



**HAL**  
open science

# Aeroacoustic Liner Impedance Metamodel from Simulation and Experimental Data Using Probabilistic Learning

Amritesh Sinha, Christian Soize, Christophe Desceliers, Guilherme Coelho-Cunha

## ► To cite this version:

Amritesh Sinha, Christian Soize, Christophe Desceliers, Guilherme Coelho-Cunha. Aeroacoustic Liner Impedance Metamodel from Simulation and Experimental Data Using Probabilistic Learning. *AIAA Journal*, 2023, 61 (11), pp.4926-4934. 10.2514/1.J062991 . hal-04182265

**HAL Id: hal-04182265**

**<https://univ-eiffel.hal.science/hal-04182265v1>**

Submitted on 19 Aug 2023

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

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

# Aeroacoustic Liner Impedance Metamodel from Simulation and Experimental Data Using Probabilistic Learning

Amritesh Sinha <sup>\*</sup>, Christian Soize <sup>†</sup> and Christophe Desceliers <sup>‡</sup>  
*Gustave Eiffel University, MSME UMR 8208 (Marne-la-Vallée, France)*

Guilherme Cunha <sup>§</sup>  
*Airbus Operations SAS (Toulouse, France)*

**In order to be able to calculate a parameterized aeroacoustic liner impedance, a robust statistical metamodel is constructed as a function of the frequency and of the control parameters that are the percentage of open area and the sound pressure level. This construction is based on the use of simulated data generated with a computationally expensive aeroacoustic model, which translates to a very small training dataset. This means that the learning process has to be used and the probabilistic learning on manifolds algorithm is chosen. Although the aeroacoustic simulation is conducted on a large aeroacoustic computational model, some approximations are introduced, generating model errors that are taken into account by a probability model in the constructed training dataset. This probability model is calibrated using dimensionless experiments available from the open literature. Despite the fact that only a small amount of data is available, a novel statistical metamodel is successfully developed for which the predictions are consistent. This statistical framework allows for exhibiting a confidence region of the parameterized aeroacoustic liner impedance, which gives an information about the level of uncertainties as a function of the frequency and the control parameters.**

## I. Introduction

**A**ir traffic has significantly grown over the past decade. With ever increasing environmental concerns, green aviation has come into prominence. In particular, aircraft noise is of significant interest to the aviation community, whether it be the question of external or internal noise relative to the aircraft. This has led to increasingly stringent requirements by the certification authorities. In modern turbofan engines (UHBR - Ultra high bypass ratio), fan noise is one of the main contributors to the overall aircraft noise. Fan noise can be characterized by broadband and tonal noise components. Acoustic liners (acoustic treatments) can be designed to tackle both the components. Tonal noise is mainly

---

<sup>\*</sup>PhD student, Gustave Eiffel University, MSME UMR 8208, 5, Boulevard Descartes, 77454, Marne-la-Vallée Cedex 2; amritesh.sinha@univ-eiffel.fr.

<sup>†</sup>Professor, Gustave Eiffel University, MSME UMR 8208, 5, Boulevard Descartes, 77454, Marne-la-Vallée Cedex 2; christian.soize@univ-eiffel.fr.

<sup>‡</sup>Professor, Gustave Eiffel University, MSME UMR 8208, 5, Boulevard Descartes, 77454, Marne-la-Vallée Cedex 2; christophe.desceliers@univ-eiffel.fr.

<sup>§</sup>Engineer, Propulsion Physics Advanced Methods, Data Integration & Advanced Methods (IPGM); guilherme.coelho-cunha@airbus.com

attenuated by resonance effects while viscous dissipation acts on both tonal and broadband noise. Noise attenuation by acoustic treatments is tuned for the blade passing frequency (BPF), whereas dissipating as much as possible the broadband component, by modifying the liner geometry or intrinsic properties. In order to be effective in absorbing fan noise, liners have to be studied in their operating environment *i.e.* in different flight conditions. In the past, many experiments have been carried out for identifying uncertainties related to liner impedance (see for instance [1–4]). High-fidelity computational models have also been developed for predicting the liner performances (see for instance [5–10]). The design of liners is of prime interest and many works have been published on this subject (see for instance [11–17]). Statistical inference such as Bayesian approaches have recently been used for statistical inverse problems related to liner impedances (see for instance [18, 19]). The acoustic performance of a liner depends on the quantities that are highly related to the operating conditions, such that velocity, mean pressure, and fluid density. Any external variation directly impacts the environment of the liners and thus the acoustic performance of the liners system. This requires to take into account uncertainties in the high-fidelity computational model of the liners system. An uncertain computational model of the liners system is presented by Dangla [16], which allows for quantifying uncertainties in aeroacoustic models of liner performance.

### **A. Objective of the Paper**

A liner is characterized by its acoustic absorption, which will be modeled by its (local) acoustic impedance adapted to the low-frequency tonal noise. In this work, the liner acoustic impedance is estimated using an aeroacoustic computational model (ACM) applied to a simplified configuration of the liner (see Section II.B). One evaluation with such an ACM is computationally expensive and consequently, the ACM cannot be used a large number of times for constructing a parameterized aeroacoustic liner impedance. This is the reason why we propose to develop a statistical metamodel of the impedance whose parameterization is frequency  $\omega$ , percentage of open area (POA), and sound pressure level (SPL). The Mach number has been set to zero to minimize the number of parameters that need to be considered for the metamodel. In fact, including a non-zero Mach number would require analyzing how the impedance changes with respect to various boundary layer properties (such as friction velocity, boundary layer thickness, local Reynolds number, bulk Mach number, and so on), which is beyond the scope of this study. Further investigations on this topic are currently underway and will be addressed in future work. Due to the use of a simplified configuration, it is of prime importance to take into account model uncertainties induced by modeling errors in the construction of the impedance model. These uncertainties on the impedance model will be introduced in the statistical metamodel of the impedance and the level of uncertainties will be identified with experiments. We thus propose to develop a statistical metamodel of the parameterized liner impedance, which is robust with respect to uncertainties. The objective of the paper is therefore to present a useful methodology to build a robust statistical metamodel of the aeroacoustic liner impedance as a function of the control parameters. The property of robustness is necessary with respect to uncertainties but also to the small

amount of data available. A methodology is thus presented to circumvent the prohibitive computational cost of building a parameterized model of the aeroacoustic liner impedance. Such a statistical metamodel can, for example, be used to carry out a parametric analysis of the impedance, to perform its acoustic performance, and can also be included in an optimization loop. Such possible applications are beyond the scope of the paper.

### **B. Why POA is chosen as a design parameter**

Let  $t$ ,  $d$ , and  $L$  be the perforated plate thickness, the hole diameter, and the length from the perforated plate to the rigid backing sheet, respectively. Parameters  $t$ ,  $d$ , and  $L$  have been held constant and equal to 1.0 mm, 0.8 mm, and 9.6 mm, respectively. It is important to note that the POA is the most important parameter when considering liner design. If POA was kept with a constant value and if diameter  $d$  was chosen as a design variable, then almost no impact would be observed on the effects of attenuation. For the range of liners considered, height  $L$  of the cavity mainly influences the reactance, and has a negligible impact on the resistance (which is mainly a function of the geometric parameters of the resistive layer). However, it is possible to separate the reactance contribution of the cavity from that of the resistive sheet, by subtracting the cotangent term from the reactance value obtained. This is done in Eq.(3). This makes the presented results independent of  $L$ . Regarding thickness  $t$ , its value has been set at 1.0 mm. This value is representative of what is used in the aeronautical industry. It should be noted that, for the design, this parameter is less important than the POA. It should also be noted that the integration constraints make the chosen thickness value a very representative value.

### **C. Sources of uncertainties and variabilities, and their consideration in the development of the metamodel**

There are two types of uncertainties: the uncertainties on the model parameters called the “model-parameter uncertainties” and the “model uncertainties” induced by the modeling errors. In addition, there are some “variabilities” in the real system, due to manufacturing process and due to small differences in the configurations: an experimental configuration of a complex system differs from the designed system and is never perfectly known. When a computational model of a complex system is developed, model-parameter uncertainties, model uncertainties, and variabilities have to be taken into account. When experiments are available, the probability model of uncertainties has to be identified by solving a statistical inverse problem. If information is not available for the construction of an informative probability model of model-parameter uncertainties (that is the case for the considered problem), model-parameter uncertainties and model uncertainties can simultaneously be taken into account using a nonparametric probability model or the output-predictive-error method as explained in [20]. The latter consists in introducing an additive or multiplicative noise on the quantities of interest of the computational model and in identifying the hyperparameters that control the probabilistic model of the noise using the experiments. It is the latter method that is proposed in this work.

#### **D. Proposed Methodology**

The development of the statistical metamodel (surrogate model) of the aeroacoustic liner impedance is based on the use of a probabilistic learning tool and a training dataset made up of a small number of points. Experimental data is directly integrated in the proposed statistical metamodel in order to take into account model uncertainties (see Section I.A). Note that there is no direct connection between the experimental values and ACM that is constructed independently of the experimental values. The experimental values are only used to calibrate the level of the noise that is added in the model to take into account uncertainties. The learning tool is the probabilistic learning on manifolds (PLoM) [21–23]. The training dataset is generated using experimental data [24] and Lavieille’s work [5] for numerical predictions of acoustic impedance. It should be noted that the development of the metamodel is not performed by using a proper experimental dataset because such dataset would require to carry out a complete experimental campaign that is currently not available. In addition, it would be a challenge to be able to transfer the statistical fluctuations from different dataset (using different experimental cases) to a given computational model. The main limitation regarding third-party experimental datasets is that we do not have any information about the liners manufacturing process. This makes it difficult to compare directly any experimental data with numerical ones, unless we have quality check information on manufactured liners. For instance, it is known that the nominal geometrical parameters and the ones that are effectively manufactured might greatly differ (by more than 20%, depending solely on the manufacturing process). This means that any direct comparison would be biased, unless it is made via statistical quantities. Experimental means varies from one research/industrial institution to another and, unless we compare numerical data to experimental ones after a careful quality check, it would be rather difficult to make a quantitative comparison of all data. In this context, we have used the available experimental database from [24] for which the experimental results are presented in a dimensionless form, covering a large family of liners. As explained, such a database does not permit to carry out a direct probabilistic inference because the experimental statistical mean values cannot be used, as it does not correspond to the case being analyzed in this paper. Only the experimental statistical fluctuations around the experimental statistical mean will be used to calibrate the noise that models uncertainties. Therefore, physics-based constraints cannot be used for the learning step. In order to circumvent this difficulty, the experimental information is integrated in the training dataset using the output predictive error model [20].

