3.3) gives the relative error as a function of $\nu < n_d - 1$ for a fixed value of n_d.

3.4. "Projection" of the target on the model

We have to specify the meaning of Eq. (2.11) that we have defined as a "projection" of the target dataset onto the model in a context of incomplete data due to the partial observability. For all $r \in \{1, \ldots, N_r\}$, let $\boldsymbol{\eta}^r_{\text{targ}} \in \mathbb{R}^{n_{\text{o}}}$ be the linear least squares solution of the equation $[A_{\text{o}}]\,\boldsymbol{\eta}^r_{\text{targ}} = \mathbf{b}^r$ with $\mathbf{b}^r = \mathbf{o}^r_{\text{targ}} - \underline{\mathbf{o}}_{\text{id}} \in \mathbb{R}^{n_{\text{o}}}$, which is the unique solution of the optimization problem

$$\boldsymbol{\eta}^r_{\text{targ}} = \min_{\boldsymbol{\eta} \in \mathbb{R}^{\nu}} \|\,[A_{\text{o}}]\,\boldsymbol{\eta} - \mathbf{b}^r\,\|\,. \tag{3.4}$$

Since $\nu < n_o$, the rank of matrix $[A_{\text{o}}] \in M_{n_{\text{o}},\nu}$ is less than or equal to ν. There is always a unique left-pseudo inverse $[A^{\text{inv}}_{\text{o}}] \in \mathbb{M}_{\nu,n_{\text{o}}}$ such that on the one hand $[A^{\text{inv}}_{\text{o}}]\,[A_{\text{o}}]$ and $[A_{\text{o}}]\,[A^{\text{inv}}_{\text{o}}]$ are symmetric matrices and on the other hand such that $[A_{\text{o}}]\,[A^{\text{inv}}_{\text{o}}]\,[A_{\text{o}}] = [A_{\text{o}}]$ and $[A_{\text{o}}]\,[A^{\text{inv}}_{\text{o}}]\,[A_{\text{o}}] = [A_{\text{o}}]$. The left pseudo-inverse can be computed using the SVD of $[A_{\text{o}}]$ and then deducing $[A^{\text{inv}}_{\text{o}}]$ by inverting the non-zero singular values. If the rank of $[A_{\text{o}}]$ is equal to ν, then all the singular values are positive and $[A_{\text{o}}]^T[A_{\text{o}}]$ is invertible. In such a case the left-pseudo inverse can be written as $[A^{\text{inv}}_{\text{o}}] = ([A_{\text{o}}]^T[A_{\text{o}}])^{-1}[A_{\text{o}}]^T$. For all $\boldsymbol{\eta} \in \mathbb{R}^{\nu}$, it is known that $\|\,[A_{\text{o}}]\,\boldsymbol{\eta} - \mathbf{b}^r\,\| \geq \|\,[A_{\text{o}}]\,\boldsymbol{\eta}^r_{\text{targ}} - \mathbf{b}^r\|$ in which $\boldsymbol{\eta}^r_{\text{targ}}$ is given by $\boldsymbol{\eta}^r_{\text{targ}} = [A^{\text{inv}}_{\text{o}}]\,\mathbf{b}^r$. Therefore, the unique solution of the optimization problem defined by Eq. (3.4) is $\boldsymbol{\eta}^r_{\text{targ}} = [A^{\text{inv}}_{\text{o}}]\,\mathbf{b}^r$, (see Eq. (2.11)). The relative error in the "projection" of the target dataset on the model can be quantified, for fixed N_r, by the mapping,

$$\nu \mapsto \text{err}_{\text{targ}}(\nu; N_r) = \frac{\sum_{r=1}^{N_r} \|([I_{n_{\text{o}}}] - [A_{\text{o}}]\,[A^{\text{inv}}_{\text{o}}])(\mathbf{o}^r_{\text{targ}} - \underline{\mathbf{o}}_{\text{id}})\|^2}{\sum_{r=1}^{N_r} \|\mathbf{o}^r_{\text{targ}} - \underline{\mathbf{o}}_{\text{id}}\|^2}\,, \tag{3.5}$$

in which $[A_{\text{o}}]\,[A^{\text{inv}}_{\text{o}}]$ is the orthogonal projector onto the range of $[A_{\text{o}}]$ and consequently, where $[I_{n_{\text{o}}}] - [A_{\text{o}}]\,[A^{\text{inv}}_{\text{o}}]$ is the orthogonal projector onto the null space of $[A_{\text{o}}]^T$.

3.5. Probabilistic learning constrained by the target dataset using a weak formulation of Fourier transform of probability measures

The construction of the learned dataset $\mathcal{D}_{\text{learn}}(\boldsymbol{\eta}_{\text{ud}}) = \{\boldsymbol{\eta}^{\ell}_{\text{ud}}, \ell = 1, \ldots, N_{\text{ud}}\}$ of the $\mathbb{R}^{\nu}$-valued random variable $\mathbf{H}_{\text{ud}}$ is carried out using the Kullback-Leibler divergence minimum principle (KLDMP) based on the prior probability measure of $\mathbf{H}$ constructed with the training dataset $\mathcal{D}_{\text{train}}(\boldsymbol{\eta}) = \{\boldsymbol{\eta}^j, j = 1, \ldots, n_d\}$ and constrained by the target dataset $\mathcal{D}_{\text{targ}}(\boldsymbol{\eta}_{\text{targ}}) = \{\boldsymbol{\eta}^r_{\text{targ}}, r = 1, \ldots, N_r\}$. It should be noted that the constraints imposed for the KLDMP must be described by statistical moments, i.e., taking the form of a mathematical expectation. In the present case, the constraints are described by realizations that constitute the points of the target dataset. Therefore, to be able to impose the constraints using realizations, we use the extension [89] of the constrained PLoM [85, 86, 87]. With this extension, the constraint, which is defined by the realizations of the target dataset, is transformed into a constraint expressed in the form of a mathematical expectation, which is necessary to implement the KLDMP. This transformation of constraint formulation can be done thanks to the use of a weak formulation of the Fourier transform of the probability measures. We refer the reader to [89] for details of this method and the associated algorithms. However, we give below a very brief summary of the essential points of this method in order to facilitate the reading.

Prior probability measure of $\mathbf{H}$. Let $P_{\mathbf{H}}(d\boldsymbol{\eta}) = p_{\mathbf{H}}(\boldsymbol{\eta})\,d\boldsymbol{\eta}$ be the prior probability measure on $\mathbb{R}^{\nu}$ of $\mathbf{H}$, whose probability density function $\boldsymbol{\eta} \mapsto p_{\mathbf{H}}(\boldsymbol{\eta}) : \mathbb{R}^{\nu} \to \mathbb{R}^{+}$ is estimated by using the Gaussian kernel-density estimation (KDE) with the training dataset $\mathcal{D}_{\text{train}}(\boldsymbol{\eta}) = \{\boldsymbol{\eta}^j, j = 1, \ldots, n_d\}$, involving the modification proposed in [20] of the classical formulation [112] for which $s_{\text{SB}} = (4/(n_d(2+\nu)))^{1/(\nu+4)}$ is the Silverman bandwidth,

$$p_{\mathbf{H}}(\boldsymbol{\eta}) = \frac{1}{n_d} \sum_{j=1}^{n_d} \frac{1}{(\sqrt{2\pi}\,\hat{s})^{\nu}} \exp\left(-\frac{1}{2\hat{s}^2} \|\,\frac{\hat{s}}{s_{\text{SB}}}\,\boldsymbol{\eta}^j - \boldsymbol{\eta}\,\|^2 \right) \quad , \quad \forall \boldsymbol{\eta} \in \mathbb{R}^{\nu}\,, \tag{3.6}$$

in which $\hat{s} = s_{\text{SB}} \left(s_{\text{SB}}^2 + (n_d - 1)/n_d\right)^{-1/2}$. With such a modification, the normalization of $\mathbf{H}$ is preserved for any value of n_d, that is to say,

$$E\{\mathbf{H}\} = \int_{\mathbb{R}^\nu} \boldsymbol{\eta}\, p_{\mathbf{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \frac{1}{2\hat{s}^2}\, \widehat{\underline{\boldsymbol{\eta}}} = \mathbf{0}_\nu\,, \tag{3.7}$$

$$E\{\mathbf{H} \otimes \mathbf{H}\} = \int_{\mathbb{R}^\nu} (\boldsymbol{\eta} \otimes \boldsymbol{\eta})\, p_{\mathbf{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \hat{s}^2\, [I_\nu] + \frac{\hat{s}^2}{s_{\text{SB}}^2} \frac{(n_d - 1)}{n_d}\, [\widehat{C}_{\mathbf{H}}] = [I_\nu]\,, \tag{3.8}$$

in which $\widehat{\boldsymbol{\eta}} \in \mathbb{R}^\nu$ and $[\widehat{C}_{\mathbf{H}}] \in \mathbb{M}_\nu^+$ are the estimates of the mean value and the covariance matrix of $\mathbf{H}$, performed with $\mathcal{D}_{\text{train}}(\boldsymbol{\eta})$. Theorem 3.1 in [84] proves that, for all $\boldsymbol{\eta}$ fixed in $\mathbb{R}^\nu$, Eq. (3.6) is a consistent estimation of the sequence $\{p_{\mathbf{H}}\}_{n_d}$ for $n_d \to +\infty$.

Representation of the constraint defined by the target dataset $\mathcal{D}_{\text{targ}}(\boldsymbol{\eta}_{\text{targ}})$. In [89], it is proven that the constraint defined by the target dataset $\mathcal{D}_{\text{targ}}(\boldsymbol{\eta}_{\text{targ}}) = \{\boldsymbol{\eta}_{\text{targ}}^r, r = 1, \ldots, N_r\}$ can be written as

$$E\{\mathbf{h}^c(\mathbf{H})\} = \mathbf{b}^c \text{ on } \mathbb{R}^{N_r}\,, \tag{3.9}$$

in which $\mathbf{h}^c(\boldsymbol{\eta}) = (h_1^c(\boldsymbol{\eta}), \ldots, h_{N_r}^c(\boldsymbol{\eta}))$ and $\mathbf{b}^c = (b_1^c, \ldots, b_{N_r}^c)$ are the vectors in $\mathbb{R}^{N_r}$, which are written, for $r \in \{1, \ldots, N_r\}$ and $\boldsymbol{\eta} \in \mathbb{R}^\nu$, as

$$h_r^c(\boldsymbol{\eta}) = \exp\left(-\frac{1}{\nu s^2} \|\boldsymbol{\eta} - \boldsymbol{\eta}_{\text{targ}}^r\|^2\right) \quad , \quad b_r^c = \frac{1}{N_r} \sum_{r'=1}^{N_r} \exp\left(-\frac{1}{\nu s^2} \|\boldsymbol{\eta}_{\text{targ}}^{r'} - \boldsymbol{\eta}_{\text{targ}}^r\|^2\right), \tag{3.10}$$

in which $s = \left(4\,(N_r(2 + \nu))^{-1}\right)^{1/(\nu+4)}$.

Updated estimate using the Kullback-Leibler divergence minimum principle under the constraint. The updated probability density function $\boldsymbol{\eta} \mapsto p_{\mathbf{H}_{\text{ud}}}(\boldsymbol{\eta})$ on $\mathbb{R}^\nu$ of the $\mathbb{R}^\nu$-valued random variable $\mathbf{H}_{\text{ud}} = (H_{\text{ud},1}, \ldots, H_{\text{ud},\nu})$ is estimated by using the KLDMP [113, 114, 115, 85, 87]. The pdf $p_{\mathbf{H}_{\text{ud}}}$ on $\mathbb{R}^\nu$, which satisfies the constraint defined by Eq. (3.9) and which is closest to $p_{\mathbf{H}}$ defined by Eq. (3.6), is the solution of the following optimization problem,

$$p_{\mathbf{H}_{\text{ud}}} = \arg \min_{p \in C_{\text{ad},p}} \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, \log\left(\frac{p(\boldsymbol{\eta})}{p_{\mathbf{H}}(\boldsymbol{\eta})}\right) d\boldsymbol{\eta}\,, \tag{3.11}$$

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

$$C_{\text{ad},p} = \left\{\boldsymbol{\eta} \mapsto p(\boldsymbol{\eta}) : \mathbb{R}^\nu \to \mathbb{R}^+\,, \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, d\boldsymbol{\eta} = 1\,, \int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta})\, p(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \mathbf{b}^c\right\}. \tag{3.12}$$

It has been proven that there exists a unique solution to the optimization problem defined by Eqs. (3.11) and (3.12), which is reformulated using Lagrange multipliers to account for the constraints in the admissible set (refer to Theorem 3 in [89] for the construction of the probability measure of $\mathbf{H}_{\text{ud}}$ and the proof of its existence and uniqueness). To generate the learned dataset $\mathcal{D}_{\text{learn}}(\boldsymbol{\eta}_{\text{ud}}) = \{\boldsymbol{\eta}_{\text{ud}}^\ell, \ell = 1, \ldots, N_{\text{ud}}\}$, an MCMC algorithm is required [116, 117, 118]. In this work, the MCMC generator used is a nonlinear Itô stochastic differential equation (ISDE) associated with the nonlinear stochastic dissipative Hamiltonian dynamical system proposed in [119] and based on [120]. This MCMC generator allows for the removal of the transient part to rapidly reach the stationary response associated with the invariant measure, for which the measure $p_{\mathbf{H}_{\text{ud}}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ is the marginal measure (refer to Theorem 4 in [89] for the mathematical analysis of this MCMC generator). The ISDE is solved using the Störmer-Verlet algorithm, which provides an efficient and accurate MCMC algorithm. This algorithm can be easily parallelized to significantly reduce the elapsed time on a multicore computer. It should be noted that this MCMC generator can be considered as belonging to the class of Hamiltonian Monte Carlo methods [121, 122], but it is distinct due to the presence of the dissipative term. Finally, the algorithm presented in [89] can be readily extended to integrate the diffusion maps tool, which forms the basis of the PLoM algorithm ([81, 84, 88]). It is this extended PLoM algorithm that will be used in the application presented in Section 4.

3.6. Predictive statistical surrogate model

Once the learned dataset $\mathcal{D}_{\text{learn}}(\mathbf{o}_{\text{ud}}, \mathbf{w}_{\text{ud}}) = \{(\mathbf{o}_{\text{ud}}^{\ell}, \mathbf{w}_{\text{ud}}^{\ell}) \in \mathbb{R}^{N_{\text{o}}} \times \mathbb{R}^{n_w} \ , \ \ell = 1, \ldots, N_{\text{ud}}\}$, which are realizations of the updated $\mathbb{R}^{N_{\text{o}}} \times \mathbb{R}^{n_w}$-valued random variable $(\mathbf{O}_{\text{ud}}, \mathbf{W}_{\text{ud}})$, has been constructed (see Eq. (2.12)), the predictive statistical surrogate model defined by Eq. (2.13) can be implemented as explained at the end of Section 2.

(i) In order to prepare the developments that will be used to present the results of the application, we are going to reshape the updated $\mathbb{R}^{N_{\text{o}}}$-random variable $\mathbf{O}_{\text{ud}}$ into a $\mathbb{M}_{n_{\text{obs}},n_{\text{freq}}}$-valued random variable $[\mathbf{O}_{\text{ud}}]$ such that $N_{\text{o}} = n_{\text{obs}} \times n_{\text{freq}}$, which makes it possible to explain the dependence according to a parameter $\omega \in \mathcal{B} = \{\omega_1, \ldots, \omega_{n_{\text{freq}}}\} \subset \mathbb{R}^+$ (for instance, ω will be the frequency and n_{obs} will be the number of considered frequency-dependent observations). The corresponding learned realizations for $[\mathbf{O}_{\text{ud}}]$ are $\{[\mathbf{o}_{\text{ud}}^{\ell}] \in \mathbb{M}_{n_{\text{obs}},n_{\text{freq}}}, \ell = 1, \ldots, N_{\text{ud}}\}$. Since the statistical conditioning of $\mathbf{O}_{\text{ud}}$ given $\mathbf{W}_{\text{ud}} = \mathbf{w}$ has been written as $\mathbf{O}_{\text{ud}}(\mathbf{w})$, the statistical conditioning of $[\mathbf{O}_{\text{ud}}]$ given $\mathbf{W}_{\text{ud}} = \mathbf{w}$ is written as $[\mathbf{O}_{\text{ud}}(\mathbf{w})]$.