#### **E. Organization of the Paper**

Section II deals with the definition of the aeroacoustic liner system and with the aeroacoustic computational model that is used to generate the simulated data. In Section III, the experimental dimensional data are derived from the experimental dimensionless data, extracted from the open literature. A comparison of the aeroacoustic computation model predictions with the dimensional experiments is presented. Section IV is devoted to the construction of the probability model of the model uncertainties, which is calibrated using the experimental dimensional data. Then the

training dataset including model uncertainties is constructed. In Section V we present the construction of the statistical metamodel of the parameterized random liner impedance and then introduce conditional statistics and the methodology to estimate it on the basis of probabilistic learning. Finally Section VI deals with the presentation of the results and its discussion.

A summary of the probabilistic learning on manifolds (PLOM) algorithm, its parameterization, and the algebraic expressions of the conditional statistics are given in Supplemental Material of this paper.

## II. Defining the Liner System and its Aeroacoustic Computational Model

### A. Definition of the liner system

We consider a perforated liner system whose scheme is shown in Fig. 1. It is constituted of a perforated plate, a rigid backing sheet, and a honeycomb constituting drainage holes. As explained above, the parameters used to control the liner system are the POA and the SPL. We then define the control parameter as  $\mathbf{w} = (w_1, w_2)$  in which  $w_1$  is POA and  $w_2$  is SPL. The control parameter will be modeled by a  $\mathbb{R}^2$ -valued random variable  $\mathbf{W} = (W_1, W_2)$  whose prior probability distribution will allow for generating samples of  $\mathbf{W}$ .

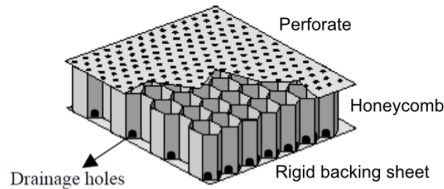
### B. Computation of the impedance using an aeroacoustic computational model

The domain decomposition used for the aeroacoustic computational model is the one proposed in [5] and is shown in Fig. 2. The left figure displays the scheme of elementary period  $d_1$ . The central figure shows the domain that is constituted of three subdomains: the outside domain, the resistive sheet domain, and the resonator domain of elementary period  $d_1$ . The right figure represents the reduced resonator domain of period  $\delta_1$ . In the outside domain, the acoustic field is described by the sum of incident and diffracted plane waves in order to compute sound pressure level (SPL) at the upper face of the resistive sheet that allows an equivalent reduced impedance  $z^{\text{acm}}(\omega; \mathbf{w}) = r^{\text{acm}}(\omega; \mathbf{w}) + \iota v^{\text{acm}}(\omega; \mathbf{w})$  that depends on the frequency  $\omega$  and on the control parameter  $\mathbf{w}$ , in which  $\iota = \sqrt{-1}$ ,  $r^{\text{acm}}(\omega; \mathbf{w})$  is the resistance, and where  $v^{\text{acm}}(\omega; \mathbf{w})$  is the reactance. In the reduced resonator domain, which belongs to the resistive sheet domain, the nonlinear Navier-Stokes equations are solved in order to well capture the viscous effects and the nonlinear phenomena at higher SPL. The period  $\delta_1$  is chosen so that the corresponding POA is verified at the resistive sheet level. Then, a bi-periodic condition is imposed. This means that any generated vortex can cross the lateral boundaries and still be correctly taken into account. Consequently, the third domain (see Fig. 2) is sufficiently wide to capture this motion. We refer the reader to the paper [5] for the details concerning this aeroacoustic computational model. It should be noted that such a simplified aeroacoustic model yields large computational cost for exploring the flight and design configurations. Nevertheless, it demonstrates the interest of the proposed methodology for constructing a robust statistical metamodel. More complex aeroacoustic computational models can be found in [25–27]. For each value of the POA and SPL, and for each sampled frequency  $\omega_k \in C_\omega = \{\omega_1, \dots, \omega_{n_\omega}\}$ , the aeroacoustic computational model computes  $r^{\text{acm}}(\omega_k)$ ,

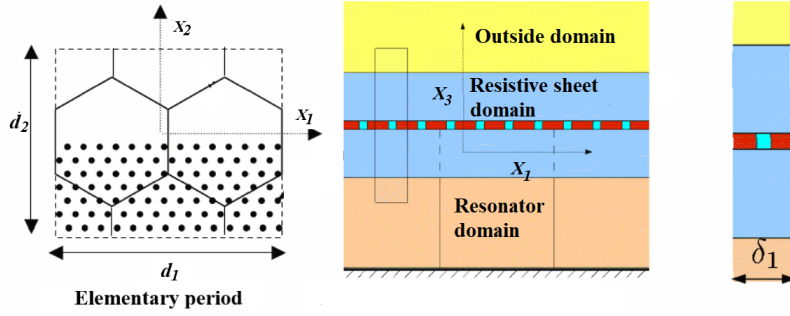
$v^{\text{acm}}(\omega_k)$ , and the hole-orifice velocity  $v_{\text{or}}^{\text{acm}}(\omega_k)$ . In the computational method used, only the reduced resonator domain is meshed for Navier-Stokes computation. For the considered liner, the mesh is made up of 41 781 vertices, 206 560 elements, and 278 514 degrees of freedom. The mesh in the refinement zone (see Fig. 3) is adapted according to the POA value. For each value of the POA and SPL, and for the seven sampled frequencies, the CPU time is 448 hours (using a 64 cores computer). The computation has been done by Airbus using the SANUMO software [5] for  $n_\omega = 7$  sampled frequencies and for  $(\text{POA}, \text{SPL}) \in [0.03, 0.1] \times [130, 145]$ . Since the resistance is very sensitive to the values of POA, in order to not artificially increase the statistical fluctuations in the conditional statistics that will be constructed using probabilistic learning, interval  $[0.03, 0.1]$  is split as  $[0.03, 0.05] \cup [0.06, 0.1]$ , the first one containing 3 points and the second one 5 points. More precisely, the larger the "diameter" of the support (domain of the values of  $\mathbf{w}$ ) of the probability measure of the control parameter  $\mathbf{w}$ , the larger the width of the confidence domain of the quantities of interest. Therefore a compromise has to be made between the choice of the points of the training dataset and its effects on the amplitude of the statistical fluctuations of the quantities of interest, which control the width of the confidence domain. An analysis has been conducted to find the partition that is proposed. Note that the SPL interval contain 6 points. In this paper, the presented analysis is performed for each POA-subinterval  $[a_{\text{POA}}, b_{\text{POA}}]$ . We therefore consider the values of  $\mathbf{w}$  belonging to the finite set

$$C_{\mathbf{w}} = \{\mathbf{w}^j, j = 1, \dots, n_d\} \subset \mathcal{S}_{\mathbf{w}} \quad , \quad \mathbf{w}^j \in \mathbb{R}^{n_w} \quad , \quad (1)$$

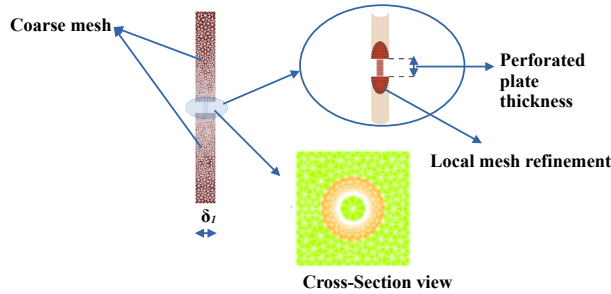
in which  $n_w = 2$  and  $n_d = 18$  for the first subinterval and  $n_d = 30$  for the second one, where the points in  $C_{\mathbf{w}}$  are considered as samples of the random variable  $\mathbf{W} = (W_1, W_2)$ , which belong to the domain  $\mathcal{S}_{\mathbf{w}} = [a_{\text{POA}}, b_{\text{POA}}] \times [130, 145] \subset \mathbb{R}^{n_w}$  ( $\mathcal{S}_{\mathbf{w}}$  can be viewed as the support of the prior probability distribution of  $\mathbf{W}$ ). We now define the ACM simulation dataset related to the points of  $C_{\mathbf{w}}$ . For  $j \in \{1, \dots, n_d\}$ , we introduce the vectors  $\mathbf{r}^{\text{acm},j} = (r^{\text{acm},j}(\omega_1), \dots, r^{\text{acm},j}(\omega_{n_\omega})) \in \mathbb{R}^{n_\omega}$  and  $\mathbf{v}^{\text{acm},j} = (v^{\text{acm},j}(\omega_1), \dots, v^{\text{acm},j}(\omega_{n_\omega})) \in \mathbb{R}^{n_\omega}$  in which  $r^{\text{acm},j}(\omega_k) = r^{\text{acm}}(\omega_k; \mathbf{w}^j)$  and  $v^{\text{acm},j}(\omega_k) = v^{\text{acm}}(\omega_k; \mathbf{w}^j)$ . Finally, for  $j \in \{1, \dots, n_d\}$ , we introduce the vector  $\mathbf{q}^{\text{acm},j} = (\mathbf{r}^{\text{acm},j}, \mathbf{v}^{\text{acm},j}) \in \mathbb{R}^{n_q} = \mathbb{R}^{n_\omega} \times \mathbb{R}^{n_\omega}$  with  $n_q = 2n_\omega$ . The points of the ACM simulation dataset  $\{\mathbf{q}^{\text{acm},1}, \dots, \mathbf{q}^{\text{acm},n_d}\}$  are the realizations of the  $\mathbb{R}^{n_q}$ -valued random variable  $\mathbf{Q}^{\text{acm}}$ .



**Fig. 1** Scheme of a perforated liner. [Reprinted from [5] with permission of Maud Lavielle].



**Fig. 2 Domain decomposition of the aeroacoustic computational model. [Reprinted from [5] with permission of Maud Lavieille].**



**Fig. 3 Mesh of reduced resonator domain: part with mesh refinement and part with coarse mesh.**

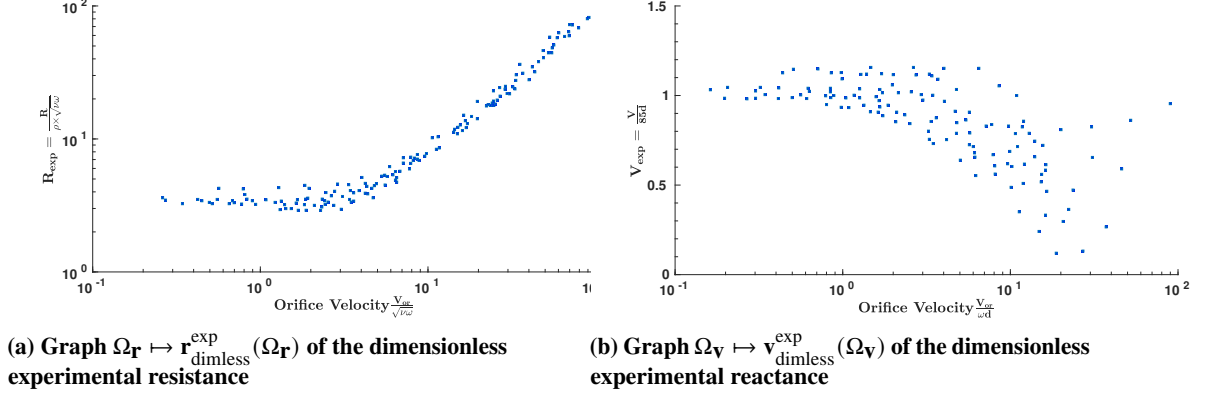
### III. Experimental Dimensional Data and Comparison with ACM Simulations

#### A. Experimental Dimensional Data

As previously mentioned (see Section I.D) experimental data compiled by Pantan and Goldman [24] are used to calibrate the level of uncertainties in the constructed statistical metamodel. The experimental dimensionless impedance  $z_{\text{dimless}}^{\text{exp}}(\Omega)$  is presented as a function of dimensionless orifice velocity  $\Omega = \Omega_{\text{r}} = v_{\text{or}}(\omega; \mathbf{w})/\sqrt{\nu\omega}$  for dimensionless resistance (see Fig. 4a) and  $\Omega = \Omega_{\text{v}} = v_{\text{or}}(\omega; \mathbf{w})/(\omega d)$  for dimensionless reactance (see Fig. 4b), in which  $\omega$  is the angular frequency,  $v_{\text{or}}(\omega; \mathbf{w})$  is the experimental frequency-dependent dimensional orifice velocity that depends on control parameter  $\mathbf{w}$ ,  $\nu$  is the kinematic viscosity, and  $d$  is the hole diameter. It should be noted that  $\Omega_{\text{r}}$  and  $\Omega_{\text{v}}$  are defined using the same frequency-dependent dimensional orifice velocity,  $v_{\text{or}}(\omega; \mathbf{w})$ . The images shown in Fig. 4 are digitized and the datapoints are extracted using Plotdigitizer software [28]. It should be noted that the resistance and reactance are not normalized in the same way in Fig. 4 as described at the beginning of Section III.A. Nevertheless, the rescaling was properly performed. For each considered value of  $\Omega_{\text{r}}$  and  $\Omega_{\text{v}}$ , the mean of the closest values of the experimental dimensionless resistance and reactance are extracted, which allows for replacing the cloud of experimental



dimensionless points into experimental dimensionless curve. In order to use these experimental data for constructing the statistical metamodel, these dimensionless data are transformed as a function of dimensionless orifice velocity  $\Omega$  into dimensional quantities as a function of the frequency,  $\omega$ . These dimensionless experimental data are transformed



**Fig. 4** Dimensionless resistance (a) and reactance (b) as a function of dimensionless orifice velocity [Compiled from [24] with permission. Copyright 1976, Acoustical Society of America].

into dimensional data that are compatible with the ACM simulation data. For the experimental dimensional resistance and reactance at sampling points  $\omega_k$  and for given value of control parameter  $\mathbf{w}$ , these transformations yield

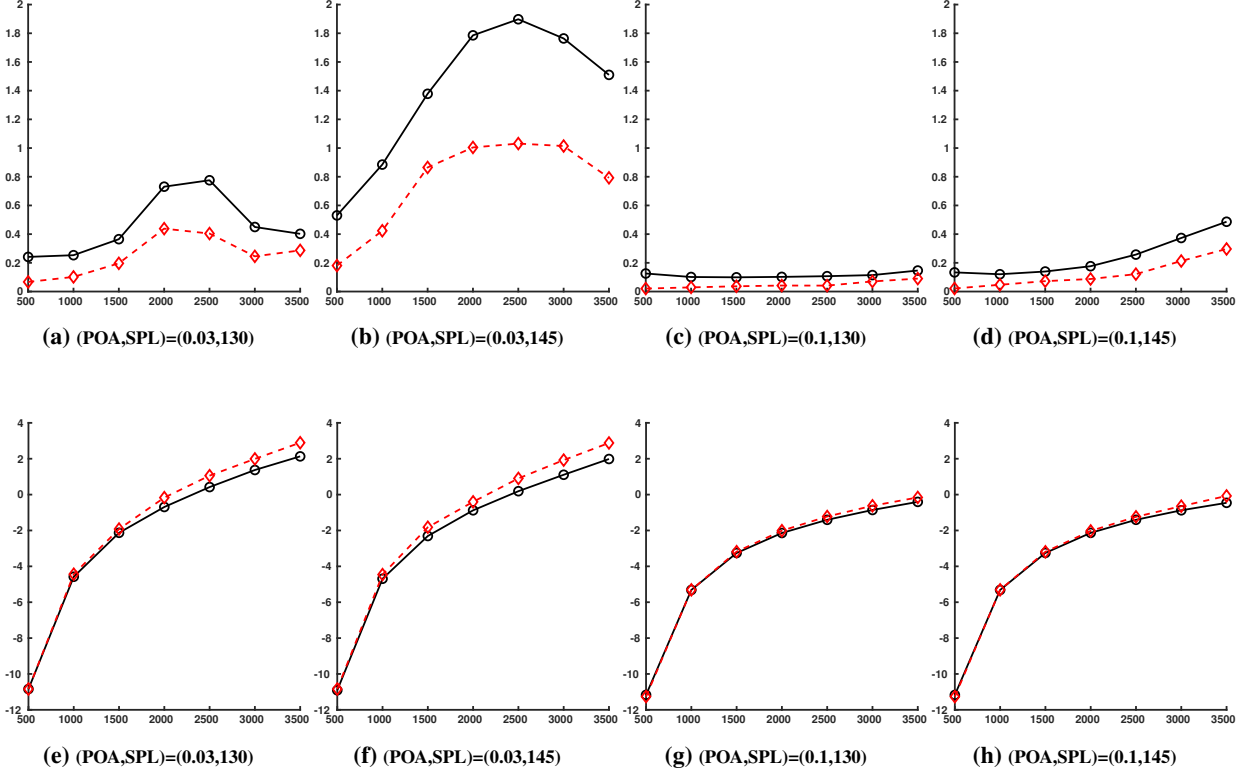
$$\mathbf{r}^{\text{exp}}(\omega_k; \mathbf{w}) = \mathbf{r}_{\text{dimless}}^{\text{exp}}(\Omega_{\mathbf{r},k}(\mathbf{w})) \times \frac{\rho \sqrt{v \omega_k}}{c \times w_1} \quad (2)$$

$$\mathbf{v}^{\text{exp}}(\omega_k; \mathbf{w}) = \left[ \left( \mathbf{v}_{\text{dimless}}^{\text{exp}}(\Omega_{\mathbf{v},k}(\mathbf{w})) \times \frac{8}{3\pi} + \frac{t}{d} \right) \frac{\rho \omega_k d}{c \times w_1} \right] - \cot \left( \frac{\omega_k L}{c} \right) \quad (3)$$

in which  $\cot$  is the cotangent,  $\omega_k$  is a sampled frequency point,  $\rho$  is the air density,  $c$  is the speed of sound,  $w_1$  is the percentage of open area (POA),  $t$  is the perforated plate thickness, and  $L$  is length from the perforated plate to the rigid backing sheet. The  $\mathbf{w}$ -dependent dimensionless frequencies are such that  $\Omega_{\mathbf{r},k}(\mathbf{w}) = v_{\text{or}}^{\text{acm}}(\omega_k; \mathbf{w}) / \sqrt{v \omega_k}$  and  $\Omega_{\mathbf{v},k}(\mathbf{w}) = v_{\text{or}}^{\text{acm}}(\omega_k; \mathbf{w}) / (\omega_k d)$ , in which  $v_{\text{or}}^{\text{acm}}(\omega_k; \mathbf{w})$  is the orifice velocity computed with ACM for a given  $\omega_k$  and  $\mathbf{w}$ . Eqs. (2) and (3) are used for converting the dimensionless data plotted in Fig.4 to the ones used in the current paper. The dimensionless factors are the ones proposed by Pantan in [24]. These parameters happen to have a link with the Crandall model, but as well as other models, such as Poiseuille's, as described in [24]. The dashed curves with diamond markers represent, in Figs. 5-(a) to (d), the experimental results for the resistance for the considered values of  $\mathbf{w}$  that is the pair (POA, SPL), and in Figs. 5-(e) to (h), the reactance for the same values of  $\mathbf{w}$ .

## B. Comparison of the ACM Predictions with Experiments

Fig. 5 shows the comparison between the ACM simulation dataset and the experimental dimensional dataset for the resistance and for the reactance, for several values of  $\mathbf{w}$ , the pair (POA, SPL). In Figs. 5-(a) to (d), the solid curves



**Fig. 5 Resistance ((a) to (d)) and reactance ((e) to (h)) versus frequency (Hz) (horizontal axis) for several values of (POA, SPL): ACM simulation (circles), experiments (diamonds).**

with circular markers represent the computational results for the resistance for different values of  $w$  that is the pair (POA, SPL). In Figs. 5-(e) to (h), the reactance is plotted for the same values of  $w$ . In these figures, the POA and SPL values correspond to its minimum and maximum. These comparisons show that the experimental mean is far from the ACM simulations for the resistance whose good prediction is of prime interest. Consequently, even if the variations are of the same type for experiments and simulations, these experimental values cannot directly be used for a probabilistic inference. Nevertheless, it will be assumed that the experimental statistical fluctuations around the experimental mean is representative of the level of model uncertainties and consequently, will be used for calibrating the level of uncertainties (see Section IV).