(ii) In addition, the results will be presented for a given set $\mathcal{W}_{\text{ref}} = \{\mathbf{w}_{\text{ref}}^1, \ldots, \mathbf{w}_{\text{ref}}^{n_{\text{ref}}}\}$ of values of the control parameter, whose n_{ref} points are defined as the "reference dataset for validation" for which the reference responses are known what allows for validating the predictions. For this reason, this subset is chosen such that $\mathcal{W}_{\text{ref}} \subset \mathcal{W}_{\text{targ}}$ but $\mathcal{W}_{\text{ref}} \not\subset \mathcal{W}_{\text{train}}$. Consequently, for any $\mathbf{w}_0$ given in $\mathcal{W}_{\text{ref}}$, there is an index $r_0 \in \{1, \ldots, N_r\}$ such that $\mathbf{o}_{\text{targ}}^{r_0} \in \mathbb{R}^{n_{\text{o}}}$ corresponds to the "measurement" performed on the dynamical system for the given value $\mathbf{w}_0 = \mathbf{w}_{\text{targ}}^{r_0}$ of the control parameter $\mathbf{w}$.

(iii) For the given value $\mathbf{w}_0$ of the control parameter, and for every observation defined by a fixed value of the observation index i in $\{1, \ldots, n_{\text{obs}}\}$ and by a fixed value of the frequency index k in $\{1, \ldots, n_{\text{freq}}\}$, the statistical surrogate model consists (see Eq. (2.13)) in estimating the probability density function $p_{Z(\mathbf{w}_0)}(z; \mathbf{w}_0)$ of the real-valued random variable $Z(\mathbf{w}_0)$ defined as the conditional random variable $Z_{\text{ud}} = [\mathbf{O}_{\text{ud}}]_{ik}$ given $\mathbf{w}_0$,

$$p_{Z(\mathbf{w}_0)}(z; \mathbf{w}_0) = p_{Z_{\text{ud}},\mathbf{W}_{\text{ud}}}(z, \mathbf{w}_0)/p_{\mathbf{W}_{\text{ud}}}(\mathbf{w}_0) \quad , \quad p_{\mathbf{W}_{\text{ud}}}(\mathbf{w}_0) = \int_{\mathbb{R}} p_{Z_{\text{ud}},\mathbf{W}_{\text{ud}}}(z, \mathbf{w}_0)\, dz \,, \tag{3.13}$$

in which $p_{Z_{\text{ud}},\mathbf{W}_{\text{ud}}}$ is the joint probability density function of Z_{ud} and $\mathbf{W}_{\text{ud}}$. It can then deduce the estimate of the confidence interval $[z^-(\mathbf{w}_0), z^+(\mathbf{w}_0)]$ of $Z(\mathbf{w}_0)$ such that

$$z^-(\mathbf{w}_0) : \int_{-\infty}^{z^-(\mathbf{w}_0)} p_{Z(\mathbf{w}_0)}(z; \mathbf{w}_0)\, dz = 1 - p_c \quad , \quad z^+(\mathbf{w}_0) : \int_{-\infty}^{z^+(\mathbf{w}_0)} p_{Z(\mathbf{w}_0)}(z; \mathbf{w}_0)\, dz = p_c \,, \tag{3.14}$$

in which p_c is a given probability level (for instance $p_c = 0.98$). The formulas to numerically estimate the confidence interval using the N_{ud} learned realizations $z_{\text{ud}}^{\ell} = [\mathbf{o}_{\text{ud}}^{\ell}]_{ik}$ and $\mathbf{w}_{\text{ud}}^{\ell}$ for $\ell = 1, \ldots, N_{\text{ud}}$, are given in Appendix A.1.

3.7. Mean-square convergence criterion of the surrogate-model prediction with respect to n_d,N_r, and N_{ud}

In order to analyze the convergence of the surrogate-model prediction with respect to the number n_d of points in training dataset $\mathcal{D}_{\text{train}}(\mathbf{y}, \mathbf{o}_{\text{id}}, \mathbf{w})$, the number N_r of points in target dataset $\mathcal{D}_{\text{targ}}(\mathbf{o}_{\text{id}}) = \{\mathbf{o}_{\text{targ}}^1, \ldots, \mathbf{o}_{\text{targ}}^{N_r}\}$, and the number $N_{\text{ud}} = n_d \times n_{\text{MC}}$ of points in learned dataset $\mathcal{D}_{\text{learn}}(\mathbf{o}_{\text{ud}}, \mathbf{w}_{\text{ud}})$, we introduce a mean-square convergence criterion of the surrogate-model prediction. For each given value $\mathbf{w}_0$ of the control parameter and for each observation $i_{\text{o}} \in \{1, \ldots, N_{\text{o}}\}$, for which the corresponding component of the target is $i \in \{1, \ldots, n_{\text{o}}\}$ and which corresponds to the "measurement" superscript $r = r_{\text{o}}$ associated with $\mathbf{w}_0$, this criterion is defined as the square of a coefficient of variation,

$$\text{cv}(\mathbf{w}_0, i_{\text{o}})^2 = \frac{\sum_{k=1}^{n_{\text{freq}}} \left(E\left\{ [\mathbf{O}_{\text{ud}}(\mathbf{w}_0)]_{i_{\text{o}}k}^2 \right\} - (E\{[\mathbf{O}_{\text{ud}}(\mathbf{w}_0)]_{i_{\text{o}}k}\})^2 \right)}{\sum_{k=1}^{n_{\text{freq}}} \left(E\{[\mathbf{O}_{\text{ud}}(\mathbf{w}_0)]_{i_{\text{o}}k}\} \right)^2} \,, \tag{3.15}$$

in which $[\mathbf{O}_{\text{ud}}(\mathbf{w}_0)]_{i_{\text{o}}k}$ is the conditional random variable given $\mathbf{W}_{\text{ud}} = \mathbf{w}_0$ such that,

$$E\{[\mathbf{O}_{\text{ud}}(\mathbf{w}_0)]_{i_{\text{o}}k}^{\alpha}\} = \int_{\mathbb{R}^+} o_k^{\alpha}\, p_{[\mathbf{O}_{\text{ud}}(\mathbf{w}_0)]_{i_{\text{o}}k}}(o_k; \mathbf{w}_0, i_{\text{o}})\, do_k \quad , \quad \alpha = 1, 2 \,, \tag{3.16}$$

in which the conditional pdf (with support $\mathbb{R}^+$) of $[\mathbf{O}_{\rm ud}(\mathbf{w}_0)]_{i_o k}$ is written as

$$p_{[\mathbf{O}_{\rm ud}(\mathbf{w}_0)]_{i_o k}}(o_k; \mathbf{w}_0, i_o) = p_{[\mathbf{O}_{\rm ud}(\mathbf{w}_0)]_{i_o k}, \mathbf{W}_{\rm ud}}(o_k, \mathbf{w}_0; i_o)/p_{\mathbf{W}_{\rm ud}}(\mathbf{w}_0) \quad , \quad \forall o_k \in \mathbb{R}^+ . \tag{3.17}$$

The formulas, which allow for numerically estimating the conditional pdf $p_{[\mathbf{O}_{\rm ud}(\mathbf{w}_0)]_{i_o k}}(\cdot\,; \mathbf{w}_0, i_o)$ and $E\{[\mathbf{O}_{\rm ud}(\mathbf{w}_0)]^\alpha_{i_o k}\}$ using the $N_{\rm ud}$ learned realizations $[\mathbf{O}_{\rm ud}(\mathbf{w}_0)]^\ell_{i_o k}$ and $\mathbf{w}^\ell_{\rm ud}$ for $\ell = 1, \ldots, N_{\rm ud}$, are given in Appendix A.2. For a fixed value of n_d, for given $\mathbf{w}_0$ and i_o, the convergence analysis, with respect to n_d, N_r, and $n_{\rm MC}$ is performed by studying the values of $\mathrm{cv}(n_d, N_r, n_{\rm MC}; \mathbf{w}_0, i_o)$.

3.8. Distance of the statistical-surrogate-model predictions to the target

In order to evaluate the quality of the predictions carried out by the statistical surrogate model, we introduce the mean-square distance between the predictions and the target for each given value $\mathbf{w}_0$ of the control parameter and for each observation $i_o \in \{1, \ldots, N_o\}$ for which there is a corresponding target $i \in \{1, \ldots, n_o\}$ (as in Section 3.7). We thus define the real-valued random variable R_{i_o} by

$$R_{i_o} = \Big(\sum_{k=1}^{n_{\rm freq}} ([\mathbf{O}_{\rm ud}]_{i_o k} - [\mathbf{o}^{r_0}_{\rm targ}]_{ik})^2\Big)/\Big(\sum_{k=1}^{n_{\rm freq}} [\mathbf{o}^{r_0}_{\rm targ}]^2_{ik}\Big) , \tag{3.18}$$

in which $[\mathbf{o}^{r_0}_{\rm targ}]$ is the reshaping of $\mathbf{o}^{r_0}_{\rm targ} \in \mathbb{R}^{n_o}$. Note that R_{i_o} is not the statistical conditioning for given $\mathbf{W}_{\rm ud} = \mathbf{w}_0$ (we have just introduced the translation $[\mathbf{o}^{r_0}_{\rm targ}]_{ik}$ that depends on $\mathbf{w}_0$). The $N_{\rm ud}$ learned realizations of R_{i_o} are computed by

$$r^\ell_{i_o} = \frac{\sum_{k=1}^{n_{\rm freq}} ([\mathbf{o}^\ell_{\rm ud}]_{i_o k} - [\mathbf{o}^{r_0}_{\rm targ}]_{ik})^2}{\sum_{k=1}^{n_{\rm freq}} [\mathbf{o}^{r_0}_{\rm targ}]^2_{ik}} \quad , \quad \ell \in \{1, \ldots, N_{\rm ud}\} . \tag{3.19}$$

We now introduce the statistical conditioning $R_{i_o}(\mathbf{w}_0)$ of R_{i_o} given $\mathbf{W}_{\rm ud} = \mathbf{w}_0$ whose conditional probability density function $r \mapsto p_{R_{i_o}(\mathbf{w}_0)}(r; \mathbf{w}_0)$ on $\mathbb{R}$ with support $\mathbb{R}^+$ is written as

$$p_{R_{i_o}(\mathbf{w}_0)}(r; \mathbf{w}_0) = p_{R_{i_o}, \mathbf{W}_{\rm ud}}(r, \mathbf{w}_0)/p_{\mathbf{W}_{\rm ud}}(\mathbf{w}_0) , \tag{3.20}$$

in which $p_{R_{i_o}, \mathbf{W}_{\rm ud}}$ is the joint probability density function of R_{i_o} and $\mathbf{W}_{\rm ud}$. For given $\mathbf{w}_0$ and i_o, the mean-square distance between the prediction and the target is defined by $d_{\rm ms}(\mathbf{w}_0, i_o) = (E\{R_{i_o}(\mathbf{w}_0)\})^{1/2}$, and consequently,

$$d_{\rm ms}(\mathbf{w}_0, i_o)^2 = \int_{\mathbb{R}^+} r\, p_{R_i(\mathbf{w}_0)}(r; \mathbf{w}_0)\, dr . \tag{3.21}$$

The formulas, which allow for numerically estimating the pdf $p_{R_i(\mathbf{w}_0)}(\cdot\,; \mathbf{w}_0)$ and the mean-square distance $d_{\rm ms}(\mathbf{w}_0, i_o)$ using the $N_{\rm ud}$ learned realizations r^ℓ_i and $\mathbf{w}^\ell_{\rm ud}$ for $\ell = 1, \ldots, N_{\rm ud}$, are given in Appendix A.3.

4. Application: nonlinear stochastic dynamics of a three-dimensional MEMS device

In this application, we partially reuse the nonlinear dynamical system presented in [66], but with different parameter values and the introduction of a vector-valued control parameter and a vector-valued random uncontrolled parameter. This system is particularly interesting because its response is highly sensitive to the nonlinearities considered, and there is a significant transfer of energy in the stochastic response outside the frequency band of the external excitation applied to the system. A 3D view of the system is shown in Figure 1-mid. We provide all the dimensions and mechanical constants values to allow for reproducibility of this application. Additionally, some values have been modified (if the system definition were based on [66], it would result in numerous back-and-forth references and make the text less readable).

4.1. Nonlinear dynamical system

(i) Definition of the nonlinear dynamical system. This 3D MEMS device considered is constituted of a mobile part made up of a square frame with a vertical beam attached to it. Its suspended part is constituted of a parallelepipedic

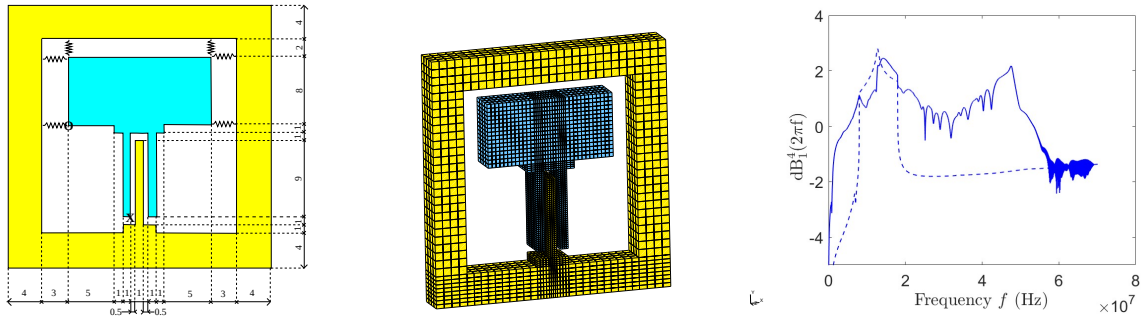


Figure 1: **Left figure**: 2D view of the MEMS device with lengths in $10^{-6}\,m$ (from [66]); observation node of coordinates $(14, 6, 0)$ microns: marked by symbol **x**; observation node of coordinates $(7, 17, 0)$ microns: marked by symbol **o**. **Mid figure**: 3D view of the finite element mesh of the nonlinear computational stochastic dynamical model (from [66]) for which the reference frame is the Cartesian coordinate system $Ox_1x_2x_3$ attached to the mobile part, with origin O located at the bottom left corner of the device, axis Ox_1 horizontal and oriented positively from left to right, axis Ox_2 vertical and oriented positively from bottom to top, axis Ox_3 perpendicular to the plane Ox_1x_2 oriented positively from bottom to top. **Right figure**: illustration of the linear response (dashed line) and of the nonlinear response (solid line) in the frequency domain and in dB of the nominal dynamical system for a given observation and for a given value of the control parameter.