## IV. Probability Model of Model Uncertainties and Training Dataset

### A. Probability Model of Model Uncertainties

As previously explained, the experimental dimensional dataset only gives a consistent information about the statistical fluctuations and not about the mean values. Consequently, a probabilistic inference (Bayesian or likelihood methods, or learning constrained by experiments) cannot be used. Nevertheless, these experimental dimensional data are used for calibrating an additive noise (model uncertainties) to the ACM simulated data yielding the training dataset used by the

probabilistic learning. In the proposed construction, the noise level is controlled by a hyperparameter  $\delta^{\text{exp}}$  (that will be a the coefficient of variation). This hyperparameter has to be estimated for any value of the control parameter  $\mathbf{w}$ , using the experimental dimensional dataset. Consequently, conditional statistics have to be used. However, the number of these experimental points is too small to obtain a good convergence of the estimate of  $\delta^{\text{exp}}$ . We then have to resample the probability distribution of the experimental points using PLoM followed by conditional statistics. The steps of the construction are the following.

(i) The first step consists in using the probabilistic learning (the PLoM algorithm summarized in Appendix A of the Supplemental Material) for which we have to define the random vector  $\mathbf{X}^{\text{exp}}$  and its training dataset for the learning process. For all  $\mathbf{w}$ , we define the vectors in  $\mathbb{R}^{n_\omega}$  such that  $\mathbf{r}^{\text{exp}}(\mathbf{w}) = (r^{\text{exp}}(\omega_1; \mathbf{w}), \dots, r^{\text{exp}}(\omega_{n_\omega}; \mathbf{w}))$  and  $\mathbf{v}^{\text{exp}} = (v^{\text{exp}}(\omega_1; \mathbf{w}), \dots, v^{\text{exp}}(\omega_{n_\omega}; \mathbf{w}))$  related to the experimental dimensional resistance and reactance, in which  $r^{\text{exp}}(\omega_k; \mathbf{w})$  and  $v^{\text{exp}}(\omega_k; \mathbf{w})$  are defined by Eqs. (2) and (3). Modeling the deterministic vector  $\mathbf{w}$  by random variable  $\mathbf{W}$  yields the  $\mathbb{R}^{n_\omega}$ -valued random variables  $\mathbf{R}^{\text{exp}} = \mathbf{r}^{\text{exp}}(\mathbf{W})$  and  $\mathbf{V}^{\text{exp}} = \mathbf{v}^{\text{exp}}(\mathbf{W})$ . Let  $\mathbf{Q}^{\text{exp}} = (\mathbf{R}^{\text{exp}}, \mathbf{V}^{\text{exp}})$  be the  $\mathbb{R}^{n_q}$ -valued random variable. Let  $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$  with  $n = n_q + n_w$ . Let  $\mathbf{X}^{\text{exp}} = (\mathbf{Q}^{\text{exp}}, \mathbf{W})$  be the  $\mathbb{R}^n$ -valued random variable whose independent realizations are  $\mathbf{x}^{\text{exp},j} = (\mathbf{q}^{\text{exp},j}, \mathbf{w}^j)$ ,  $j = 1, \dots, n_d$  with  $\mathbf{w}^j \in C_w$ , where  $\mathbf{q}^{\text{exp},j}$  is the realization of  $\mathbf{Q}^{\text{exp}}$ , which is calculated by  $\mathbf{q}^{\text{exp},j} = (\mathbf{r}^{\text{exp}}(\mathbf{w}^j), \mathbf{v}^{\text{exp}}(\mathbf{w}^j))$ . The training dataset for  $\mathbf{X}^{\text{exp}}$  is represented by the matrix  $[x_d^{\text{exp}}] = [\mathbf{x}^{\text{exp},1} \dots \mathbf{x}^{\text{exp},n_d}] \in \mathbb{M}_{n,n_d}$ . For generating  $n_{\text{ar}} = n_d \times n_{\text{MC}}$  learned realizations  $\mathbf{x}_{\text{ar}}^{\text{exp},\ell}$ ,  $\ell = 1, \dots, n_{\text{ar}}$  of  $\mathbf{X}^{\text{exp}}$  with  $n_{\text{MC}} \gg 1$ , the PLoM algorithm is used. We then deduced the  $n_{\text{ar}}$  learned realizations  $(\mathbf{q}_{\text{ar}}^{\text{exp},\ell}, \mathbf{w}_{\text{ar}}^\ell) = \mathbf{x}_{\text{ar}}^{\text{exp},\ell}$  for all  $\ell$  in  $\{1, \dots, n_{\text{ar}}\}$ . As explained in Subsection II.B, the database is split in two parts, one corresponding to  $n_d = 18$  and the other to  $n_d = 30$ , referred to as  $\mathcal{D}_1$  and  $\mathcal{D}_2$ , respectively. After the experimental dataset has been extracted using the methodology described in Subsection III.A a total of  $18 + 30$  datapoints is obtained corresponding to control parameters of the finite set defined by Eq. (1). This is not sufficient to have a good convergence of the vector-valued hyperparameter  $\delta^{\text{exp}}(\mathbf{w}_o)$  that has to be estimated using conditional statistics. Consequently, a resampling of the probability distribution of each experimental dataset  $\mathcal{D}_1$  and  $\mathcal{D}_2$  has been carried out using PLoM and adequate convergence has been obtained for  $n_{\text{ar}} = 20\,000$  and  $n_{\text{ar}} = 40\,000$ , respectively.

(ii) For  $i$  in  $\{1, \dots, n_q\}$ , the experimental conditional coefficient of variation  $\delta_i^{\text{exp}}(\mathbf{w}_o)$  of the component  $Q_i^{\text{exp}}$  of  $\mathbf{Q}^{\text{exp}}$ , given  $\mathbf{W} = \mathbf{w}_o$  in  $S_w$ , is defined by

$$\delta_i^{\text{exp}}(\mathbf{w}_o) = \frac{\sigma_{Q_i^{\text{exp}}}(\mathbf{w}_o)}{m_{Q_i^{\text{exp}}}(\mathbf{w}_o)}, \quad (4)$$

in which  $m_{Q_i^{\text{exp}}}(\mathbf{w}_o) = E\{Q_i^{\text{exp}} | \mathbf{W} = \mathbf{w}_o\}$  and  $\sigma_{Q_i^{\text{exp}}}^2(\mathbf{w}_o) = E\{(Q_i^{\text{exp}})^2 | \mathbf{W} = \mathbf{w}_o\} - m_{Q_i^{\text{exp}}}^2(\mathbf{w}_o)$  in which the conditional mathematical expectation are estimated using the learned realizations  $\{(\mathbf{q}_{\text{ar}}^{\text{exp},\ell}, \mathbf{w}_{\text{ar}}^\ell), \ell = 1, \dots, n_{\text{ar}}\}$  and the Gaussian Kernel Density Estimation (KDE) method (the integration being explicitly calculated, see Appendix B of the

Supplemental Material). We introduce the vector  $\delta^{\text{exp}}(\mathbf{w}_o) = (\delta_1^{\text{exp}}, \dots, \delta_{n_q}^{\text{exp}})$ .

(iii) The random vector  $\mathbf{Q}^{\text{acm}}$  is transformed into a random vector  $\mathbf{Q}$  whose components  $\{Q_i, i = 1, \dots, n_q\}$  are written as

$$Q_i = (1 + B_i)Q_i^{\text{acm}}, \quad i \in \{1, \dots, n_q\}, \quad (5)$$

in which  $1 + B_i$  is a multiplicative random noise that allows model uncertainties to be taken into account. Vector  $\mathbf{Q}$ , which represents the  $\mathbf{W}$ -parameterized and frequency-dependent random liner impedance, will be named the random quantity of interest (QoI).

In order to construct the random variables  $B_1, \dots, B_{n_q}$ , we introduce a vector  $\delta(\mathbf{w}_o) = (\delta_1(\mathbf{w}_o), \dots, \delta_{n_q}(\mathbf{w}_o))$  in which  $\delta_i(\mathbf{w}_o)$  is the coefficient of variation of the conditional random variable  $1 + B_i$  given  $\mathbf{W} = \mathbf{w}_o$ . For the experiments, the parameters presenting variabilities are  $d/\Delta$ ,  $t/d$ ,  $v_{\text{or}}$ ,  $\rho/c$ , and  $L$ , in which  $\Delta = \sqrt{v/\omega}$ . It is assumed that the model uncertainties in the ACM are due to variabilities and uncertainties generated by the same parameters, which justifies the use of the experiments for generating the modeling errors. It should be noted that the components of vectors  $\delta(\mathbf{w}_o)$  and  $\delta^{\text{exp}}(\mathbf{w}_o)$  represent the variations with respect to frequency  $\omega$  for the resistance and reactance. We then choose for the frequency variations of  $\delta(\mathbf{w}_o)$ , the frequency variations of  $\delta^{\text{exp}}(\mathbf{w}_o)$ . If the level of uncertainties were chosen equal for the experiments and for the model, then we would have  $\delta(\mathbf{w}_o) = \delta^{\text{exp}}(\mathbf{w}_o)$ . However, we are interested in performing a sensitivity analysis with respect to the level of model uncertainties. We thus introduce a global parameter  $a_{\text{unc}} \in [0, 1]$  to quantify the level of uncertainties, and consequently, we write  $\delta(\mathbf{w}_o) = a_{\text{unc}} \delta^{\text{exp}}(\mathbf{w}_o)$ . Three values of  $a_{\text{unc}}$  will be considered: small uncertainties  $a_{\text{unc}} = 0.2$ , medium uncertainties  $a_{\text{unc}} = 0.5$ , and large uncertainties  $a_{\text{unc}} = 1.0$ . For all  $i \in \{1, \dots, n_q\}$ , the real-valued random variable  $B_i$  is thus defined by

$$B_i = \delta_i(\mathbf{W}) 2\sqrt{3} \left( U_i - \frac{1}{2} \right), \quad (6)$$

in which  $U_1, \dots, U_{n_q}$  are  $n_q$  independent real-valued random variable uniformly distributed on  $[0, 1]$  and independent of  $\mathbf{W}$ . It can then be shown that  $B_i$  is a centered random variable and that coefficient of variation of the conditional random variable  $1 + B_i$  given  $\mathbf{W} = \mathbf{w}_o$  is  $\delta_i(\mathbf{w}_o)$ .