solid with two attached vertical beams. The suspended part is attached to the mobile part by a 3D suspension made of 20 springs. The geometry is described in a Cartesian coordinate system $Ox_1x_2x_3$ (reference frame) that is attached to its mobile part and is defined in caption of Figure 1. The external width of the square frame is $30\times10^{-6}\,m$, its external height is $31\times10^{-6}\,m$, and its depth is $4\times10^{-6}\,m$. The other dimensions are given in microns in Figure 1-left. The stiffnesses of the springs and suspension depend on the axis along which they act: $\underline{k}_{s1} = 4\ N/m$ (along Ox_1), $\underline{k}_{s2} = 6\ N/m$ (along Ox_2), and $\underline{k}_{s3} = 1.5\ N/m$ (along Ox_3). The suspended and mobile parts are made of a homogeneous, orthotropic, linear elastic material whose nominal mechanical properties in the aforementioned reference frame are those of a standard (100) silicon wafer [123]. In the Voigt notation, its nominal orthotropic elasticity matrix is $[\underline{C}^{\mathrm{sym}}_{\mathrm{elas}}] \in \mathbb{M}^+_6$ for which the Young moduli are $\underline{E}_{11} = \underline{E}_{22} = 169\times10^9\ N/m^2$ and $\underline{E}_{33} = 130\times10^9\ N/m^2$; the Poisson ratios are $\underline{\nu}_{23} = 0.36$, $\underline{\nu}_{31} = 0.28$, and $\underline{\nu}_{12} = 0.064$; the shear moduli are $\underline{G}_{23} = \underline{G}_{31} = 79.6\times10^9\ N/m^2$ and $\underline{G}_{12} = 50.9\times10^9\ N/m^2$. The nominal mass density of the silicon material is $2\,330\ Kg/m^3$. A nonlinear elastic material is inserted between the aforementioned beams. Its constitutive equation corresponds to a cubic, elastic, restoring force with elastic constant k_b whose nominal value is $\underline{k}_b = 2\times10^{12}\ N/m$. A zero x_1-, x_2-, and x_3-displacement boundary conditions with respect to the reference frame $Ox_1x_2x_3$ are prescribed at the base of the mobile part of the device. In the reference frame, the following time-dependent, square integrable, and real-valued x_1-acceleration is applied to the base,

$$\Gamma(t) = \Gamma_0\,\{\sin(t(\omega_c + \Delta\omega_c/2)) - \sin(t(\omega_c - \Delta\omega_c/2))\}/(\pi\,t) \quad , \quad \forall t \in [t_0, T]\,,$$

where Γ_0 is the amplitude whose nominal value is $\underline{\Gamma}_0 = 120\,m/s^2$, where $\omega_c = 2\,\pi\times13\times10^6\ rad/s$ is the central angular frequency, and where $\Delta\omega_c = 2\,\pi\times10\times10^6\ rad/s$ is the angular frequency bandwidth. The energy of the excitation signal is mainly concentrated in the frequency band $[-\omega_e, -\omega_{\min}]\cup[\omega_{\min}, \omega_e]$, where $\omega_{\min} = \omega_c - \Delta\omega_c/2 = 2\pi\times8\times10^6\ rad/s$ and $\omega_e = \omega_c + \Delta\omega_c/2 = 2\pi\times18\times10^6\ rad/s$. At time t_0, the device is at rest (its displacement and velocity fields are zero). In all analyses, the time-interval of analysis is $[t_0, T]$ with $t_0 = -2.7778\times10^{-5}\ s$ and $T = 4.6403\times10^{-5}\ s$. For this value of T, the device is returned to its zero equilibrium with a small relative error. If the dynamical system was linear, the energy of the response signal would be concentrated in the same frequency band $[\omega_{\min}, \omega_e]$ as that of the energy of the excitation signal. Due to the nonlinearity however, part of the energy of the excitation signal is transferred outside its frequency band and consequently, the frequency band of the response is not $[\omega_{\min}, \omega_e]$ but $[0, \omega_{\max}]$ in which $\omega_{\max}$ has been identified as $\omega_{\max} = 2\pi\times72\times10^6\ rad/s$ (see figure 1-right. The time discretization is performed with a sampling time-step $\Delta t = \pi/\omega_{\max} = 6.9444\times10^{-9}\ s$. This yields $n_{\mathrm{time}} = 10\,682$ time points in the time-interval $[t_0, T]$. For computing the Fourier transform of the observations (x_1-acceleration at observed points) from the time-discretized responses, the sampling frequency step is set to $\Delta\omega = 2\pi\times13\,481\ rad/s$, yielding also $n_{\mathrm{freq}} = 10\,682$ frequency points in the frequency band of analysis $[-\omega_{\max}, \omega_{\max}]$. The results will be presented in the frequency domain for the frequency band $\mathcal{B}_a = [0, \omega_a]$, where $\omega_a = 2\,\pi\times70\times10^6\ rad/s$.

(ii) Control parameter $\mathbf{w}$. There are two control parameters and consequently, $\mathbf{w} = (w_1, w_2) \in C_w \subset \mathbb{R}^{n_w}$ with $n_w = 2$ and where the admissible set of $\mathbf{w}$ is $C_w = [0.5\,, 1] \times [0.5\,, 0.9]$. Component w_1 allows for controlling the amplitude Γ_0 of the x_1-acceleration applied to the base, such that $\Gamma_0 = w_1\, \underline{\Gamma}_0$. Component w_2 allows for controlling the elastic constant k_b of the nonlinear elastic material, such that $k_b = w_2\, \underline{k}_b$.

(iii) Uncontrolled random parameter $\mathbf{U}$. As we have explained in Section 2, the level of statistical fluctuations of $\mathbf{U}$ makes it possible to control the "diameter" of the generated family of computational models. We choose this uncontrolled random parameter $\mathbf{U}$ as the components of the random elasticity matrix $[\mathbf{C}_{\text{elas}}]$ with values in $\mathbb{M}_6^+$ whose mean value is $[\underline{C}_{\text{elas}}^{\text{sym}}] \in \mathbb{M}_6^+$ and whose level of statistical fluctuations is controlled by a hyperparameter δ_{train} that will be fixed to 0.2. This value has been identified as large enough to generate a training dataset with sufficiently large statistical fluctuations to correctly update the probability measure of $\mathbf{H}_{\text{ud}}$ under the constraint of the target dataset. The mean model is orthotropic and we could used a random orthotropic model as presented in [124]. Nevertheless, for the reasons given above, it is better adapted to use for the statistical fluctuations an anisotropic model, the one introduced in [125] yielding $[\mathbf{C}_{\text{elas}}] = [\underline{L}_{\text{elas}}^{\text{sym}}]^T\, [\mathbf{G}]\, [\underline{L}_{\text{elas}}^{\text{sym}}]$ in which $[\underline{C}_{\text{elas}}^{\text{sym}}] = [\underline{L}_{\text{elas}}^{\text{sym}}]^T\, [\underline{L}_{\text{elas}}^{\text{sym}}]$ and where the probability measure, the hyperparameter δ_{train}, and the random generator of random matrix $[\mathbf{G}]$ with values in $\mathbb{M}_6^+$ is given Page 103 of [126], and is such that $E\{[\mathbf{G}]\} = [I_6]$. This random generator allows for calculating n_d independent realizations $\{\mathbf{u}^j, j = 1, \ldots n_d\}$ of $\mathbf{U}$.

(iv) Observation $\boldsymbol{O}(\boldsymbol{w})$ *and identification observation* $\mathbf{O}_{\text{id}}(\mathbf{w})$. The observation and the identification observation are relative to the x_1-acceleration in the frequency domain of $n_{\text{obs}} = 744$ observed nodes of the finite element mesh (see Paragraph (v) below) distributed across various locations of the MEMS device, in particular on the boundaries of the vertical beams of the suspended and mobile parts.

(iv-1) Observation $\mathbf{O}(\mathbf{w})$. The frequency sampling is $n_{\text{freq}}^{\text{obs}} = 5\,193$ frequency points $\omega_k^{\text{obs}} = \omega_1 + (k-1)\,\Delta\omega$ for $k \in \{1, \ldots, n_{\text{freq}}^{\text{obs}}\}$ and with $\omega_1 = 2\pi \times 13\,481\ rad/s$ and $\Delta\omega = 2\pi \times 13\,481\ rad/s$. We defined the matrix-valued observation $[\mathbf{O}(\mathbf{w})]$ with values in $\mathbb{M}_{n_{\text{obs}}, n_{\text{freq}}^{\text{obs}}}$ such that, for $m \in \{1, \ldots, n_{\text{obs}}\}$ and $k \in \{1, \ldots, n_{\text{freq}}^{\text{obs}}\}$, $[\mathbf{O}(\mathbf{w})]_{mk} = \log((\omega_k^{\text{obs}})^2\, |\widehat{Y}_{i_m}(\omega_k^{\text{obs}})|)$ with $\widehat{\mathbf{Y}}(\omega_k^{\text{obs}}) = \int_{t_0}^T \exp(-i\omega_k^{\text{obs}} t)\, \mathbf{Y}(t)\, dt$. In this formula, i_m is the dof number corresponding to the x_1-displacement of the m-th observed node. Observation $\mathbf{O}(\mathbf{w})$, which is obtained by reshaping $[\mathbf{O}(\mathbf{w})]$, is with values in $\mathbb{R}^{N_{\text{o}}}$ with $N_{\text{o}} = n_{\text{obs}} \times n_{\text{freq}}^{\text{obs}} = 3\,863\,592$ and is used for presenting the predictions obtained with the statistical surrogate model. Nevertheless, in order to keep the number of figures within a reasonable limit, two observations are selected among all 744 possible ones. These observations are relative to two observed nodes in the suspended part, the first one having coordinates $(14, 6, 0)$ microns, located at the down left corner of the elastic beam of the suspended part and the second one having coordinates $(7, 17, 0)$ microns, located at the down left corner of the massive suspended part (see Figure 1-left). These two observations are the $\mathbb{R}^{n_{\text{freq}}^{\text{obs}}}$-valued random variable denoted by $\mathbf{O}_{\text{obs}}^1(\mathbf{w})$ and $\mathbf{O}_{\text{obs}}^2(\mathbf{w})$.

(iv-2) Identification observation $\mathbf{O}_{\text{id}}(\mathbf{w})$. There are $n_{\text{freq}}^{\text{id}} = 519$ frequency points $\omega_k^{\text{id}} = \omega_1 + (k-1)\,\delta\omega^{\text{id}}$ for $k \in \{1, \ldots, n_{\text{freq}}^{\text{id}}\}$ and with $\delta\omega^{\text{id}} = 2\pi \times 134\,810\ rad/s$ (there are 10 times less frequency points). We defined the matrix-valued identification observation $[\mathbf{O}_{\text{id}}(\mathbf{w})]$ with values in $\mathbb{M}_{n_{\text{obs}}, n_{\text{freq}}^{\text{id}}}$ such that, for $m \in \{1, \ldots, n_{\text{obs}}\}$ and $k \in \{1, \ldots, n_{\text{freq}}^{\text{id}}\}$, $[\mathbf{O}_{\text{id}}(\mathbf{w})]_{mk} = \log((\omega_k^{\text{id}})^2\, |\widehat{Y}_{i_m}(\omega_k^{\text{id}})|)$ with $\widehat{\mathbf{Y}}(\omega_k^{\text{id}}) = \int_{t_0}^T \exp(-i\omega_k^{\text{id}} t)\, \mathbf{Y}(t)\, dt$. Identification observation $\mathbf{O}_{\text{id}}(\mathbf{w})$, which is obtained by reshaping $[\mathbf{O}_{\text{id}}(\mathbf{w})]$, is with values in $\mathbb{R}^{n_{\text{o}}}$ with $n_{\text{o}} = n_{\text{obs}} \times n_{\text{freq}}^{\text{id}} = 386\,136$ and is used to construct the learned dataset.

4.2. Computational model

The finite element mesh is shown in Figure 1-mid. There are 7 328 eight-nodes solid elements, 10 675 nodes, and $n_y = 32\,025$ dofs. There are 205 of these nodes, which belong to the base of the mobile part of the device: at each of these nodes, all displacement dofs are constrained to zero in the moving reference frame $Ox_1x_2x_3$, due to the boundary conditions. Hence, there are 615 zero Dirichlet conditions. The governing equation for the relative displacement vector, $\mathbf{Y}(t; \mathbf{w})$, is of the type of Eq. (3.1) and written as

$$[M]\, \ddot{\mathbf{Y}}(t; \mathbf{w}) + [D]\, \dot{\mathbf{Y}}(t; \mathbf{w}) + [K(\mathbf{U})]\, \mathbf{Y}(t; \mathbf{w}) + \mathbf{f}_{\text{NL}}(\mathbf{Y}(t; \mathbf{w}); w_2) = \mathbf{f}(t; w_1) \quad , \quad t \in]t_0, T]\,, \tag{4.1}$$

in which the external force $\mathbf{f}$ is deterministic and depends on the control parameter w_1 via the amplitude $\Gamma_0 = w_1\, \underline{\Gamma}_0$ of the imposed acceleration. The internal nonlinear elastic force $\mathbf{f}_{\text{NL}}$, which are generated by the nonlinear elastic

material inserted between the elastic beams, depends on the control parameter w_2 via the elastic constant $k_b = w_2\,\underline{k}_b$. The stiffness matrix $[K(\mathbf{U})]$ of the suspended and mobile parts depends on the uncontrolled random parameter $\mathbf{U}$ that is the reshaping of the random elasticity matrix $[\mathbf{C}_{\text{elas}}]$ of the linear elastic silicon. The mass matrix $[M]$ and the damping matrix $[D]$ is independent of $\mathbf{w}$ and $\mathbf{U}$. Damping matrix $[D]$ is constructed using the global damping model described in Appendix A of [66], which is applied to the nominal model and for which the damping rate is $\xi_d = 0.02$. The nominal computational model is given by Eq. (4.1) in which $\mathbf{U}$ is replaced by the deterministic vector $\underline{\mathbf{u}}$ corresponding to the reshaping of the nominal elasticity matrix $[\underline{C}^{\text{sym}}_{\text{elas}}]$.

4.3. Target dataset and reference dataset for validation

The set $\mathcal{W}_{\text{targ}} = \{\mathbf{w}^1_{\text{targ}}, \ldots, \mathbf{w}^{N_r}_{\text{targ}}\}$ of the considered values of the control parameters for generating the target dataset $\mathcal{D}_{\text{targ}}(\mathbf{o}_{\text{id}}) = \{\mathbf{o}^1_{\text{targ}}, \ldots, \mathbf{o}^{N_r}_{\text{targ}}\}$ defined in Section 2.2, results from an experimental plan of points chosen on a Cartesian grid of N_r points in the admissible set C_w defined in Section 4.1(ii). For each value of N_r, the nodes of this 2D-grid is approximatively uniform and lightly modified in order to contain four points $\mathbf{w}_{\text{ref}} \in C_w \subset \mathbb{R}^{n_w}$ defined in the reference dataset $\mathcal{W}_{\text{ref}} = \{\mathbf{w}^1_{\text{ref}} = (0.6, 0.6), \mathbf{w}^2_{\text{ref}} = (0.6, 0.8), \mathbf{w}^3_{\text{ref}} = (0.9, 0.6), \mathbf{w}^4_{\text{ref}} = (0.9, 0.8)\}$ of the reference control parameters. The considered values of N_r are 4, 16, 30, 63, 104, and 208. It should be noted that it is normal to choose the set $\mathcal{W}_{\text{ref}} \subset \mathcal{W}_{\text{targ}}$ because the reference responses that are used to validate the predictions made with the statistical surrogate model must correspond to values of the control parameter for which the response "measurements" were made. For each fixed value of N_r, target dataset $\mathcal{D}_{\text{targ}}(\mathbf{o}_{\text{id}})$ is generated using the methodology presented in Section 2.2 and the deterministic nominal computational model defined by Eq. (3.20) in which the nominal value $\mathbf{U} = \underline{\mathbf{u}}$ has been replaced by another value $\mathbf{U} = \mathbf{u}_{\text{ref}}$ corresponding to a modification of the elasticy constants in elasticity matrix $[\underline{C}^{\text{sym}}_{\text{elas}}]$. For the validation, the reference dataset is then defined by $\mathcal{D}_{\text{ref}}(\mathbf{o}) = \left\{\{\mathbf{O}^i_{\text{obs}}(\mathbf{w}^1_{\text{ref}}), \ldots, \mathbf{O}^i_{\text{obs}}(\mathbf{w}^4_{\text{ref}})\}, i = 1, 2\right\}$,

4.4. Constructing the training dataset

The prior probability model of the control parameter $\mathbf{w}$ is the $\mathbb{R}^{n_w}$-valued random variable $\mathbf{W}$ for which its probability measure is uniform on the admissible set C_w defined in Section 4.1(ii). Consequently, for given n_d, the set $\mathcal{W}_{\text{targ}} = \{\mathbf{w}^1, \ldots \mathbf{w}^{n_d}\}$ is made up of n_d realizations drawn from this uniform distribution on C_w. The considered values of n_d are 100, 200, and 300. For each fixed value of n_d, the training dataset $\mathcal{D}_{\text{train}}(\mathbf{y}, \mathbf{o}_{\text{id}}, \mathbf{w})$ is generated using the computational model defined by Eq. (3.20) and the methodology presented in Section 2.