## B. Training Dataset Including Model Uncertainties

To well represent the statistical fluctuations induced by model uncertainties in the training dataset, for each realization of  $\mathbf{W}$  we will assign  $M_d$  realizations of the random variable  $\mathbf{B}$ . Consequently, for all  $i$  in  $\{1, \dots, n_q\}$ , the  $N_d = n_d \times M_d$  realizations  $\{\{q_i^{1,m}, \dots, q_i^{n_d,m}\}, m = 1, \dots, M_d\}$  of the random variable  $Q_i$  defined by Eq. (5) are computed by the equation

$$q_i^{j,m} = (1 + b_i^{j,m}) q_i^{\text{acm},j}, \quad (7)$$

in which  $b_i^{j,m}$  is the realization of  $B_i$  defined by Eq. (6) and such that

$$b_i^{j,m} = \delta_i(\mathbf{w}^j) 2\sqrt{3} (u_i^{j,m} - \frac{1}{2}), \quad (8)$$

in which  $\{\{u_i^{1,m}, \dots, u_i^{n_d,m}\}, m = 1, \dots, M_d\}$  are  $N_d = n_d \times M_d$  independent realizations of random variable  $U_i$ . It should be noted that for  $m$  fixed  $\{b_i^{j,m}, j = 1, \dots, n_d\}$  are  $n_d$  independent realizations of  $B_i$ , but the realizations in the couple  $(j, m)$  are dependent. We then define the random vector  $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$  with values in  $\mathbb{R}^n$ , in which  $\mathbf{Q}$  is defined by Eq. (5), whose  $(j, m)$ -th realization is  $\mathbf{x}^{j,m} = (\mathbf{q}^{j,m}, \mathbf{w}^{j,m})$  in which  $\mathbf{q}^{j,m}$  is given by Eq. (7) and where  $\mathbf{w}^{j,m} = \mathbf{w}^j$  for all  $(j, m)$ . For the probabilistic learning, the training dataset is represented by the matrix  $[x_d] = [\mathbf{x}^{1,1} \dots \mathbf{x}^{n_d,1}, \dots, \mathbf{x}^{1,M_d} \dots \mathbf{x}^{n_d,M_d}] \in \mathbb{M}_{n,N_d}$ .

## V. Construction of the Statistical Surrogate Model of the Parameterized Random Liner Impedance

### A. Random Manifold Associated with the Parameterized Random Liner Impedance

The random QoI  $\mathbf{Q}$  defined in Section IV.B is written as  $\mathbf{Q} = (\mathbf{R}, \mathbf{V})$  in which  $\mathbf{R} = (R_1, \dots, R_{n_\omega})$  and  $\mathbf{V} = (V_1, \dots, V_{n_\omega})$  are the  $\mathbb{R}^{n_\omega}$ -valued random variables representing the frequency-dependent resistance and reactance. This  $\mathbb{R}^{n_q}$ -valued random variable  $\mathbf{Q}$  can be written as  $\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbf{U})$  in which  $\mathbf{W}$  is the random control parameter with values in  $\mathbb{R}^{n_w}$ ,  $\mathbf{U}$  is the random uncontrolled parameter with values in  $\mathbb{R}^{n_u}$  with  $n_u = n_q$  (see Eq. (8)), and where  $(\mathbf{w}, \mathbf{u}) \mapsto \mathbf{f}(\mathbf{w}, \mathbf{u})$  is a deterministic implicit mapping from  $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$  into  $\mathbb{R}^{n_q}$ . We then define the random mapping  $\mathbf{w} \mapsto \mathbf{F}(\mathbf{w})$  such that for all  $\mathbf{w}$  in  $\mathbb{R}^{n_w}$ ,  $\mathbf{F}(\mathbf{w}) = \mathbf{f}(\mathbf{w}, \mathbf{U})$ . The graph  $\{(\mathbf{F}(\mathbf{w}), \mathbf{w}), \mathbf{w} \in \mathcal{S}_w \subset \mathbb{R}^{n_w}\}$  defines a random manifold in  $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ , in which  $n = n_q + n_w$ . The random variable  $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$  with values in  $\mathbb{R}^n$  is related to this random manifold because  $\mathbf{X}$  can also be rewritten as  $\mathbf{X} = (\mathbf{F}(\mathbf{W}), \mathbf{W})$ . In Section IV.B, we have constructed the training set of  $\mathbf{X}$  made up of  $N_d$  points  $\mathbf{x}^j = (\mathbf{q}^j, \mathbf{w}^j) \in \mathbb{R}^n$ , represented by matrix  $[x_d] \in \mathbb{M}_{n,N_d}$ .

### B. Statistical Surrogate Model

Taking into account the definition of random QoI  $\mathbf{Q}$  defined in Section V.A, the objective of the statistical metamodel of the parameterized random linear impedance is to construct the conditional probability distribution  $P_{\mathbf{Q}|\mathbf{W}}(d\mathbf{q}|\mathbf{w}_o) = p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w}_o) d\mathbf{q}$  of  $\mathbf{Q}$  given  $\mathbf{W} = \mathbf{w}_o$  for any  $\mathbf{w}_o$  in  $\mathcal{S}_w \subset \mathbb{R}^{n_w}$ , in which  $p_{\mathbf{Q}|\mathbf{W}}$  is the conditional probability density function on  $\mathbb{R}^{n_q}$  given  $\mathbf{W}$ . This conditional probability distribution completely defines the statistical metamodel. However, we can only estimate this conditional probability distribution using the nonparametric statistics and sufficiently large set of realizations of  $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ . Since the training dataset of  $\mathbf{X}$ , represented by matrix  $[x_d]$ , is constituted of a small number  $N_d$  of realizations, we will perform a resampling of  $\mathbf{X}$  using a probabilistic learning on manifolds in order to generate a large number  $N_{ar} \gg N_d$  of learned realizations of  $\mathbf{X}$  (see Section V.C). The connection

between the probability density function of  $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$  and the conditional probability density function of interest is the following

$$p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w}_o) = \frac{1}{p_{\mathbf{W}}(\mathbf{w}_o)} p_{\mathbf{Q},\mathbf{W}}(\mathbf{q}, \mathbf{w}_o), \quad (9)$$

in which  $p_{\mathbf{Q},\mathbf{W}}$  is the joint probability density function on  $\mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$  of random variables  $\mathbf{Q}$  and  $\mathbf{W}$ , and where  $p_{\mathbf{W}}(\mathbf{w}_o) = \int_{\mathbb{R}^{n_q}} p_{\mathbf{Q},\mathbf{W}}(\mathbf{q}, \mathbf{w}_o) d\mathbf{q}$  is the probability density function of  $\mathbf{W}$  at point  $\mathbf{w}_o \in \mathcal{S}_w$ .

From an engineering point of view, we are interested in deriving conditional statistics from conditional probability distribution  $p_{\mathbf{Q}|\mathbf{W}}$ , such as the conditional mean values and the conditional confidence regions. The results presented in Section VI will be the conditional mathematical expectation,

$$E\{\mathbf{Q}|\mathbf{W} = \mathbf{w}_o\} = \int_{\mathbb{R}^{n_q}} \mathbf{q} p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w}_o) d\mathbf{q}, \quad (10)$$

for any  $\mathbf{w}_o$  in  $\mathcal{S}_w$ . We will also present the conditional confidence region of the frequency-dependent random resistance  $R_1, \dots, R_{n_\omega}$  and reactance  $V_1, \dots, V_{n_\omega}$ . For  $k$  fixed in  $\{1, \dots, n_\omega\}$ , let  $Q_i$  be the component of  $\mathbf{Q}$  representing either resistance  $R_k$  or reactance  $V_k$ . Then the lower bound  $q_i^-$  and the upper bound  $q_i^+$  of the conditional confidence interval of  $Q_i$  given  $\mathbf{W} = \mathbf{W}_o$  for a probability level  $p_c$  are defined by

$$q_i^+ : \text{Proba}\{Q_i \leq q_i^+ | \mathbf{W} = \mathbf{w}_o\} = p_c, \quad (11)$$

$$q_i^- : \text{Proba}\{Q_i^- \leq q_i^- | \mathbf{W} = \mathbf{w}_o\} = 1 - p_c, \quad (12)$$

and where the probability in Eqs. (11) and (12) are calculated with the conditional cumulative distribution function,

$$\text{Proba}\{Q_i \leq q_i^* | \mathbf{W} = \mathbf{w}_o\} = \int_{-\infty}^{q_i^*} p_{Q_i|\mathbf{W}}(q_i^* | \mathbf{w}_o) dq_i, \quad (13)$$

in which the conditional pdf  $p_{Q_i|\mathbf{W}}$  is derived from conditional pdf  $p_{\mathbf{Q}|\mathbf{W}}$  by an integration on  $\mathbb{R}^{n_q-1}$ .

### C. Generation of Learned Realizations to Estimate the Statistical Surrogate Model

In order to estimate the conditional statistics defined by Eqs. (10) to (13) of the statistical metamodel, we need to generate the learned dataset constituted of a large number  $N_{\text{ar}}$  of learned realizations  $(\mathbf{q}_{\text{ar}}^\ell, \mathbf{w}_{\text{ar}}^\ell)$  of random variable  $(\mathbf{Q}, \mathbf{W})$  using the available information defined by the training dataset, represented by matrix  $[x_d]$ , for which the columns are the  $N_d$  points  $x^j \in \mathbb{R}^n$ . As previously explained, we need a probabilistic learning algorithm and we propose to use the probabilistic learning on manifolds (PLoM) [21–23] for which the algorithm is summarized in the Appendix A of the Supplemental Materials of the paper. Once the learned realizations have been generated, the joint probability density function  $p_{\mathbf{Q},\mathbf{W}}$  of  $\mathbf{Q}$  and  $\mathbf{W}$  is estimated using the multivariate Gaussian Kernel Density Estimation method.

The resulting explicit expression allows for performing exact multiple integration with respect to coordinate vector  $\mathbf{Q}$ . We then obtain explicit algebraic expression for the estimate of Eqs. (10) to (13), which only depends on the learned dataset. These algebraic expressions are given in the Appendix B of the Supplemental Materials of the paper.

## VI. Results and Discussion

The conditional statistics defined in Section V.B are estimated using the learned dataset generated as explained in Section V.C with the following values of the parameters:  $n_d$  defined in Section II.B,  $M_d = 15$ ,  $N_d = n_d \times M_d$ ,  $N_{ar} = N_d \times n_{MC}$  with  $n_{MC} = 6000$ , and  $a_{unc} = 0.5$ . Two types of results are shown.

- (i) The first one is related to the conditional statistics for which the control parameter  $\mathbf{w}$  belongs to  $C_w$  (the training dataset is based on these values).
- (ii) For the second presented analysis, the control parameter does not belong to  $C_w$  (and consequently, does not belong to the training dataset). Recalling that in the training dataset, the  $n_d$  points represent the values of resistance and the reactance as a function of the frequency and the control-parameters samples. Since  $n_d$  is very small, all these points have been kept to construct the training set, and it was not possible to keep a part of these points to perform a quality assessment (or cross-validation). So this second analysis has to be seen as predictions performed by the statistical metamodel, for which quality can only be evaluated by coherence.