4.5. Reduced representation of $\boldsymbol{Y}$, training dataset $\mathcal{D}_{\text{train}}(\boldsymbol{x})$, and reduced representation of $\boldsymbol{X}$

(i) SVD-based data compression for computing $[V]$. The reduced representation of $\{\mathbf{Y}(t; \mathbf{w}), t \in \mathcal{J}, \mathbf{w} \in C_w\}$, defined in Section 2.5(i), is constructed with the $(t, \mathbf{w})$-independent ROB in $\mathbb{R}^{n_y}$ represented by matrix $[V] \in \mathbb{M}_{n_y, n_q}$ that is computed as explained in Section 3.2. The computation of $[V]$ is carried out with $n_{\text{snps}} = n_{\text{time}} = 10\,682$ and $n_q^{\max} = 20$. For $n_d \in \{100, 200, 300\}$, the value of $N_d = n_d \times n_{\text{snps}}$ are 1 068 200, 2 136 400, and 3 204 600, respectively, and Figure 2-left displays the graph of function $n_q \mapsto \text{err}_{\text{comp}}(n_q; N_d)$ defined by Eq. (3.2). The three curves are almost superimposed. Using the criterion defined by Eq. (3.2) with $\text{err}_{\text{comp}} = 10^{-5}$ yields $n_q = 11$ that is independent of the three considered values of n_d. This small value of n_q shows the efficiency of the POD method for reducing data.

(ii) Constructing training dataset $\mathcal{D}_{\text{train}}(\mathbf{x})$ and PCA-based reduced representation of $\mathbf{X}$. The training dataset $\mathcal{D}_{\text{train}}(\mathbf{x})$ is then constructed using Section 2.5(ii) and for $j \in \{1, \ldots, n_d\}$, its points $\mathbf{x}^j \in \mathbb{R}^{n_x}$ are defined by Eq. (2.6) with $n_x = 10\,682 \times 12 + 386\,136 + 2 = 514\,322$ that is, in this case, independent of n_d. The reduced representation of $\mathbf{X}$ using PCA is performed as explained in Section 3.3. Note that, for this application, n_x is independent of N_d, but the points of $\mathcal{D}_{\text{train}}(\mathbf{x})$ depend on n_d, and consequently, for a fixed value of the tolerance ε_{PCA}, dimension ν is n_d-dependent. For $n_d \in \{100, 200, 300\}$, Figure 2-right displays the graph of function $\nu \mapsto \text{err}_{\text{PCA}}(\nu\,; n_d)$ defined by Eq. (3.3). Choosing $\varepsilon_{\text{PCA}} = 0.01$ yields for ν the values 44, 57, and 62, respectively. It should be noted that the chosen value of $\varepsilon_{\text{PCA}} = 10^{-2}$ might appear to large. In fact, calculations were made with smaller values (up to 10^{-6}) and showed that, for this application, there was no significant impact on the predictions made by the statistical surrogate model. This choice makes it possible to reduce the computational costs.

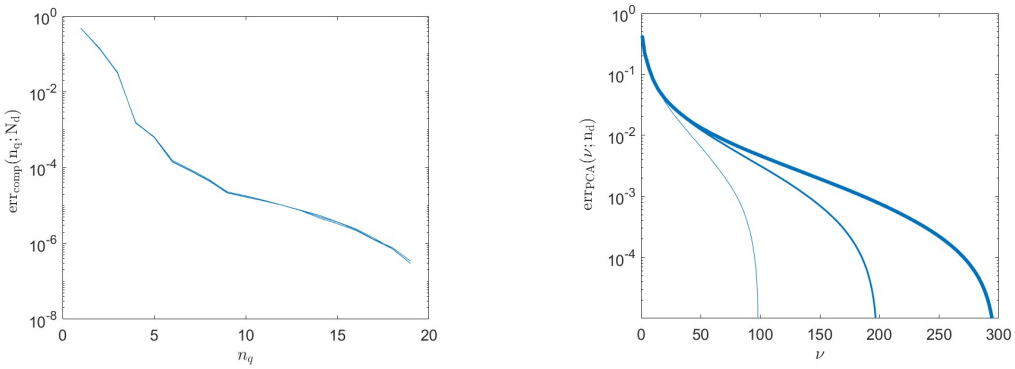


Figure 2: Left figure: for $n_d \in \{100, 200, 300\}$ yielding $N_d \in \{1\,068\,200, 2\,136\,400, 3\,204\,600\}$, graph of function $n \mapsto \mathrm{err}_{\mathrm{comp}}(n; N_d)$ (the three curves are almost superimposed). Right figure: for $n_d = 100$ (thin line), 200 (mid thickness line), 300 (thick line), graph of function $\nu \mapsto \mathrm{err}_{\mathrm{PCA}}(\nu\,; n_d)$.

4.6. Convergence analysis of the "projection" of the target on the model

(i) For $n_d = 300$, $\varepsilon_{\mathrm{PCA}} = 10^{-2}$, and $N_r = 208$, Figure 3-left shows the graph of function $\nu \mapsto \mathrm{err}_{\mathrm{targ}}(\nu\,; N_r)$ defined by Eq. (3.5) for which $\nu \in \{1, \ldots, 299\}$, and which allows for analyzing the convergence of the "projection" of the target on the model. The dimension ν of the reduced representation of $\mathbf{X}$ is constructed with the PCA of $\mathbf{X}$ and the truncation error is given by the function $\nu \mapsto \mathrm{err}_{\mathrm{PCA}}(\nu\,; n_d)$ defined by Eq. (3.3) (see Figure 2-right). It can be seen that for $\nu = 62$ the PCA error is $\mathrm{err}_{\mathrm{PCA}}(62\,; 300) = 0.00997$ and the corresponding value of the projection error of the target on the model is $\mathrm{err}_{\mathrm{targ}}(62\,; 208) = 0.1126$. It should be noted that, for $\nu = 62$, this projection error of the target on the model is small enough not to significantly impact the updating of the learned probability measure of $\mathbf{H}_{\mathrm{ud}}$ obtained by the probabilistic learning under the constraint defined by the projected target. Note also that for this case ($n_d = 300$), the asymptotic value of the error, obtained for $\nu = 299$ is $\mathrm{err}_{\mathrm{targ}}(299\,; 208) = 0.094$. Since there is a partial observability (744 components of the 32 025 components of $\mathbf{Y}(\cdot\,; \mathbf{w})$ are only used to construct the identification observation), which induces incomplete data. This level, 0.094, of the projection error of the target could be reduced in considerably increasing the number n_d of points in the training dataset, that would not be coherent with the framework of the proposed methodology for which a small training dataset is assumed (small value of n_d).

(ii) In addition, for $\varepsilon_{\mathrm{PCA}}$ fixed to the value 10^{-2} and for $N_r = 208$, Figure 3-right displays the graph of function $n_d \mapsto \mathrm{err}_{\mathrm{targ}}(\nu(n_d)\,; N_r)$ for $n_d \in \{100, 200, 300\}$ whose corresponding values of $\nu(n_d)$ are $\{44, 57, 62\}$. It can be seen that for these fixed values of N_r and $\varepsilon_{\mathrm{PCA}}$, the projection error stays sufficiently small.

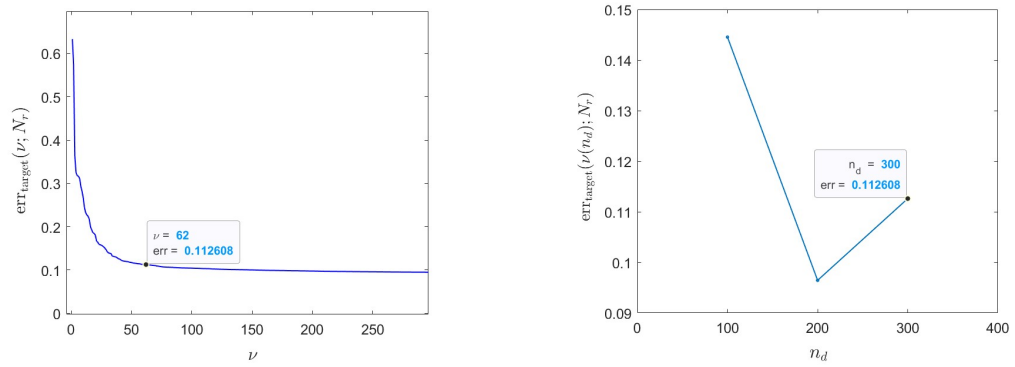


Figure 3: For $n_d = 300$, $\varepsilon_{\mathrm{PCA}} = 10^{-6}$, and $N_r = 208$, graph of function $\nu \mapsto \mathrm{err}_{\mathrm{targ}}(\nu\,; N_r)$ (left figure). For $\varepsilon_{\mathrm{PCA}} = 10^{-2}$ and $N_r = 208$, graph of function $n_d \mapsto \mathrm{err}_{\mathrm{targ}}(\nu(n_d)\,; N_r)$ with $n_d \in \{100, 200, 300\}$ and $\nu(n_d) \in \{44, 57, 62\}$ (right figure).

4.7. Construction of the learned dataset $\mathcal{D}_{\text{learn}}(\boldsymbol{\eta}_{\text{ud}})$ using PLoM constrained by the target dataset

The methodology presented in Sections 2.5(v) and 3.5 is used for constructing the learned dataset $\mathcal{D}_{\text{learn}}(\boldsymbol{\eta}_{\text{ud}})$. In addition the PLoM method is used, *i.e.* the MCMC generator, which is based on an Itô equation derived from a dissipative Hamiltonian formulation, is projected on the diffusion maps basis (see the methodology in [81, 84] and the last version of the used algorithm in [88]). Let ε_{opt} be the parameter of the kernel used for constructing the transition matrix defined in Section 5.2 of [84] and let $m_{\text{opt}} = \nu + 1$ be the dimension of the diffusion maps basis (see [88]). Then for $\varepsilon_{\text{PCA}} = 10^{-2}$, for $n_d = 100, 200, 300$, yielding $\nu = 44, 57, 62$ respectively, we have $\varepsilon_{\text{opt}} = 106.5, 150.5, 192.5$ respectively, and the distributions of the eigenvalues $\alpha \mapsto \Lambda_\alpha$ of the transition matrix are shown in Figure 4-left. As explained in Section 3.5, the optimization problem defined by Eqs. (3.11) and (3.12) is solve with the algorithm presented in Sections 6.3 to 6.6 of [89]. For $N_r = 208$, for the three considered values of n_d, and for $n_{\text{MC}} = 10\,000$, Figure 4-right shows the graphs of the error function $i \mapsto \text{err}(i)$ defined by Eq. (6.36) of [89], which allows for controlling the convergence of the iteration algorithm as a function of the iteration number i in order to estimate the optimal value of the vector-valued Lagrange multiplier. The results show in Figure 5 are obtained for $\varepsilon_{\text{PCA}} = 10^{-2}$,

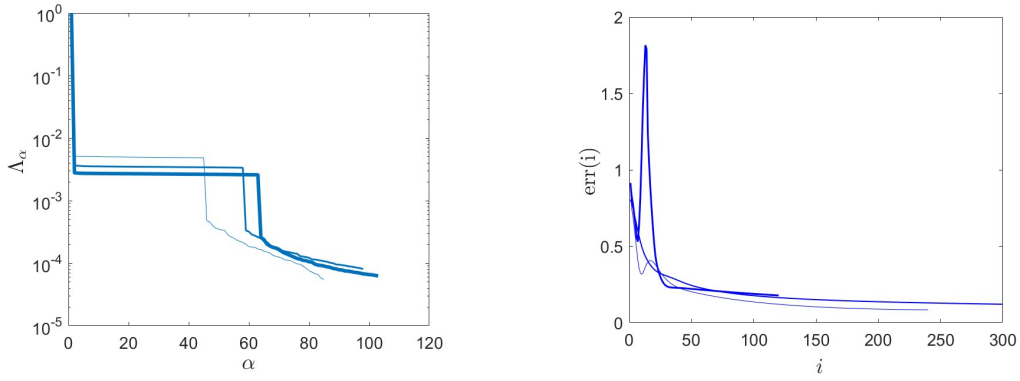


Figure 4: For $\varepsilon_{\text{PCA}} = 10^{-2}$ and $n_d = 100$ (thin line), 200 (medium line), and 300 (thick line); graph of function $\alpha \mapsto \Lambda_\alpha$ (left figure). For $N_r = 208$ and for $n_{\text{MC}} = 10\,000$, graph of function $i \mapsto \text{err}(i)$ (right figure).

$n_d = 300$, $n_{\text{MC}} = 10\,000$, $N_{\text{ud}} = n_d \times n_{\text{MC}} = 3\,000\,000$ learned realizations, and $N_r = 208$. Figure 5-left shows the cloud of the learned realizations $\{(\eta^\ell_{\text{ud},1}, \eta^\ell_{\text{ud},2}, \eta^\ell_{\text{ud},3}), \ell = 1, \ldots, N_{\text{ud}}\}$ (red points) and also the cloud of the $n_d = 300$ points $\{(\eta^j_1, \eta^j_2, \eta^j_3), j = 1, \ldots, n_d\}$ (blue points) of the training dataset. It can be seen that the constraints defined by the target dataset has a strong effect; the prior probability measure of $\mathbf{H}$ is strongly modified by the constraints yielding the updated probability measure of $\mathbf{H}_{\text{ud}}$. As an illustration, Figure 5-right displays the pdf $\eta_3 \mapsto p_{H_3}(\eta_3)$ estimated with the n_d points of the training set and the pdf $\eta_3 \mapsto p_{H_{\text{ud},3}}(\eta_3)$ estimated with the N_{ud} points of the learned dataset (solid line). It can be seen the strong effect of the constraints defined by the target set.

4.8. Construction of the large learned dataset $\mathcal{D}_{\text{learn}}(\boldsymbol{o}_{\text{ud}}, \boldsymbol{w}_{\text{ud}})$ and convergence analysis

The construction of the large learned dataset $\mathcal{D}_{\text{learn}}(\mathbf{o}_{\text{ud}}, \mathbf{w}_{\text{ud}})$ defined by Eq. (2.12) is carried out by using Sections 2.5(vi) and 3.6.

The convergence analysis with respect to N_r is carried out using the coefficient of variation $\text{cv}(\mathbf{w}^j_{\text{targ}}, i; N_r)$ defined by Eq. (3.15), for the points of the reference dataset $\mathcal{D}_{\text{ref}}(\mathbf{o}_{\text{ud}}) = \{\{\mathbf{O}^i_{\text{ud,obs}}(\mathbf{w}^1_{\text{ref}}), \ldots, \mathbf{O}^i_{\text{ud,obs}}(\mathbf{w}^4_{\text{ref}})\}, i = 1, 2\}$ defined in Section 4.3. For $n_d = 300$, $\varepsilon_{\text{PCA}} = 10^{-2}$, and for $n_{\text{MC}} = 10\,000$, Figure 6-left for observation $i = 1$ and Figure 6-right for observation $i = 2$ display the graph of function $N_r \mapsto \text{cv}(\mathbf{w}^j_{\text{ref}}, i; N_r)$ for the four values $j = 1, 2, 3, 4$ of the reference control parameters. In Figure 6, it can be seen that the coefficient of variation is decreasing with respect to N_r when $N_r \geq 60$ for observation 1 and $N_r \geq 100$ for observation 2. Fluctuations in the convergence for the small values of N_r can be observed. This is due to the fact that the coefficient of variation is sensitive to the points chosen to define the target dataset. Nevertheless, such a sensitivity is lost when N_r increases because the distribution of points in the target dataset tends to become more homogeneous. Clearly, if the distribution of points is homogeneous for a given value of N_r, then it would also be homogeneous for a large value of N_r.

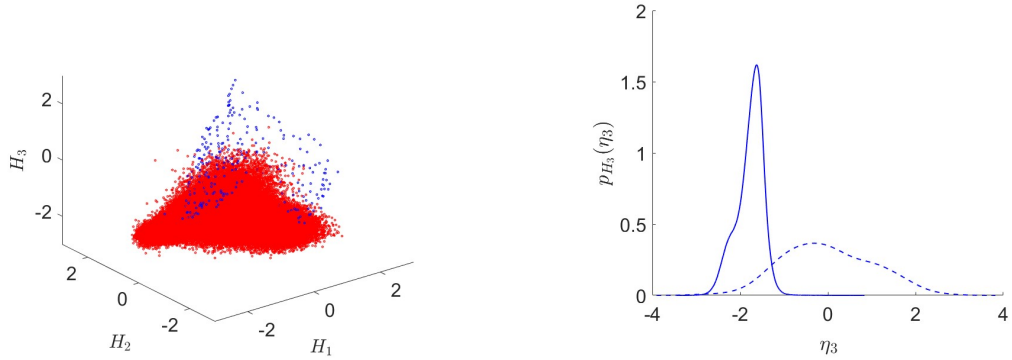


Figure 5: Learned dataset for $\varepsilon_{\text{PCA}} = 10^{-2}$, $n_d = 300$, $n_{\text{MC}} = 10\,000$, $N_{\text{ud}} = 3\,000\,000$, and $N_r = 208$. Left figure: cloud of the learned realizations $\{(\eta^\ell_{\text{ud},1}, \eta^\ell_{\text{ud},2}, \eta^\ell_{\text{ud},3}), \ell = 1, \ldots, N_{\text{ud}}\}$ (red points) and cloud of the $n_d = 300$ points $\{(\eta^j_1, \eta^j_2, \eta^j_3), j = 1, \ldots, n_d\}$ (blue points) of the training dataset. Right figure: pdf $\eta_3 \mapsto p_{H_3}(\eta_3)$ estimated with the training set (dashed line) and pdf $\eta_3 \mapsto p_{H_{\text{ud},3}}(\eta_3)$ estimated with the learned dataset (solid line).

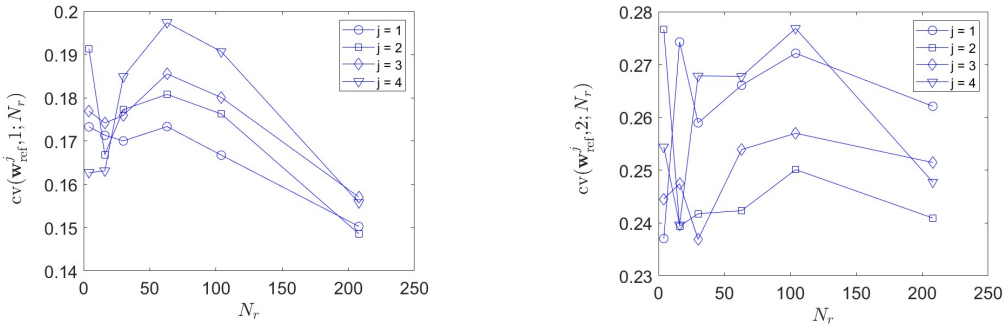


Figure 6: For $n_d = 300$, $\varepsilon_{\text{PCA}} = 10^{-2}$, and $n_{\text{MC}} = 10\,000$, convergence analysis with respect to N_r: graph of function $N_r \mapsto \text{cv}(\mathbf{w}^j_{\text{ref}}, i; N_r)$ for observations $\mathbf{O}^1_{\text{ud,obs}}(\mathbf{w}^j_{\text{ref}})$ (left figure) and $\mathbf{O}^2_{\text{ud,obs}}(\mathbf{w}^j_{\text{ref}})$ (right figure), for the four values $j = 1, 2, 3, 4$ of the reference control parameters.

Similarly, for $N_r = 208$, the convergence analysis with respect to $n_d \in \{100, 200, 300\}$ for $\varepsilon_{\text{PCA}} = 10^{-2}$ and $n_{\text{MC}} = 10\,000$ is carried out using the coefficient of variation $\text{cv}(\mathbf{w}^j_{\text{ref}}, i; n_d)$. Figure 7-left for observation $i = 1$ and Figure 7-right for observation $i = 2$ display the graph of function $n_d \mapsto \text{cv}(\mathbf{w}^j_{\text{ref}}, i; n_d)$ for the four values $j = 1, 2, 3, 4$ of the reference control parameters. Figure 7 shows that, for $n_d \geq 100$, the coefficient of variation is decreasing with respect to n_d (except for the first point $\mathbf{w}^1_{\text{ref}}$) for observation 1 while is decreasing for observation 2 for j equal to 2 and 4. It is clear that there is no reason for the convergence with respect to n_d to be monotonous for $n_d \geq 1$. Since $\varepsilon_{\text{PCA}} = 10^{-2}$ is fixed for all values of n_d, the reduced dimension ν varies according to n_d.

4.9. Distance between the surrogate-model prediction and the target

For $n_d = 300$, $\varepsilon_{\text{PCA}} = 10^{-2}$, and $n_{\text{MC}} = 10\,000$, we have estimated the square $d_{\text{ms}}(\mathbf{w}^j_{\text{ref}}, i; N_r)^2$ of the distance defined by Eq. (3.21) between the surrogate-model prediction and the target, for the two observations $\{\mathbf{O}^i_{\text{ud,obs}}(\mathbf{w}^j_{\text{ref}}), i = 1, 2\}$ and for the four values $\{\mathbf{w}^j_{\text{ref}}, j = 1, 2, 3, 4\}$ of the reference control parameters. Figure 8-left for observation 1 and Figure 8-right for observation 2 displays the set of the values $d_{\text{ms}}(\mathbf{w}^j_{\text{ref}}, i; N_r)^2$ for $N_r \in \{4, 16, 30, 63, 104, 208\}$. These figures show that, for $N_r = 208$, the values belong to the interval $[0.012\,, 0.075]$ for observation 1 and to the interval $[0.023\,, 0.12]$ for obseravtion 2. These figures quantify the quality of the predictive statistical surrogate model and it can be seen that the prediction is good enough. For $\varepsilon_{\text{PCA}} = 10^{-2}$, $n_d = 300$, $n_{\text{MC}} = 10\,000$, $N_{\text{ud}} = 3\,000\,000$, and $N_r = 208$, the graph of the conditional pdf $r \mapsto p_{R_i(\mathbf{w}^j_{\text{ref}})}(r; \mathbf{w}^j_{\text{ref}})$, defined Eq. (3.20), of the random variable R_i given $\mathbf{W}_{\text{ud}} = \mathbf{w}^j_{\text{ref}}$, for the two observations $\{\mathbf{O}^i_{\text{ud,obs}}(\mathbf{w}^j_{\text{ref}}), i = 1, 2\}$ and for the four values $\{\mathbf{w}^j_{\text{ref}}, j = 1, 2, 3, 4\}$ of the reference control parameters, is displayed in Figure 9-left for observation 1 (with a bimodal pdf for $j = 3$) and in Figure 9-right for observation 2 (with a multimodal pdf for $j = 4$).

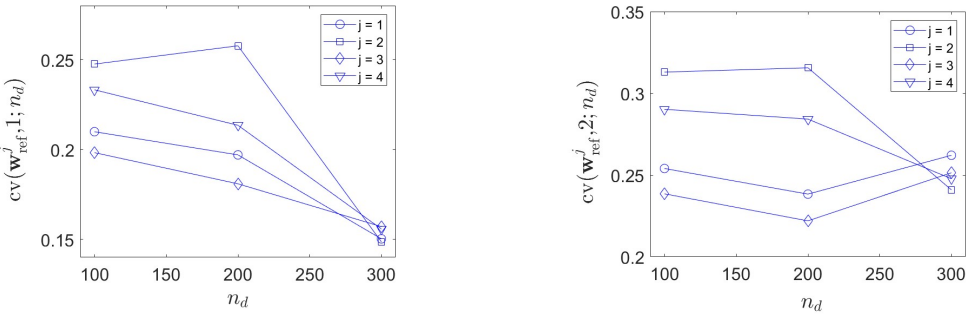


Figure 7: For $N_r = 208$, for $\varepsilon_{\mathrm{PCA}} = 10^{-2}$, and for $n_{\mathrm{MC}} = 10\,000$, convergence analysis with respect to n_d: graph of function $n_d \mapsto \mathrm{cv}(\mathbf{w}^j_{\mathrm{ref}}, i; n_d)$ for observations $\mathbf{O}^1_{\mathrm{ud,obs}}(\mathbf{w}^j_{\mathrm{ref}})$ (left figure) and $\mathbf{O}^2_{\mathrm{ud,obs}}(\mathbf{w}^j_{\mathrm{ref}})$ (right figure), for the four values $j = 1, 2, 3, 4$ of the reference control parameters.

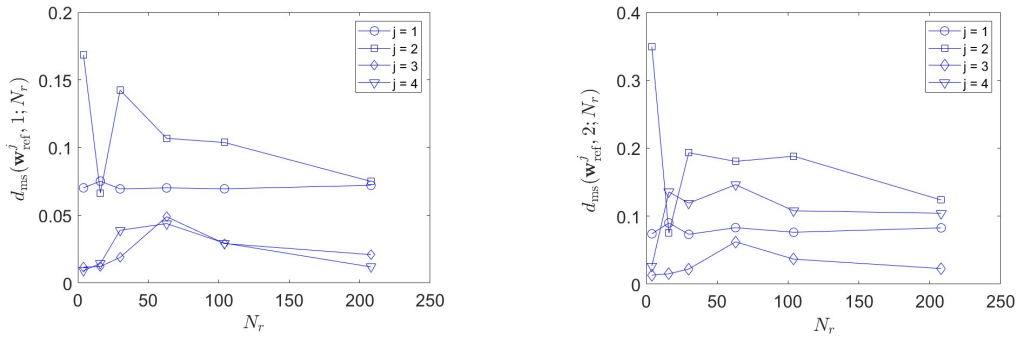


Figure 8: For $n_d = 300$, $\varepsilon_{\mathrm{PCA}} = 10^{-2}$, $n_{\mathrm{MC}} = 10\,000$, and for $N_r \in \{4, 16, 30, 63, 104, 208\}$, set of the values $d_{\mathrm{ms}}(\mathbf{w}^j_{\mathrm{ref}}, i; N_r)^2$ of the square of the distance between the surrogate-model prediction and the target for observations $\mathbf{O}^1_{\mathrm{ud,obs}}(\mathbf{w}^j_{\mathrm{ref}})$ (left figure) and $\mathbf{O}^2_{\mathrm{ud,obs}}(\mathbf{w}^j_{\mathrm{ref}})$ (right figure), for the four values $j = 1, 2, 3, 4$ of the reference control parameters.

4.10. Prediction of the statistical surrogate model

The prediction of the statistical surrogate model is carried out by using Section 3.6(iii) with $\varepsilon_{\mathrm{PCA}} = 10^{-2}$, $n_d = 300$, $n_{\mathrm{MC}} = 10\,000$, $N_{\mathrm{ud}} = 3\,000\,000$, and $N_r = 208$. For these values of the main parameters of the algorithms, the projection error of the target dataset on the model stays sufficiently small and is converged with respect to ν (see Section 4.6), the convergence is reasonnably reached with respect to n_d and N_r (see Section 4.8), and the quality of the predictive surrogate model is good enough (see Section 4.9). The prediction of the statistical surrogate model consists in estimating the conditional confidence regions for the two observations $\{\mathbf{O}^i_{\mathrm{ud,obs}}(\mathbf{w}^j_{\mathrm{ref}}), i = 1, 2\}$ and for the four values $\{\mathbf{w}^j_{\mathrm{ref}}, j = 1, 2, 3, 4\}$ of the reference control parameters. Taking into account the definition of these two observations given in Section 4.1(iv-1), we introduce the frequency dependent function $\omega \mapsto \mathrm{dB}^1_{\mathrm{j}}(\omega)$ and $\omega \mapsto \mathrm{dB}^2_{\mathrm{j}}(\omega)$ such that $\mathrm{dB}^{\mathrm{i}}_{\mathrm{j}}(\omega_{\mathrm{k}}) = \{\mathbf{O}^{\mathrm{i}}_{\mathrm{ud,obs}}(\mathbf{w}^{\mathrm{j}}_{\mathrm{ref}})\}_{\mathrm{k}}$ for $k = 1, \ldots, n^{\mathrm{obs}}_{\mathrm{freq}}$. For each observation $i \in \{1, 2\}$ and for each value $\mathbf{w}^j_{\mathrm{ref}}$ with $j \in \{1, 2, 3, 4\}$ of the reference control parameter, the conditional confidence region is estimated with a probability level $p_c = 0.98$. Figures 10 and 11 display, for observations 1 and 2 respectively, the conditional confidence region of $f \mapsto \mathrm{dB}^{\mathrm{i}}_{\mathrm{j}}(2\pi\mathrm{f})$ and also the deterministic functions for the training and for the target corresponding to $\mathbf{w} = \mathbf{w}^j_{\mathrm{ref}}$. These figures show the the prediction is good. In particular, it can be seen in Figures 10-(c) and (d) and in Figures 11-(c) and -(d. The effects of the learning under the constraints defined by the target dataset are very visible: the target line is inside the conditional confidence region whereas the training line is outside this region.

5. Conclusion

We have presented a novel methodology that addresses the challenging problem of constructing predictive statistical surrogate models for parameterized uncertain nonlinear computational models. These difficulties arise from

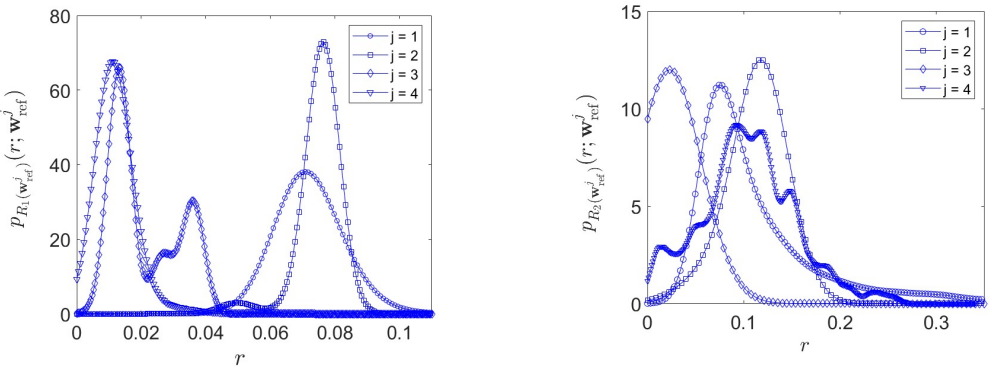
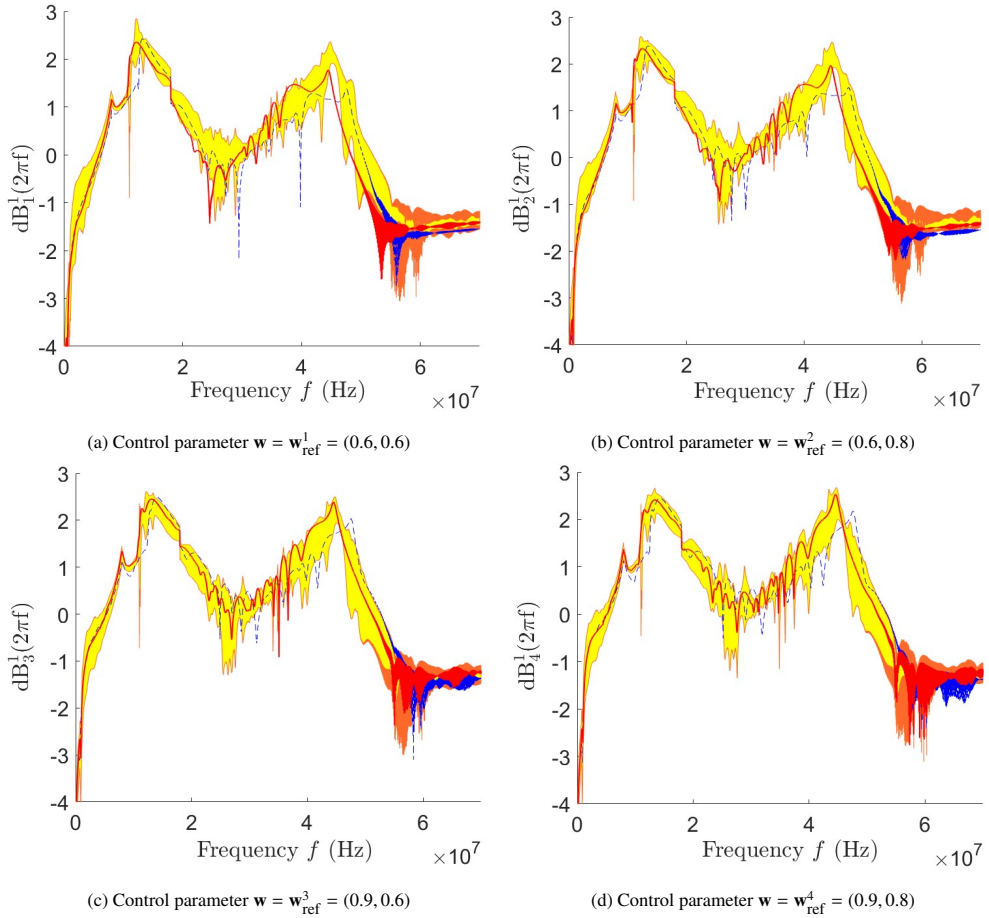