### A. Predictions of the statistical metamodel for which the control parameter belongs to the training dataset

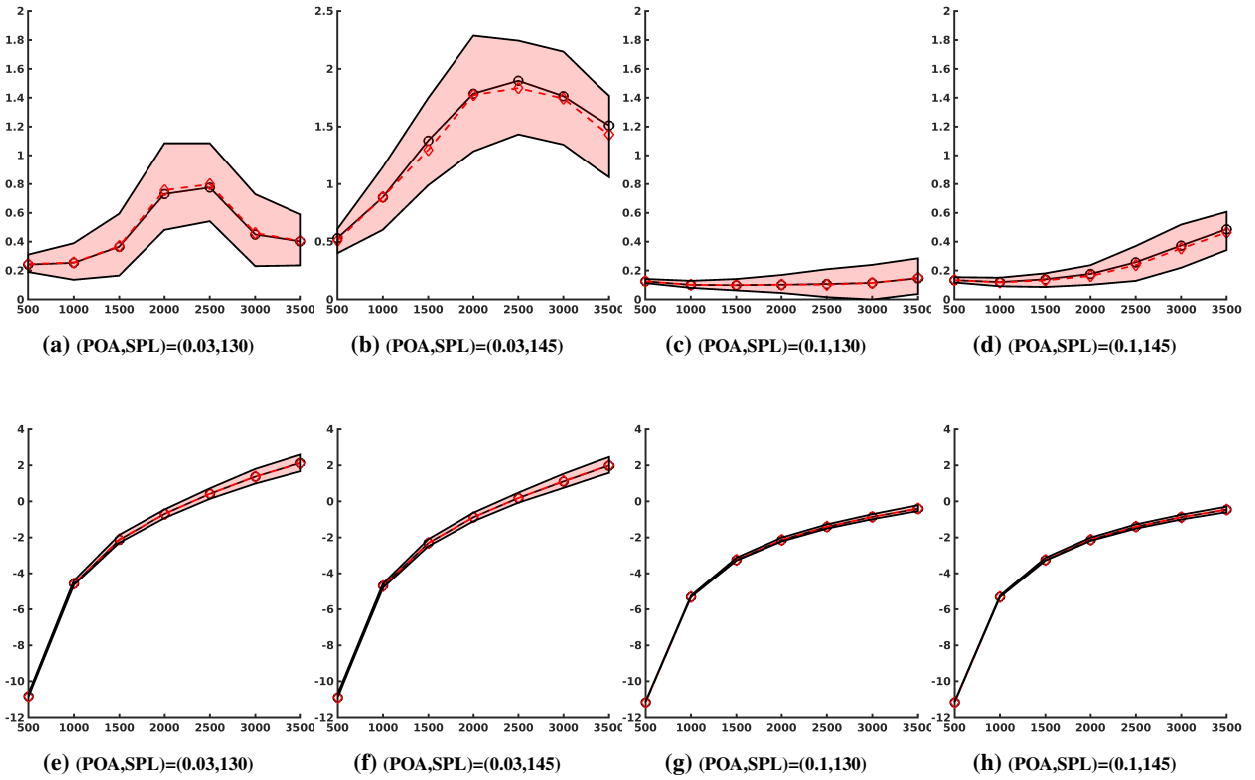
For several values of the control parameters (POA, SPL) that belong to the training dataset, Fig. 6 shows the resistance (figures (a) to (d)) and the reactance (figures (e) to (h)) as a function of frequency. In each plot, it can be seen the curve corresponding to the ACM simulation data. On the other hand, using the training dataset that includes model uncertainties, the curve of the learning-based conditional mathematical expectation is plotted and the learning-based conditional confidence region for a probability level  $p_c = 0.98$  is plotted.

### B. Predictions of the statistical metamodel for which the control parameter does not belong to the training dataset

For this case, the nature of the presented results are the same as those presented in Subsection VI.A but the control parameters (POA, SPL) do not belong to the training dataset. Fig. 7 shows the resistance (figures (a) to (d)) and the reactance (figures (e) to (h)) as a function of frequency.

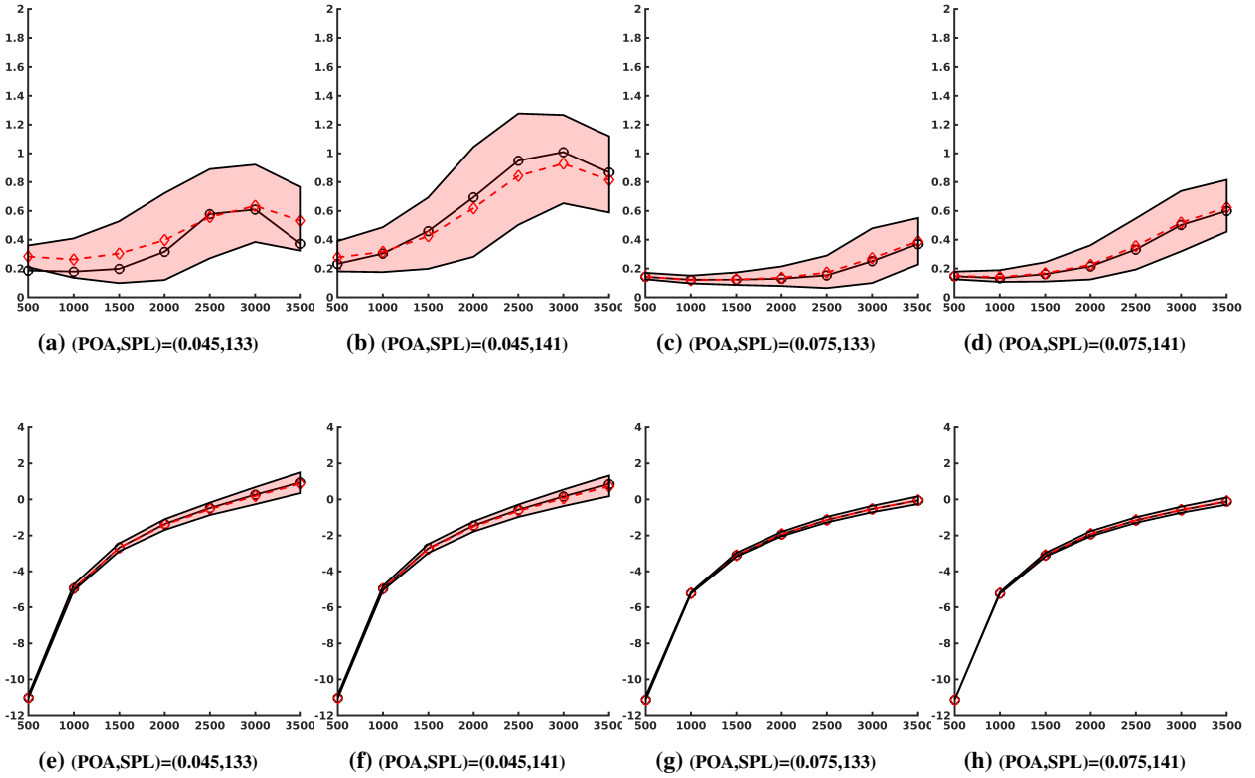
### C. Discussion

It can be seen that the dispersion of the resistance is larger than the reactance. This means that the resistance is more sensitive to statistical fluctuations than the reactance. This dispersion is due to two factors. The first one is directly correlated to the contents of the training dataset without model uncertainties (the ACM simulation data), which contributes to the dispersion of the resistance. The second one is due to the model uncertainties that have been included



**Fig. 6** Statistical metamodel for  $(POA, SPL)$  belonging to the training dataset. Resistance ((a) to (d)) and reactance ((e) to (h)) versus frequency (Hz) (horizontal axis): ACM simulation (circles), learning-based conditional mean (diamonds) and conditional confidence region (shaded domain).





**Fig. 7** Statistical metamodel for  $(POA, SPL)$  not belonging to the training dataset. Resistance ((a) to (d)) and reactance ((e) to (h)) versus frequency (Hz) (horizontal axis): ACM simulation (circles), learning-based conditional mean (diamonds) and conditional confidence region (shaded domain).

in the training dataset. The analyses performed have shown that the contributions of the statistical fluctuations in the ACM simulation data is more dominant than those induced by the model uncertainties. For the considered values of the control parameters the dispersion of the reactance always stays small, while, the dispersion of the resistance strongly depends on the control-parameters values. It should be noted that, for each given value of the control parameter, the estimated conditional expectation of the resistance and reactance is close to the ACM simulation data. This proximity is due to the fact the random variable  $\mathbf{B}$  related to the model uncertainties is centered and, in addition, in the probabilistic learning process, the mean value corresponds to the ACM simulation data. Finally, Fig. 7 related to the second analysis, shows that the predictions performed by the statistical metamodel are coherent with respect to those shown in Fig. 6.

## **VII. Conclusion**

We have presented a methodology for constructing a robust statistical metamodel of the aeroacoustic liner impedance as a function of the frequency and on the main control parameters, the POA and the SPL. In this work the Mach number has been chosen to be zero. It corresponds to an important configuration that has to be studied. For nonzero Mach number, the frequency evolutions and the amplitudes of the aeroacoustic impedance are different and require additional works, using the presented methodology. This construction is based on computationally expensive aeroacoustic model to generate simulated data, which yields a small training dataset. This means that a learning process has to be used and we have chosen the PLoM algorithm. Although the aeroacoustic simulation is conducted with a large aeroacoustic computational model, some approximations have been introduced, generating model errors. A probability model of these model errors has been developed to construct the training dataset. In order to calibrate the model errors, we have used dimensionless experiments available from the open literature. In addition, we have also introduced a sensitivity parameter to the level of model uncertainties. The results that have been presented correspond to a medium value of model uncertainties. Despite the fact that we have very small amount of data, we have succeeded in proposing a robust statistical metamodel that is novel and whose predictions are consistent. This statistical framework allows for exhibiting a confidence region, which gives an information about the level of uncertainties about the aeroacoustic liner impedance as a function of the frequency and the control parameters. Finally, it should be noted that the main contribution of this paper is the methodology presented, which is independent of the choice of the aeroacoustic computational model (more or less simplified) and of the choice of experiments used to estimate the level of uncertainties. Concerning the application presented, the experimental data and the aeroacoustic computation could be replaced by others.

## **Acknowledgments**

The works presented here were partially funded by the French Civil Aviation Authority (DGAC - Direction Générale de l'Aviation Civile) under the framework of the Mambo Project, in which Airbus has partnered with the Laboratory MSME of Université Gustave Eiffel.