Figure 9: For $\varepsilon_{\text{PCA}} = 10^{-2}$, $n_d = 300$, $n_{\text{MC}} = 10\,000$, $N_{\text{ud}} = 3\,000\,000$, and $N_r = 208$, graph of the conditional pdf $r \mapsto p_{R_i(\mathbf{w}^j_{\text{ref}})}(r; \mathbf{w}^j_{\text{ref}})$ for observation $\mathbf{O}^1_{\text{ud,obs}}(\mathbf{w}^j_{\text{ref}})$ (left figure) and for observation $\mathbf{O}^2_{\text{ud,obs}}(\mathbf{w}^j_{\text{ref}})$ (right figure), for the four values $\{\mathbf{w}^j_{\text{ref}}, j = 1, 2, 3, 4\}$ of the reference control parameters.



(a) Control parameter $\mathbf{w} = \mathbf{w}^1_{\text{ref}} = (0.6, 0.6)$

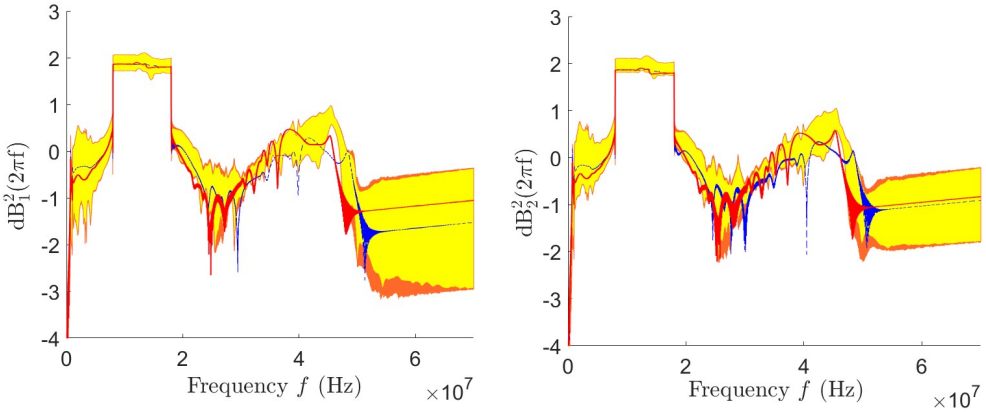
(b) Control parameter $\mathbf{w} = \mathbf{w}^2_{\text{ref}} = (0.6, 0.8)$

(c) Control parameter $\mathbf{w} = \mathbf{w}^3_{\text{ref}} = (0.9, 0.6)$

(d) Control parameter $\mathbf{w} = \mathbf{w}^4_{\text{ref}} = (0.9, 0.8)$

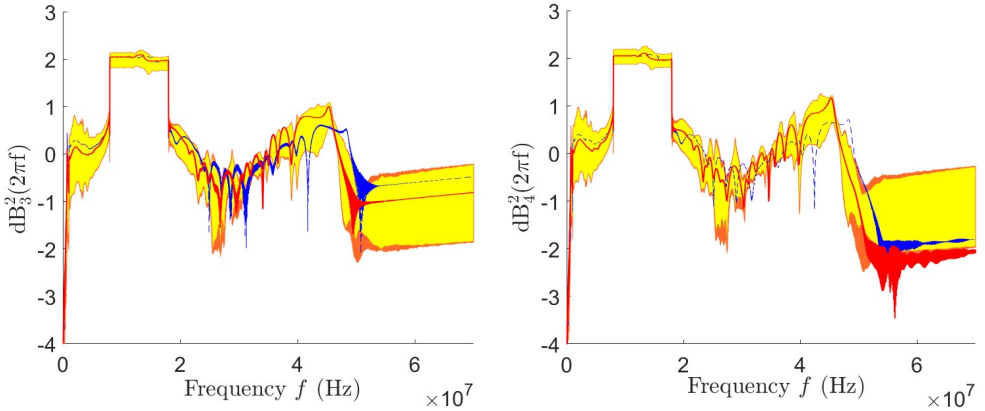
Figure 10: Observation 1: conditional confidence region of $f \mapsto \text{dB}^1_j(2\pi\text{f})$ for $\mathbf{W} = \mathbf{w}^j_{\text{ref}}$ (yellow region with orange edges), training (blue dashed line) and target (red solid thick line) corresponding to $\mathbf{w} = \mathbf{w}^j_{\text{ref}}$.

the chosen framework, which corresponds to real-world situations involving large stochastic computational models of complex systems encountered in engineering sciences. The main challenges are primarily attributed to the high



(a) Control parameter $\mathbf{w} = \mathbf{w}^1_{\mathrm{ref}} = (0.6, 0.6)$

(b) Control parameter $\mathbf{w} = \mathbf{w}^2_{\mathrm{ref}} = (0.6, 0.8)$



(c) Control parameter $\mathbf{w} = \mathbf{w}^3_{\mathrm{ref}} = (0.9, 0.6)$

(d) Control parameter $\mathbf{w} = \mathbf{w}^4_{\mathrm{ref}} = (0.9, 0.8)$

Figure 11: Observation 2: conditional confidence region of $f \mapsto \mathrm{dB}_{\mathrm{j}}^2(2\pi\mathrm{f})$ for $\mathbf{W} = \mathbf{w}^j_{\mathrm{ref}}$ (yellow region with orange edges), training (blue dashed line) and target (red solid thick line) corresponding to $\mathbf{w} = \mathbf{w}^j_{\mathrm{ref}}$.

dimensionality of the considered under-observed nonlinear uncertain computational model, its partial observability leading to incomplete data in the target dataset of the identification observations, the non-injectivity of the nonlinear mapping used to compute the identification observations from the stochastic responses of the computational model, and, above all, the limited availability of a small training dataset. The proposed approach is purely probabilistic. The surrogate model is not directly represented by an algebraic model but indirectly represented by a probability measure, with its generator facilitating the construction of a large learned dataset. We have provided an effective description of the statistical surrogate model, which allows for rapid computation of output statistics in an online context using only conditional statistics that explore the learned dataset. For any given value of the vector-valued control parameter, this statistical surrogate model provides the statistics of any observation of the stochastic computational model, including mean values, variances, probability density functions, and confidence intervals. The presented developments have been illustrated through a representative application that highlight all the aforementioned difficulties. The obtained results contribute to the validation of the proposed approach.

Acknowledgments

The authors acknowledge partial funding from DOE SciDAC FASTMath institute.

Appendix A. Formulas for conditional statistics

In this Appendix, we give the explicit formulas for computing the conditional statistics related to the predictive statistical surrogate model defined in Section 3.6(iii), related to the mean-square convergence criterion of the surrogate-model prediction defined in Section 3.7, and related to the distance of the statistical-surrogate-model predictions to the target defined in Section 3.8. We reuse the notations of these three sections, but for simplifying the writing subscript "ud" is removed. Consequently, for k fixed in $\{1,\dots,n_{\text{freq}}\}$, the N_{ud} learned realizations $\mathbf{w}_{\text{ud}}^\ell$ of $\mathbf{W}_{\text{ud}}$, for $\ell = 1,\dots,N_{\text{ud}}$, are simply rewritten as $\mathbf{w}^\ell$ for $\mathbf{W}$ with $\ell = 1,\dots,N$. Considering $\mathbf{W} = (W_1,\dots,W_{n_w})$, $\mathbf{w} = (w_1,\dots,w_{n_w})$, and $\mathbf{w}_0 = (w_{0,1},\dots,w_{0,n_w})$, for $j = 1,\dots,n_w$, let $\underline{w}_j = \frac{1}{N}\sum_{\ell=1}^N w_j^\ell$ and let $\sigma_j = \frac{1}{N-1}\sum_{\ell=1}^N (w_j^\ell - \underline{w}_j)^2$ (if $\sigma_j = 0$, then σ_j is set to 1). Let $\widetilde{W}_j = (W_j - \underline{w}_j)/\sigma_j$ whose realizations are $\tilde{w}_j^\ell = (w_j^\ell - \underline{w}_j)/\sigma_j$ for $\ell = 1,\dots,N$, and let $\tilde{w}_{0,j} = (w_{0,j} - \underline{w}_j)/\sigma_j$.

Appendix A.1. Formulas for conditional statistics of the predictive statistical surrogate model

We reuse the notations of Section 3.6(iii), but simplifying the writing, for k fixed in $\{1,\dots,n_{\text{freq}}\}$, the N_{ud} learned realizations $z_{\text{ud}}^\ell = [\mathbf{o}_{\text{ud}}^\ell]_{ik}$ of Z_{ud} for $\ell = 1,\dots,N_{\text{ud}}$, are simply rewritten as z^ℓ for Z. The associated conditional random variable $Z_{\text{ud}}(\mathbf{w}_0)$ given $\mathbf{W} = \mathbf{w}_0$ is also simply rewritten as $Z(\mathbf{w}_0)$. Let $\underline{z} = \frac{1}{N}\sum_{\ell=1}^N z^\ell$ and let $\sigma = \frac{1}{N-1}\sum_{\ell=1}^N (z^\ell - \underline{z})^2$ (if $\sigma = 0$, then σ is set to 1). Let $\widetilde{Z} = (Z - \underline{z})/\sigma$ whose realizations are $\tilde{z}^\ell = (z^\ell - \underline{z})/\sigma$ for $\ell = 1,\dots,N$. The conditional probability density function $z \mapsto p_{Z|\mathbf{W}}(z\,|\,\mathbf{w}_0)$ of Z given $\mathbf{W} = \mathbf{w}_0$ is given by

$$p_{Z|\mathbf{W}}(z\,|\,\mathbf{w}_0) = \frac{1}{\sigma\sqrt{2\pi}\,s}\frac{\sum_{\ell=1}^N \exp\left(-\frac{1}{2s^2}\left((\tilde{z}-\tilde{z}^\ell)^2 + \|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)\right)}{\sum_{\ell=1}^N \exp\left(-\frac{1}{2s^2}\|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)} \quad , \quad \tilde{z} = (z-\underline{z})/\sigma\,,$$

in which $s = (4/(N(2+n)))^{1/(n+4)}$ with $n = n_{\text{freq}} + n_w$. Let $z^* \mapsto F_{Z|\mathbf{W}}(z^*\,|\,\mathbf{w}_0)$ on $\mathbb{R}$ be the conditional cumulative distribution function of $Z(\mathbf{w}_0)$ given $\mathbf{W} = \mathbf{w}_0$, which is defined by

$$F_{Z|\mathbf{W}}(z^*\,|\,\mathbf{w}_0) = \text{Proba}\{Z \le z^*\,|\,\mathbf{w}_0\} = \int_{-\infty}^{z^*} p_{Z|\mathbf{W}}(z\,|\,\mathbf{w}_0)\,dz\,.$$

We then have,

$$F_{Z|\mathbf{W}}(z^*\,|\,\mathbf{w}_0) = \frac{\sum_{\ell=1}^N \widetilde{F}^\ell(\tilde{z}^*)\,\exp\left(-\frac{1}{2s^2}\|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)}{\sum_{\ell=1}^N \exp\left(-\frac{1}{2s^2}\|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)} \quad , \quad \tilde{z}^* = (z^* - \underline{z})/\sigma\,,$$

in which $\widetilde{F}^\ell(\tilde{z}^*)$ is written as

$$\widetilde{F}^\ell(\tilde{z}^*) = \frac{1}{2} + \frac{1}{2}\,\text{erf}\left(\frac{1}{s\sqrt{2}}(\tilde{z}^* - \tilde{z}^\ell)\right)\,,$$

with $\text{erf}(y) = (2/\sqrt{\pi})\int_0^y e^{-t^2}\,dt$. For each k fixed in $\{1,\dots,n_{\text{freq}}\}$, the conditional upper bound $z^+(\mathbf{w}_0) = \text{Proba}\{Z \le z^+(\mathbf{w}_0)\,|\,\mathbf{w}_0\}$ and the conditional lower bound $z^-(\mathbf{w}_0) = \text{Proba}\{Z \le z^-(\mathbf{w}_0)\,|\,\mathbf{w}_0\}$ of the confidence interval $[z^-(\mathbf{w}_0)\,, z^+(\mathbf{w}_0)]$ of the random variable $Z(\mathbf{w}_0)$ given $\mathbf{W} = \mathbf{w}_0$, for a given probability level $p_c \in]0,1[$ (for instance, $p_c = 0.98$) are computed by solving the equations,

$$z^+(\mathbf{w}_0) = \text{arg}_z\{F_{Z|\mathbf{W}}(z\,|\,\mathbf{w}_0) = p_c\} \quad , \quad z^-(\mathbf{w}_0) = \text{arg}_z\{F_{Z|\mathbf{W}}(z\,|\,\mathbf{w}_0) = 1 - p_c\}\,.$$

Appendix A.2. Formulas for mean-square convergence criterion of the surrogate-model prediction

We reuse the notations of Section 3.7 but again as in Appendix Appendix A.1, for simplifying the writing we define $\mathbf{Q} = (Q_1,\dots,Q_{n_{\text{freq}}})$ with $Q_k = [\mathbf{O}_{\text{ud}}]_{i_o k}$ for $k = 1,\dots,n_{\text{freq}}$. Therefore, the N learned realizations of $\mathbf{Q}$ are denoted by $\mathbf{q}^\ell$ for $\ell = 1,\dots,N$. The associated conditional random variable of $\mathbf{Q}$ given $\mathbf{W} = \mathbf{w}_0$ is written as $\mathbf{Q}(\mathbf{w}_0)$. Let $\underline{\mathbf{q}} = \frac{1}{N}\sum_{\ell=1}^N \mathbf{q}^\ell$ and let $\sigma_{q,k} = \frac{1}{N-1}\sum_{\ell=1}^N (q_k^\ell - \underline{q}_k)^2$ (if $\sigma_{q,k} = 0$, then $\sigma_{q,k}$ is set to 1). Let $\widetilde{Q}_k = (Q_k - \underline{q}_k)/\sigma_{q,k}$ whose realizations are $\tilde{q}_k^\ell = (q_k^\ell - \underline{q}_k)/\sigma_{q,k}$ for $\ell = 1,\dots,N$ and $k = 1,\dots,n_{\text{freq}}$. The square of the coefficient of variation $\text{cv}(\mathbf{w}_0, i_o)^2$ defined by Eq. (3.15) is written as

$$\text{cv}(\mathbf{w}_0, i_o)^2 = \frac{\sum_{k=1}^{n_{\text{freq}}}\left(E\{(Q_k)^2\,|\,\mathbf{W} = \mathbf{w}_0\} - (E\{Q_k\,|\,\mathbf{W} = \mathbf{w}_0\})^2\right)}{\sum_{k=1}^{n_{\text{freq}}}(E\{Q_k\,|\,\mathbf{W} = \mathbf{w}_0\})^2}\,,$$

in which

$$E\{Q_k \,|\, \mathbf{W} = \mathbf{w}_0\} = \underline{q}_k + \sigma_{q,k}\, E\{\widetilde{Q}_k \,|\, \widetilde{\mathbf{W}} = \tilde{\mathbf{w}}_0\}$$
$$E\{(Q_k)^2 \,|\, \mathbf{W} = \mathbf{w}_0\} = (\underline{q}_k)^2 + 2\,\underline{q}_k\,\sigma_{q,k}\, E\{\widetilde{Q}_k \,|\, \widetilde{\mathbf{W}} = \tilde{\mathbf{w}}_0\} + (\sigma_{q,k})^2\, E\{(\widetilde{Q}_k)^2 \,|\, \widetilde{\mathbf{W}} = \tilde{\mathbf{w}}_0\}\,,$$

and where for $\alpha = 1$ or 2,

$$E\{(\widetilde{Q}_k)^\alpha \,|\, \widetilde{\mathbf{W}} = \tilde{\mathbf{w}}_0\} = \gamma_\alpha + \frac{\sum_{\ell=1}^N (\tilde{q}_k^\ell)^\alpha \, \exp\left(-\frac{1}{2s^2}\|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)}{\sum_{\ell=1}^N \exp\left(-\frac{1}{2s^2}\|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)}\,,$$

in which $s = (4/(N(2+n)))^{1/(n+4)}$ with $n = n_{\text{freq}} + n_w$ and where $\gamma_1 = 0$ and $\gamma_2 = s^2$.

Appendix A.3. Formulas for the distance of the statistical-surrogate-model predictions to the target

We reuse the notations of Section 3.8, removing subscript i_o when no confusion is possible. Let $\underline{r} = \frac{1}{N}\sum_{\ell=1}^N r_{i_o}^\ell$ and let $\sigma = \frac{1}{N-1}\sum_{\ell=1}^N (r_{i_o}^\ell - \underline{r})^2$ (if $\sigma = 0$, then σ is set to 1). Let $\widetilde{R}_{i_o} = (R_{i_o} - \underline{r})/\sigma$ whose realizations are $\tilde{r}_{i_o}^\ell = (r_{i_o}^\ell - \underline{r})/\sigma$ for $\ell = 1, \ldots, N$. The conditional probability density function $r \mapsto p_{R_{i_o}\,|\,\mathbf{W}}(r \,|\, \mathbf{w}_0)$ of R_{i_o} given $\mathbf{W} = \mathbf{w}_0$ is written as

$$p_{R_{i_o}\,|\,\mathbf{W}}(r \,|\, \mathbf{w}_0) = \frac{1}{\sigma\sqrt{2\pi}\,s} \frac{\sum_{\ell=1}^N \exp\left(-\frac{1}{2s^2}\left((\tilde{r} - \tilde{r}_{i_o}^\ell)^2 + \|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)\right)}{\sum_{\ell=1}^N \exp\left(-\frac{1}{2s^2}\|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)} \quad , \quad \tilde{r} = (r - \underline{r})/\sigma\,,$$

in which $s = (4/(N(2+n)))^{1/(n+4)}$ with $n = 1 + n_w$. The mean-square distance $d_{\text{ms}}(\mathbf{w}_0, i_o) = (E\{R_{i_o}(\mathbf{w}_0)\})^{1/2}$ defined by Eq. (3.21) is such that

$$d_{\text{ms}}(\mathbf{w}_0, i_o)^2 = \frac{\sum_{\ell=1}^N \tilde{r}_{i_o}^\ell \, \exp\left(-\frac{1}{2s^2}\|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)}{\sum_{\ell=1}^N \exp\left(-\frac{1}{2s^2}\|\tilde{\mathbf{w}}_0 - \tilde{\mathbf{w}}^\ell\|^2\right)}\,.$$

Conflict of interest

The authors declare that he has no conflict of interest.

References

[1] N. Queipo, R. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, K. Tucker, Surrogate-based analysis and optimization, Progress in Aerospace Science 41 (1) (2005) 1–28. doi:10.1016/j.paerosci.2005.02.001.

[2] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, K. Crombecq, A surrogate modeling and adaptive sampling toolbox for computer based design, Journal of Machine Learning Research 11 (68) (2010) 2051–2055.

[3] R. Alizadeh, J. K. Allen, F. Mistree, Managing computational complexity using surrogate models: a critical review, Research in Engineering Design 31 (3) (2020) 275–298. doi:10.1007/s00163-020-00336-7.

[4] J. L. Peixoto, A property of well-formulated polynomial regression models, The American Statistician 44 (1) (1990) 26–30. doi:10.2307/2684952.

[5] E. Ostertagová, Modelling using polynomial regression, Procedia Engineering 48 (2012) 500–506. doi:10.1016/j.proeng.2012.09.545.

[6] J. P. Kleijnen, Regression and kriging metamodels with their experimental designs in simulation: a review, European Journal of Operational Research 256 (1) (2017) 1–16. doi:10.1016/j.ejor.2016.06.041.

[7] J. P. Kleijnen, Kriging metamodeling in simulation: A review, European Journal of Operational Research 192 (3) (2009) 707–716. doi:10.1016/j.ejor.2007.10.013.

[8] V. Dubourg, B. Sudret, J.-M. Bourinet, Reliability-based design optimization using kriging surrogates and subset simulation, Structural and Multidisciplinary Optimization 44 (5) (2011) 673–690. doi:10.1007/s00158-011-0653-8.

[9] P. Kersaudy, B. Sudret, N. Varsier, O. Picon, J. Wiart, A new surrogate modeling technique combining kriging and polynomial chaos expansions–application to uncertainty analysis in computational dosimetry, Journal of Computational Physics 286 (2015) 103–117.

[10] J. Qian, J. Yi, Y. Cheng, J. Liu, Q. Zhou, A sequential constraints updating approach for kriging surrogate model-assisted engineering optimization design problem, Engineering with Computers 36 (3) (2020) 993–1009.

[11] Y. Zhou, Z. Lu, J. Hu, Y. Hu, Surrogate modeling of high-dimensional problems via data-driven polynomial chaos expansions and sparse partial least square, Computer Methods in Applied Mechanics and Engineering 364 (2020) 112906. doi:10.1016/j.cma.2020.112906.

[12] R. Ghanem, P. D. Spanos, Stochastic Finite Elements: a Spectral Approach, Springer-Verlag, New York, 1991.

[13] C. Soize, R. Ghanem, Physical systems with random uncertainties: chaos representations with arbitrary probability measure, SIAM Journal on Scientific Computing 26 (2) (2004) 395–410. doi:10.1137/S1064827503424505.

[14] A. Doostan, R. Ghanem, J. Red-Horse, Stochastic model reduction for chaos representations, Computer Methods in Applied Mechanics and Engineering 196 (37-40) (2007) 3951–3966. doi:10.1016/j.cma.2006.10.047.
[15] S. Das, R. Ghanem, J. C. Spall, Asymptotic sampling distribution for polynomial chaos representation from data: a maximum entropy and fisher information approach, SIAM Journal on Scientific Computing 30 (5) (2008) 2207–2234. doi:10.1137/060652105.
[16] C. Soize, R. Ghanem, Reduced chaos decomposition with random coefficients of vector-valued random variables and random fields, Computer Methods in Applied Mechanics and Engineering 198 (21-26) (2009) 1926–1934. doi:10.1016/j.cma.2008.12.035.
[17] C. Soize, C. Desceliers, Computational aspects for constructing realizations of polynomial chaos in high dimension, SIAM Journal on Scientific Computing 32 (5) (2010) 2820–2831. doi:10.1137/100787830.
[18] O. G. Ernst, A. Mugler, H.-J. Starkloff, E. Ullmann, On the convergence of generalized polynomial chaos expansions, ESAIM: Mathematical Modelling and Numerical Analysis 46 (2) (2012) 317–339. doi:10.1051/m2an/2011045.
[19] R. Tipireddy, R. Ghanem, Basis adaptation in homogeneous chaos spaces, Journal of Computational Physics 259 (2014) 304–317. doi:10.1016/j.jcp.2013.12.009.
[20] C. Soize, Polynomial chaos expansion of a multimodal random vector, SIAM-ASA Journal on Uncertainty Quantification 3 (1) (2015) 34–60. doi:10.1137/140968495.
[21] S. Abraham, M. Raisee, G. Ghorbanias, F. Contino, C. Lacor, A robust and efficient stepwise regression method for building sparse polynomial chaos expansions, Journal of Computational Physics 332 (2017) 461–474. doi:10.1016/j.jcp.2016.12.015.
[22] C. Thimmisetty, P. Tsilifis, R. Ghanem, Homogeneous chaos basis adaptation for design optimization under uncertainty: Application to the oil well placement problem, Artificial Intelligence for Engineering Design, Analysis and Manufacturing 31 (3 (Uncertainty Quantification for Engineering Design)) (2017) 265–276. doi:10.1017/S0890060417000166.
[23] C. Soize, R. Ghanem, Polynomial chaos representation of databases on manifolds, Journal of Computational Physics 335 (2017) 201–221. doi:10.1016/j.jcp.2017.01.031.
[24] C. Desceliers, R. Ghanem, C. Soize, Maximum likelihood estimation of stochastic chaos representations from experimental data, International Journal for Numerical Methods in Engineering 66 (6) (2006) 978–1001. doi:10.1002/nme.1576.
[25] Y. M. Marzouk, H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems, Journal of Computational Physics 228 (6) (2009) 1862–1902. doi:10.1016/j.jcp.2008.11.024.
[26] M. Arnst, R. Ghanem, C. Soize, Identification of Bayesian posteriors for coefficients of chaos expansions, Journal of Computational Physics 229 (9) (2010) 3134–3154. doi:10.1016/j.jcp.2009.12.033.
[27] C. Soize, Identification of high-dimension polynomial chaos expansions with random coefficients for non-Gaussian tensor-valued random fields using partial and limited experimental data, Computer methods in applied mechanics and engineering 199 (33-36) (2010) 2150–2164. doi:10.1016/j.cma.2010.03.013.
[28] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Identification of polynomial chaos representations in high dimension from a set of realizations, SIAM Journal on Scientific Computing 34 (6) (2012) A2917–A2945. doi:10.1137/11084950X.
[29] B. V. Rosić, A. Litvinenko, O. Pajonk, H. G. Matthies, Sampling-free linear Bayesian update of polynomial chaos representations, Journal of Computational Physics 231 (17) (2012) 5761–5787. doi:10.1016/j.jcp.2012.04.044.
[30] R. Madankan, P. Singla, T. Singh, P. D. Scott, Polynomial-chaos-based Bayesian approach for state and parameter estimations, Journal of Guidance, Control, and Dynamics 36 (4) (2013) 1058–1074. doi:10.2514/1.58377.
[31] B. Chen-Charpentier, D. Stanescu, Parameter estimation using polynomial chaos and maximum likelihood, International Journal of Computer Mathematics 91 (2) (2014) 336–346. doi:10.1080/00207160.2013.809069.
[32] A. H. Elsheikh, I. Hoteit, M. F. Wheeler, Efficient bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates, Computer Methods in Applied Mechanics and Engineering 269 (2014) 515–537. doi:10.1016/j.cma.2013.11.001.
[33] L. Giraldi, O. P. Le Maître, K. T. Mandli, C. N. Dawson, I. Hoteit, O. M. Knio, Bayesian inference of earthquake parameters from buoy data using a polynomial chaos-based surrogate, Computational Geosciences 21 (4) (2017) 683–699.
[34] P. Tsilifis, R. Ghanem, Bayesian adaptation of chaos representations using variational inference and sampling on geodesics, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 474 (2217) (2018) 20180285. doi:10.1098/rspa.2018.0285.
[35] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, Comptes Rendus Mécanique 336 (6) (2008) 518–523. doi:10.1016/j.crme.2008.02.013.
[36] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, Journal of Computational Physics 230 (6) (2011) 2345–2367. doi:10.1016/j.jcp.2010.12.021.
[37] Q. Shao, A. Younes, M. Fahs, T. A. Mara, Bayesian sparse polynomial chaos expansion for global sensitivity analysis, Computer Methods in Applied Mechanics and Engineering 318 (2017) 474–496. doi:10.1016/j.cma.2017.01.033.
[38] N. Luthen, S. Marelli, B. Sudret, Sparse polynomial chaos expansions: Literature survey and benchmark, SIAM/ASA Journal on Uncertainty Quantification 9 (2) (2021) 593–649. doi:10.1137/20M1315774.
[39] P. Tsilifis, R. Ghanem, Reduced Wiener chaos representation of random fields via basis adaptation and projection, Journal of Computational Physics 341 (2017) 102–120. doi:10.1016/j.jcp.2017.04.009.
[40] Z. Liu, D. Lesselier, B. Sudret, J. Wiart, Surrogate modeling based on resampled polynomial chaos expansions, Reliability Engineering & System Safety 202 (2020) 107008. doi:10.1016/j.ress.2020.107008.
[41] K. Kontolati, D. Loukrezis, K. R. dos Santos, D. G. Giovanis, M. D. Shields, Manifold learning-based polynomial chaos expansions for high-dimensional surrogate models, International Journal for Uncertainty Quantification 12 (4) (2022). doi:10.1615/Int.J.UncertaintyQuantification.2022039936.
[42] J.-L. Akian, L. Bonnet, H. Owhadi, É. Savin, Learning best kernels from data in gaussian process regression. with application to aerodynamics, Journal of Computational Physics 470 (2022) 111595. doi:10.1016/j.jcp.2022.111595.
[43] A. Nouy, O. P. L. Maitre, Generalized spectral decomposition for stochastic nonlinear problems, Journal of Computational Physics 228 (1) (2009) 202–235.
[44] A. Nouy, Low-rank tensor methods for model order reduction, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), Handbook of Uncertainty Quantification, Springer, Cham, Switzerland, 2017, Ch. 25, pp. 857–882.