## References

- [1] Jones, M., Parrott, T., and Watson, W., “Uncertainty and sensitivity analyses of a two-parameter impedance prediction model,” *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)*, 2008, p. 2928. <https://doi.org/10.2514/6.2008-2928>.
- [2] Nark, D. M., Jones, M. G., and Piot, E., “Assessment of axial wave number and mean flow uncertainty on acoustic liner impedance education,” *2018 AIAA/CEAS Aeroacoustics Conference*, 2018, p. 3444. <https://doi.org/10.2514/6.2018-3444>.
- [3] Brown, M., Jones, M., and Watson, W., “Uncertainty analysis of the grazing flow impedance tube,” *18th AIAA/CEAS Aeroacoustics Conference (33rd AIAA Aeroacoustics Conference)*, 2012, p. 2296. <https://doi.org/10.2514/6.2012-2296>.
- [4] Spillere, A. M. N., Bonomo, L. A., Cordioli, J. A., and Brambley, E. J., “Experimentally testing impedance boundary conditions for acoustic liners with flow: Beyond upstream and downstream,” *Journal of Sound and Vibration*, Vol. 489, 2020, p. 115676. <https://doi.org/10.1016/j.jsv.2020.115676>.
- [5] Lavieille, M., Abboud, T., Bennani, A., and Balin, N., “Numerical simulations of perforate liners: Part I - Model description and impedance validation,” *19th AIAA/CEAS Aeroacoustics Conference*, 2013. <https://doi.org/10.2514/6.2013-2269>.
- [6] Van Antwerpen, B., Detandt, Y., Copiello, D., Rosseel, E., and Gaudry, E., “Performance improvements and new solution strategies of Actran/TM for nacelle simulations,” *20th AIAA/CEAS Aeroacoustics Conference*, 2014, p. 2315. <https://doi.org/10.2514/6.2014-2315>.
- [7] Pascal, L., Piot, E., and Casalis, G., “A new implementation of the extended Helmholtz resonator acoustic liner impedance model in time domain CAA,” *Journal of Computational Acoustics*, Vol. 24, No. 01, 2016, pp. 1663–1674. <https://doi.org/10.1142/S0218396X15500150>.
- [8] Casadei, L., Deniau, H., Piot, E., and Node-Langlois, T., “Time-Domain Impedance Boundary Condition Implementation in a CFD solver and validation against experimental data of acoustical liners,” *eForum Acusticum*, 2020, pp. 359–366. <https://doi.org/10.48465/fa.2020.0088>.
- [9] Dangla, V., Soize, C., Cunha, G., Mosson, A., Kassem, M., and Van Den Nieuwenhof, B., “Stochastic computational model of 3D acoustic noise predictions for nacelle liners,” *AIAA Aviation 2020 Forum*, 2020, p. 2545. <https://doi.org/10.2514/6.2020-2545>.
- [10] Winkler, J., Mendoza, J. M., Reimann, C. A., Homma, K., and Alonso, J. S., “High fidelity modeling tools for engine liner design and screening of advanced concepts,” *International Journal of Aeroacoustics*, Vol. 20, No. 5-7, 2021, pp. 530–560. <https://doi.org/10.1177/1475472X211023884>.
- [11] van Den Nieuwenhof, B., Detandt, Y., Lielens, G., Rosseel, E., Soize, C., Dangla, V., Kassem, M., and Mosson, A., “Optimal design of the acoustic treatments damping the noise radiated by a turbo-fan engine,” *23rd AIAA/CEAS Aeroacoustics Conference*, 2017, p. 4035. <https://doi.org/10.2514/6.2017-4035>.

- [12] Nark, D. M., and Jones, M. G., “Design of an advanced inlet liner for the quiet technology demonstrator 3,” *25th AIAA/CEAS Aeroacoustics Conference*, 2019, p. 2764. <https://doi.org/10.2514/6.2019-2764>.
- [13] Sutliff, D. L., Nark, D. M., Jones, M. G., and Schiller, N. H., “Design and acoustic efficacy of a broadband liner for the inlet of the DGEN aero-propulsion research turbofan,” *25th AIAA/CEAS Aeroacoustics Conference*, 2019, p. 2582. <https://doi.org/10.2514/6.2019-2582>.
- [14] Chambers, A. T., Manimala, J. M., and Jones, M. G., “Design and optimization of 3D folded-core acoustic liners for enhanced low-frequency performance,” *AIAA Journal*, Vol. 58, No. 1, 2020, pp. 206–218. <https://doi.org/10.2514/1.J058017>.
- [15] Özkaya, E., Gauger, N. R., Hay, J. A., and Thiele, F., “Efficient Design Optimization of Acoustic Liners for Engine Noise Reduction,” *AIAA Journal*, Vol. 58, No. 3, 2020, pp. 1140–1156. <https://doi.org/10.2514/1.J057776>.
- [16] Dangla, V., Soize, C., Cunha, G., Mosson, A., Kassem, M., and Van den Nieuwenhof, B., “Robust Three-Dimensional Acoustic Performance Probabilistic Model for Nacelle Liners,” *AIAA Journal*, Vol. 59, No. 10, 2021, pp. 4195–4211. <https://doi.org/10.2514/1.J060299>.
- [17] Spillere, A. M., Braga, D. S., Seki, L. A., Bonomo, L. A., Cordioli, J. A., Rocamora Jr, B. M., Greco Jr, P. C., dos Reis, D. C., and Coelho, E. L., “Design of a single degree of freedom acoustic liner for a fan noise test rig,” *International Journal of Aeroacoustics*, Vol. 20, No. 5-7, 2021, pp. 708–736. <https://doi.org/10.1177/1475472X211023831>.
- [18] Buot de l’Épine, Y., Chazot, J.-D., and Ville, J.-M., “Bayesian identification of acoustic impedance in treated ducts,” *The Journal of the Acoustical Society of America*, Vol. 138, No. 1, 2015, pp. EL114–EL119. <https://doi.org/10.1121/1.4923013>.
- [19] Roncen, R., Mery, F., Piot, E., and Simon, F., “Statistical inference method for liner impedance education with a shear grazing flow,” *AIAA Journal*, Vol. 57, No. 3, 2019, pp. 1055–1065. <https://doi.org/10.2514/1.J057559>.
- [20] Soize, C., *Uncertainty quantification*, Springer, 2017. <https://doi.org/10.1007/978-3-319-54339-0>.
- [21] Soize, C., and Ghanem, R., “Data-driven probability concentration and sampling on manifold,” *Journal of Computational Physics*, Vol. 321, 2016, p. 242–258. <https://doi.org/10.1016/j.jcp.2016.05.044>.
- [22] Soize, C., and Ghanem, R., “Probabilistic Learning on Manifolds,” *Foundations of Data Science*, Vol. 2, No. 3, 2020, pp. 279–307. <https://doi.org/10.3934/fods.2020013>.
- [23] Soize, C., and Ghanem, R., “Probabilistic learning on manifolds (PLoM) with partition,” *International Journal for Numerical Methods in Engineering*, Vol. 123, No. 1, 2022, pp. 268–290. <https://doi.org/10.1002/nme.6856>.
- [24] Panton, R. L., and Goldman, A. L., “Correlation of nonlinear orifice impedance,” *The Journal of the Acoustical Society of America*, Vol. 60, No. 6, 1976, pp. 1390–1396. <https://doi.org/10.1121/1.381232>.
- [25] Palani, S., Murray, P., McAlpine, A., and Richter, C., “Optimisation of slanted septum core and multiple folded cavity acoustic liners for aero-engines,” *AIAA Aviation 2021 Forum*, 2021, pp. 2021–2172. <https://doi.org/10.2514/6.2021-2172>.

- [26] Palani, S., Murray, P., McAlpine, A., Sasaki, D., and Richter, C., “Slanted septum and multiple folded cavity liners for broadband sound absorption,” *International Journal of Aeroacoustics*, Vol. 20, No. 5-7, 2021, pp. 633–661. <https://doi.org/10.1177/1475472X211023835>.
- [27] Palani, S., Murray, P., McAlpine, A., Knepper, K. M., and Richter, C., “Assessment of novel acoustic liners for aero-engine applications with sheared mean flow,” *28th AIAA/CEAS Aeroacoustics 2022 Conference*, 2022, pp. 2022–2900. <https://doi.org/10.2514/6.2022-2900>.
- [28] “PlotDigitizer: Version 3.1.5,” 2023. URL <https://plotdigitizer.com>.

# SUPPLEMENTAL MATERIAL

## Aeroacoustic Liner Impedance Metamodel from Simulation and Experimental Data Using Probabilistic Learning

Amritesh Sinha<sup>\*</sup>, Christian Soize<sup>†</sup> and Christophe Desceliers<sup>‡</sup>  
*Université Gustave Eiffel, MSME UMR 8208 (Marne-la-Vallée, France)*

Guilherme Cunha<sup>§</sup>  
*Airbus Operations SAS (Toulouse, France)*

### I. Introduction

This Supplemental Material is a companion paper that presents Appendix A entitled: *Summary of the probabilistic learning on manifolds (PLoM) algorithm and its parameterization* and Appendix B entitled: *Algebraic expressions of the conditional statistics*.

### II. Appendix A: Summary of the probabilistic learning on manifolds (PLoM) algorithm and its parameterization

The PLoM approach [1–3], which has specifically been developed for small data (as opposed to big data) starts from a training set  $\mathcal{D}_d$  made up of a relatively small number  $N_d$  of points. It is assumed that  $\mathcal{D}_d$  is generated with an underlying stochastic manifold related to a  $\mathbb{R}^n$ -valued random variable  $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ , defined on a probability space  $(\Theta, \mathcal{T}, \mathcal{P})$ , in which  $\mathbf{Q}$  is the quantity of interest that is a  $\mathbb{R}^{n_q}$ -random variable, where  $\mathbf{W}$  is the control parameter that is a  $\mathbb{R}^{n_w}$ -random variable, and where  $n = n_q + n_w$ . Another  $\mathbb{R}^{n_u}$ -valued random variable  $\mathbf{U}$  defined on  $(\Theta, \mathcal{T}, \mathcal{P})$  can also be considered, which is an uncontrolled parameter and/or a noise. Random variable  $\mathbf{Q}$  is assumed to be written as  $\mathbf{Q} = \mathbf{f}(\mathbf{U}, \mathbf{W})$  in which the measurable mapping  $\mathbf{f}$  is not explicitly known. The joint probability distribution  $P_{\mathbf{W}, \mathbf{U}}(d\mathbf{w}, d\mathbf{u})$  of  $\mathbf{W}$  and  $\mathbf{U}$  is assumed to be given. The non-Gaussian probability measure  $P_{\mathbf{X}}(\mathbf{x}) = P_{\mathbf{Q}, \mathbf{W}}(d\mathbf{q}, d\mathbf{w})$  of  $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$  is concentrated in a region of  $\mathbb{R}^n$  for which the only available information is the cloud of the points of training set

---

<sup>\*</sup>PhD student, Univ Gustave Eiffel, MSME UMR 8208, 5, Boulevard Descartes, 77454, Marne-la-Vallée Cedex 2; amritesh.sinha@univ-eiffel.fr.