[45] Z. Ghahramani, M. I. Jordan, Learning from incomplete data, MIT A.I. Memo 1509 (C.B.C.L. Paper No. 108) (1994) 1–12.
[46] O. Harel, The estimation of r 2 and adjusted r 2 in incomplete data sets using multiple imputation, Journal of Applied Statistics 36 (10) (2009) 1109–1118. doi:10.1080/02664760802553000.
[47] I. A. Gheyas, L. S. Smith, A neural network-based framework for the reconstruction of incomplete data sets, Neurocomputing 73 (16-18) (2010) 3039–3065. doi:10.1016/j.neucom.2010.06.021.
[48] M. Hittawe, S. Langodan, O. Beya, I. Hoteit, O. Knio, Efficient sst prediction in the red sea using hybrid deep learning-based approach, in: 2022 IEEE 20th International Conference on Industrial Informatics (INDIN), IEEE, 2022, pp. 107–117. doi:10.1109/INDIN51773.2022.9976090.
[49] Z. Xue, H. Wang, Effective density-based clustering algorithms for incomplete data, Big Data Mining and Analytics 4 (3) (2021) 183–194.
[50] W. Jiang, M. Bogdan, J. Josse, S. Majewski, B. Miasojedow, V. Ročková, T. Group, Adaptive bayesian SLOPE: model selection with incomplete data, Journal of Computational and Graphical Statistics 31 (1) (2022) 113–137. doi:10.1080/10618600.2021.1963263.
[51] W. Stacklies, H. Redestig, M. Scholz, D. Walther, J. Selbig, pcamethods - a bioconductor package providing pca methods for incomplete data, Bioinformatics 23 (9) (2007) 1164–1167. doi:10.1093/bioinformatics/btm069.
[52] J. Podani, T. Kalapos, B. Barta, D. Schmera, Principal component analysis of incomplete data–a simple solution to an old problem, Ecological Informatics 61 (2021) 101235. doi:10.1016/j.ecoinf.2021.101235.
[53] C. F. Caiafa, J. Solé-Casals, P. Marti-Puig, S. Zhe, T. Tanaka, Decomposition methods for machine learning with small, incomplete or noisy datasets, Applied Sciences 10 (23) (2020) 8481. doi:10.3390/app10238481.
[54] C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, Subspace estimation from unbalanced and incomplete data matrices: l2,8 statistical guarantees, The Annals of Statistics 49 (2) (2021) 944–967. doi:10.1214/20-AOS1986.
[55] M. A. Grepl, Y. Maday, N. C. Nguyen, A. T. Patera, Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations, ESAIM: Mathematical Modelling and Numerical Analysis 41 (3) (2007) 575–605. doi:10.1051/m2an:2007031.
[56] S. Chaturantabut, D. C. Sorensen, Nonlinear model reduction via discrete empirical interpolation, SIAM Journal on Scientific Computing 32 (5) (2010) 2737–2764.
[57] K. Carlberg, C. Bou-Mosleh, C. Farhat, 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) (2011) 155–181. doi:10.1002/nme.3050.
[58] K. Carlberg, C. Farhat, J. Cortial, D. Amsallem, The gnat method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows, Journal of Computational Physics 242 (2013) 623–647. doi:10.1016/j.jcp.2013.02.028.
[59] C. Farhat, P. Avery, T. Chapman, J. Cortial, Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting for computational efficiency, International Journal for Numerical Methods in Engineering 98 (9) (2014) 625–662. doi:10.1002/nme.4668.
[60] P. Benner, S. Gugercin, K. Willcox, A survey of projection-based model reduction methods for parametric dynamical systems, SIAM review 57 (4) (2015) 483–531. doi:10.1137/130932715.
[61] C. Farhat, T. Chapman, P. Avery, 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) (2015) 1077–1110. doi:10.1002/nme.4820.
[62] A. Paul-Dubois-Taine, D. Amsallem, An adaptive and efficient greedy procedure for the optimal training of parametric reduced-order models, International Journal for Numerical Methods in Engineering 102 (5) (2015) 1262–1292. doi:10.1002/nme.4759.
[63] C. Farhat, S. Grimberg, A. Manzoni, A. Quarteroni, Computational bottlenecks for PROMS: precomputation and hyperreduction, Model order reduction, Berlin: De Gruyter (2021) 181–244doi:10.1515/9783110671490-005.
[64] D. Jones, C. Snider, A. Nassehi, J. Yon, B. Hicks, Characterising the digital twin: A systematic literature review, CIRP Journal of Manufacturing Science and Technology 29 (2020) 36–52. doi:10.1016/j.cirpj.2020.02.002.
[65] R. Ghanem, C. Soize, L. Mehrez, V. Aitharaju, Probabilistic learning and updating of a digital twin for composite material systems, International Journal for Numerical Methods in Engineering 123 (13) (2022) 3004–3020. doi:10.1002/nme.6430.
[66] C. Soize, C. Farhat, A nonparametric probabilistic approach for quantifying uncertainties in low-dimensional and high-dimensional nonlinear models, International Journal for Numerical Methods in Engineering 109 (6) (2017) 837–888. doi:10.1002/nme .5312.
[67] C. Farhat, A. Bos, P. Avery, C. Soize, Modeling and quantification of model-form uncertainties in eigenvalue computations using a stochastic reduced model, AIAA Journal 56 (3) (2018) 1198–1210. doi:10.2514/1.J056314.
[68] C. Farhat, R. Tezaur, T. Chapman, P. Avery, C. Soize, Feasible probabilistic learning method for model-form uncertainty quantification in vibration analysis, AIAA Journal 57 (11) (2019) 4978–4991. doi:10.2514/1.J057797.
[69] H. Wang, J. Guilleminot, C. Soize, Modeling uncertainties in molecular dynamics simulations using a stochastic reduced-order basis, Computer Methods in Applied Mechanics and Engineering 354 (2019) 37–55.
[70] C. Soize, C. Farhat, Probabilistic learning for modeling and quantifying model-form uncertainties in nonlinear computational mechanics, International Journal for Numerical Methods in Engineering 117 (2019) 819–843. doi:10.1002/nme.5980.
[71] M.-J. Azzi, C. Ghnatios, P. Avery, C. Farhat, Acceleration of a physics-based machine learning approach for modeling and quantifying model-form uncertainties and performing model updating, Journal of Computing and Information Science in Engineering 23 (1) (2022) 011009. doi:10.1115/1.4055546.
[72] H. Zhang, J. Guilleminot, A riemannian stochastic representation for quantifying model uncertainties in molecular dynamics simulations, Computer Methods in Applied Mechanics and Engineering 403 (2023) 115702.
[73] A. Talwalkar, S. Kumar, H. Rowley, Large-scale manifold learning, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8. doi:10.1109/CVPR.2008.4587670.
[74] A. C. Öztireli, M. Alexa, M. Gross, Spectral sampling of manifolds, ACM Transactions on Graphics (TOG) 29 (6) (2010) 1–8. doi:10.1145/1882261.1866190.
[75] Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: An introduction, Handbook of uncertainty quantification (2016) 1–41doi:10.1007/978-3-319-11259-6_23-1.
[76] M. D. Parno, Y. M. Marzouk, Transport map accelerated markov chain Monte Carlo, SIAM/ASA Journal on Uncertainty Quantification

6 (2) (2018) 645–682. doi:10.1137/17M1134640.
[77] G. Perrin, C. Soize, N. Ouhbi, Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints, Computational Statistics & Data Analysis 119 (2018) 139–154. doi:10.1016/j.csda.2017.10.005.
[78] Y. Kevrekidis, Manifold learning for parameter reduction, Bulletin of the American Physical Society 65 (2020). doi:10.1016/j.jcp.2019.04.015.
[79] S. Pan, K. Duraisamy, Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability, SIAM Journal on Applied Dynamical Systems 19 (1) (2020) 480–509. doi:10.1137/19M1267246.
[80] I. Kalogeris, V. Papadopoulos, Diffusion maps-based surrogate modeling: An alternative machine learning approach, International Journal for Numerical Methods in Engineering 121 (4) (2020) 602–620. doi:10.1002/nme.6236.
[81] C. Soize, R. Ghanem, Data-driven probability concentration and sampling on manifold, Journal of Computational Physics 321 (2016) 242–258. doi:10.1016/j.jcp.2016.05.044.
[82] C. Soize, R. Ghanem, C. Safta, X. Huan, Z. P. Vane, J. C. Oefelein, G. Lacaze, H. N. Najm, Q. Tang, X. Chen, Entropy-based closure for probabilistic learning on manifolds, Journal of Computational Physics 388 (2019) 528–533. doi:10.1016/j.jcp.2018.12.029.
[83] C. Soize, R. Ghanem, C. Desceliers, Sampling of Bayesian posteriors with a non-Gaussian probabilistic learning on manifolds from a small dataset, Statistics and Computing 30 (5) (2020) 1433–1457. doi:10.1007/s11222-020-09954-6.
[84] C. Soize, R. Ghanem, Probabilistic learning on manifolds, Foundations of Data Science 2 (3) (2020) 279–307. doi:10.3934/fods.2020013.
[85] C. Soize, R. Ghanem, Physics-constrained non-Gaussian probabilistic learning on manifolds, International Journal for Numerical Methods in Engineering 121 (1) (2020) 110–145. doi:10.1002/nme.6202.
[86] C. Soize, R. Ghanem, Probabilistic learning on manifolds constrained by nonlinear partial differential equations for small datasets, Computer Methods in Applied Mechanics and Engineering 380 (2021) 113777. doi:10.1016/j.cma.2021.113777.
[87] C. Soize, Probabilistic learning inference of boundary value problem with uncertainties based on Kullback-Leibler divergence under implicit constraints, Computer Methods in Applied Mechanics and Engineering 395 (2022) 115078. doi:10.1016/j.cma.2022.115078.
[88] C. Soize, R. Ghanem, Probabilistic learning on manifolds (PLoM) with partition, International Journal for Numerical Methods in Engineering 123 (1) (2022) 268–290. doi:10.1002/nme.6856.
[89] C. Soize, Probabilistic learning constrained by realizations using a weak formulation of fourier transform of probability measures, Computational Statistics (2022) 1–30doi:10.1007/s00180-022-01300-w.
[90] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464. doi:10.1111/1467-9868.00294.
[91] Y. M. Marzouk, H. N. Najm, L. A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, Journal of Computational Physics 224 (2) (2007) 560–586. doi:10.1016/j.jcp.2006.10.010.
[92] J. E. Gentle, Computational statistics, Springer, New York, 2009. doi:10.1007/978-0-387-98144-4.
[93] A. M. Stuart, Inverse problems: a Bayesian perspective, Acta Numerica 19 (2010) 451–559. doi:10.1017/S0962492910000061.
[94] H. Owhadi, C. Scovel, T. Sullivan, On the brittleness of Bayesian inference, SIAM Review 57 (4) (2015) 566–582. doi:10.1137/130938633.
[95] H. G. Matthies, E. Zander, B. V. Rosić, A. Litvinenko, O. Pajonk, Inverse problems in a Bayesian setting, in: Computational Methods for Solids and Fluids, Vol. 41, Springer, 2016, pp. 245–286. doi:10.1007/978-3-319-27996-1_10.
[96] M. Dashti, A. M. Stuart, The Bayesian approach to inverse problems, in: R. Ghanem, D. Higdon, O. Houman (Eds.), Handbook of Uncertainty Quantification, Springer, Cham, Switzerland, 2017, Ch. 10, pp. 311–428. doi:10.1007/978-3-319-12385-1_7.
[97] R. Ghanem, D. Higdon, H. Owhadi, Handbook of Uncertainty Quantification, Vol. 1 to 3, Springer, Cham, Switzerland, 2017. doi:10.1007/978-3-319-12385-1.
[98] A. Spantini, T. Cui, K. Willcox, L. Tenorio, Y. Marzouk, Goal-oriented optimal approximations of Bayesian linear inverse problems, SIAM Journal on Scientific Computing 39 (5) (2017) S167–S196. doi:10.1137/16M1082123.
[99] G. Perrin, C. Soize, Adaptive method for indirect identification of the statistical properties of random fields in a Bayesian framework, Computational Statistics 35 (1) (2020) 111–133. doi:10.1007/s00180-019-00936-5.
[100] R. Ghanem, C. Soize, Probabilistic nonconvex constrained optimization with fixed number of function evaluations, International Journal for Numerical Methods in Engineering 113 (4) (2018) 719–741. doi:10.1002/nme.5632.
[101] C. Soize, Design optimization under uncertainties of a mesoscale implant in biological tissues using a probabilistic learning algorithm, Computational Mechanics 62 (3) (2018) 477–497. doi:10.1007/s00466-017-1509-x.
[102] R. Ghanem, C. Soize, C. Thimmisetty, Optimal well-placement using probabilistic learning, Data-Enabled Discovery and Applications 2 (1) (2018) 1–16. doi:10.1007/s41688-017-0014-x.
[103] R. Ghanem, C. Soize, C. Safta, X. Huan, G. Lacaze, J. C. Oefelein, H. N. Najm, Design optimization of a scramjet under uncertainty using probabilistic learning on manifolds, Journal of Computational Physics 399 (2019) 108930. doi:10.1016/j.jcp.2019.108930.
[104] E. Capiez-Lernout, C. Soize, Nonlinear stochastic dynamics of detuned bladed disks with uncertain mistuning and detuning optimization using a probabilistic machine learning tool, International Journal of Non-Linear Mechanics 143 (2022) 104023. doi:10.1016/j.ijnonlinmec.2022.104023.
[105] J. O. Almeida, F. A. Rochinha, A probabilistic learning approach applied to the optimization of wake steering in wind farms, Journal of Computing and Information Science in Engineering 23 (1) (2023) 011003. doi:10.1115/1.4054501.
[106] J. Guilleminot, J. E. Dolbow, Data-driven enhancement of fracture paths in random composites, Mechanics Research Communications 103 (2020) 103443. doi:10.1016/j.mechrescom.2019.103443.
[107] M. Arnst, C. Soize, K. Bulthies, Computation of sobol indices in global sensitivity analysis from small data sets by probabilistic learning on manifolds, International Journal for Uncertainty Quantification 11 (2) (2021) 1–23. doi:10.1615/Int.J.UncertaintyQuantification.2020032674.
[108] C. Soize, A. Orcesi, Machine learning for detecting structural changes from dynamic monitoring using the probabilistic learning on manifolds, Structure and Infrastructure Engineering Journal 17 (10) (2021) 1418–1430. doi:10.1080/15732479.2020.1811991.
[109] K. Zhong, J. G. Navarro, S. Govindjee, G. G. Deierlein, Surrogate modeling of structural seismic response using Probabilistic Learning on Manifolds, Earthquake Engineering and Structural Dynamics 52 (8) (2023) 2407–2428. doi:10.1002/eqe.3839.

[110] J. O. Almeida, F. A. Rochinha, Uncertainty quantification of waterflooding in oil reservoirs computational simulations using a probabilistic learning approach, Journal of Computing and Information Science in Engineering 13 (4) (2023) 1–22. doi:10.1615/Int.J.UncertaintyQuantification.2023041042.
[111] G. H. Golub, C. F. Van Loan, Matrix Computations, Second Edition, Johns Hopkins University Press, Baltimore and London, 1993.
[112] A. Bowman, A. Azzalini, Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-Plus Illustrations, Vol. 18, Oxford University Press, Oxford: Clarendon Press, New York, 1997. doi:10.1007/s001800000033.
[113] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86. doi:10.1214/aoms/1177729694.
[114] J. N. Kapur, H. K. Kesavan, Entropy Optimization Principles with Applications, Academic Press, San Diego, 1992.
[115] T. M. Cover, J. A. Thomas, Elements of Information Theory, Second Edition, John Wiley & Sons, Hoboken, 2006.
[116] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Vol. 160, Springer Science & Business Media, 2005. doi:10.1007/b138659.
[117] C. Robert, G. Casella, Monte Carlo Statistical Methods, Springer Science & Business Media, 2005. doi:10.1007/978-1-4757-4145-2.
[118] J. C. Spall, Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, Vol. 65, John Wiley & Sons, 2005.
[119] C. Soize, 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) (2008) 1583–1611. doi:10.1002/nme.2385.
[120] C. Soize, The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions, Vol. Series on Advances in Mathematics for Applied Sciences: Vol 17, World Scientific, Singapore, 1994. doi:10.1142/2347.
[121] R. Neal, MCMC using hamiltonian dynamics, in: S. Brooks, A. Gelman, G. Jones, X.-L. Meng (Eds.), Handbook of Markov Chain Monte Carlo, Chapman and Hall-CRC Press, Boca Raton, 2011, Ch. 5, pp. 1–51. doi:10.1201/b10905-6.
[122] M. Girolami, B. Calderhead, Riemann manifold Langevin and Hamiltonian Monte Carlo methods, Journal of the Royal Statistics Society 73 (2) (2011) 123–214. doi:10.1111/j.1467-9868.2010.00765.x.
[123] M. A. Hopcroft, W. D. Nix, T. W. Kenny, What is the young's modulus of silicon?, Journal of microelectromechanical systems 19 (2) (2010) 229–238. doi:10.1109/JMEMS.2009.2039697.
[124] J. Guilleminot, C. Soize, On the statistical dependence for the components of random elasticity tensors exhibiting material symmetry properties, Journal of Elasticity 111 (2) (2013) 109–130. doi:10.1007/s10659-012-9396-z.
[125] C. Soize, Non Gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators, Computer Methods in Applied Mechanics and Engineering 195 (1-3) (2006) 26–64. doi:10.1016/j.cma.2004.12.014.
[126] C. Soize, Uncertainty Quantification. An Accelerated Course with Advanced Applications in Computational Engineering, Springer, New York, 2017. doi:10.1007/978-3-319-54339-0.