<sup>†</sup>Professor, Univ Gustave Eiffel, MSME UMR 8208, 5, Boulevard Descartes, 77454, Marne-la-Vallée Cedex 2; christian.soize@univ-eiffel.fr.

<sup>‡</sup>Professor, Univ Gustave Eiffel, MSME UMR 8208, 5, Boulevard Descartes, 77454, Marne-la-Vallée Cedex 2; christophe.desceliers@univ-eiffel.fr.

<sup>§</sup>Engineer, Propulsion Physics Advanced Methods, Data Integration & Advanced Methods (1PGM); guilherme.coelho-cunha@airbus.com

$\mathcal{D}_d$ . The PLoM method makes it possible to generate the learned set  $\mathcal{D}_{\text{ar}}$  for  $\mathbf{X}$  whose  $n_{\text{mc}} \gg N_d$  points (learned realizations) are generated by the non-Gaussian probability measure that is estimated using the training set. The concentration of the probability measure is preserved thanks to the use of a diffusion-maps basis that allows to enrich the available information from the training set. Using the learned set  $\mathcal{D}_{\text{ar}}$ , PLoM allows for carrying out any conditional statistics such as  $\mathbf{w} \mapsto E\{\xi(\mathbf{Q})|\mathbf{W} = \mathbf{w}\}$  from  $C_w$  in  $\mathbb{R}^{n_\xi}$ , in which  $\xi$  is a given measurable mapping from  $\mathbb{R}^{n_q}$  into  $\mathbb{R}^{n_\xi}$ , that is to say to construct statistical surrogate models (metamodels) in a probabilistic framework.

The training dataset  $\mathcal{D}_d$  is made up of the  $N_d$  independent realizations  $\mathbf{x}_d^j = (\mathbf{q}_d^j, \mathbf{w}_d^j)$  in  $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$  for  $j \in \{1, \dots, N_d\}$  of random variable  $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ . The PLoM method allows for generating the learned dataset  $\mathcal{D}_{\text{ar}}$  made up of  $N_{\text{ar}} \gg N_d$  learned realizations  $\{\mathbf{x}_{\text{ar}}^\ell, \ell = 1, \dots, N_{\text{ar}}\}$  of random vector  $\mathbf{X}$ . As soon as the learned dataset has been constructed, the learned realizations for  $\mathbf{Q}$  and  $\mathbf{W}$  can be extracted as  $(\mathbf{q}_{\text{ar}}^\ell, \mathbf{w}_{\text{ar}}^\ell) = \mathbf{x}_{\text{ar}}^\ell$  for  $\ell = 1, \dots, N_{\text{ar}}$ .

(A.1) *Reduced representation.* The  $N_d$  independent realizations  $\{\mathbf{x}_d^j, j = 1, \dots, N_d\}$  are represented by the matrix  $[x_d] = [\mathbf{x}_d^1 \dots \mathbf{x}_d^{N_d}]$  in  $\mathbb{M}_{n, N_d}$ . Let  $[\mathbf{X}] = [\mathbf{X}^1, \dots, \mathbf{X}^{N_d}]$  be the random matrix with values in  $\mathbb{M}_{n, N_d}$ , whose columns are  $N_d$  independent copies of random vector  $\mathbf{X}$ . Using the PCA of  $\mathbf{X}$ , random matrix  $[\mathbf{X}]$  is written as,

$$[\mathbf{X}] = [\underline{x}] + [\varphi] [\mu]^{1/2} [\mathbf{H}], \quad (1)$$

in which  $[\mathbf{H}] = [\mathbf{H}^1, \dots, \mathbf{H}^{N_d}]$  is a  $\mathbb{M}_{\nu, N_d}$ -valued random matrix, where  $\nu \leq n$ , and where  $[\mu]$  is the  $(\nu \times \nu)$  diagonal matrix of the  $\nu$  positive eigenvalues of the empirical estimate of the covariance matrix of  $\mathbf{X}$ . The  $(n \times \nu)$  matrix  $[\varphi]$  is made up of the associated eigenvectors such  $[\varphi]^T [\varphi] = [I_\nu]$ . The matrix  $[\underline{x}]$  in  $\mathbb{M}_{n, N_d}$  has identical columns, each one being equal to the empirical estimate  $\underline{\mathbf{x}} \in \mathbb{R}^n$  of the mean value of random vector  $\mathbf{X}$ . The columns of  $[\mathbf{H}]$  are  $N_d$  independent copies of a random vector  $\mathbf{H}$  with values in  $\mathbb{R}^\nu$ . The realization  $[\eta_d] = [\eta_d^1 \dots \eta_d^{N_d}] \in \mathbb{M}_{\nu, N_d}$  of  $[\mathbf{H}]$  is computed by  $[\eta_d] = [\mu]^{-1/2} [\varphi]^T ([x_d] - [\underline{x}])$ . The value  $\nu$  is classically calculated in order that the  $L^2$ - error function  $\nu \mapsto \text{err}_{\mathbf{X}}(\nu)$  defined by

$$\text{err}_{\mathbf{X}}(\nu) = 1 - \frac{\sum_{\alpha=1}^{\nu} \mu_\alpha}{E\{\|\mathbf{X}\|^2\}}, \quad (2)$$

be smaller than  $\varepsilon_{\text{PCA}}$ . If  $\nu < n$ , then there is a statistical reduction.

(A.2) *Construction of a reduced-order diffusion-maps basis.* For preserving the concentration of the learned realizations in the region in which the points of the training dataset are concentrated, the PLoM relies on the diffusion-maps method [4, 5]. This is an algebraic basis of vector space  $\mathbb{R}^{N_d}$ , which is constructed using the diffusion maps. Let  $[K]$  and  $[b]$  be the matrices such that, for all  $i$  and  $j$  in  $\{1, \dots, N_d\}$ ,  $[K]_{ij} = \exp\{-(4 \varepsilon_{\text{DM}})^{-1} \|\eta_d^i - \eta_d^j\|^2\}$  and  $[b]_{ij} = \delta_{ij} b_i$  with  $b_i = \sum_{j=1}^{N_d} [K]_{ij}$ , in which  $\varepsilon_{\text{DM}} > 0$  is a smoothing parameter. The eigenvalues  $\lambda_1, \dots, \lambda_{N_d}$  and the associated eigenvectors  $\psi^1, \dots, \psi^{N_d}$  of the right-eigenvalue problem  $[\mathbb{P}] \psi^\alpha = \lambda_\alpha \psi^\alpha$  are such that  $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{N_d}$  and are computed by solving the generalized eigenvalue problem  $[K] \psi^\alpha = \lambda_\alpha [b] \psi^\alpha$  with the normalization  $\langle [b] \psi^\alpha, \psi^\beta \rangle = \delta_{\alpha\beta}$ . The eigenvector  $\psi^1$  associated with  $\lambda_1 = 1$  is a constant vector. For a given integer  $\kappa \geq 0$ , the diffusion-maps basis  $\{\mathbf{g}^1, \dots, \mathbf{g}^\alpha, \dots, \mathbf{g}^{N_d}\}$  is a vector basis of  $\mathbb{R}^{N_d}$  defined by  $\mathbf{g}^\alpha = \lambda_\alpha^\kappa \psi^\alpha$ . For a given integer  $m$ , the reduced-order diffusion-maps basis of order  $m$  is

defined as the family  $\{\mathbf{g}^1, \dots, \mathbf{g}^m\}$  that is represented by the matrix  $[g_m] = [\mathbf{g}^1 \dots \mathbf{g}^m] \in \mathbb{M}_{N_d, m}$  with  $\mathbf{g}^\alpha = (g_1^\alpha, \dots, g_{N_d}^\alpha)$  and  $[g_m]_{\ell\alpha} = g_\ell^\alpha$ . This basis depends on two parameters,  $\varepsilon_{\text{DM}}$  and  $m$ , which have to be identified. It is proven in [2], that the PLoM method does not depend on  $\kappa$  that can therefore be chosen to 0.

We have to find the optimal value  $m_{\text{opt}} \leq N_d$  of  $m$  and the smallest value  $\varepsilon_{\text{opt}} > 0$  of  $\varepsilon_{\text{DM}}$  such that (see [6])

$$1 = \lambda_1 > \lambda_2(\varepsilon_{\text{opt}}) \simeq \dots \simeq \lambda_{m_{\text{opt}}}(\varepsilon_{\text{opt}}) \gg \lambda_{m_{\text{opt}}+1}(\varepsilon_{\text{opt}}) \geq \dots \geq \lambda_{N_d}(\varepsilon_{\text{opt}}) > 0, \quad (3)$$

with an amplitude jump equal to an order of magnitude (a factor 10 as demonstrated in [2]) between  $\lambda_{m_{\text{opt}}}(\varepsilon_{\text{opt}})$  and  $\lambda_{m_{\text{opt}}+1}(\varepsilon_{\text{opt}})$ . A further in-depth analysis makes it possible to state the following algorithm to estimate  $\varepsilon_{\text{opt}}$  and  $m_{\text{opt}}$ . Let  $\varepsilon_{\text{DM}} \mapsto \text{Jump}(\varepsilon_{\text{DM}})$  be the function on  $]0, +\infty[$  defined by

$$\text{Jump}(\varepsilon_{\text{DM}}) = \lambda_{m_{\text{opt}}+1}(\varepsilon_{\text{DM}}) / \lambda_2(\varepsilon_{\text{DM}}). \quad (4)$$

The algorithm is the following:

- set the value of  $m$  to  $m_{\text{opt}} = \nu + 1$ ;
- identify the smallest possible value  $\varepsilon_{\text{opt}}$  of  $\varepsilon_{\text{DM}}$  in order that  $\text{Jump}(\varepsilon_{\text{opt}}) \leq 0.1$  and such that Equation (3) be verified.

(A.3) *Reduced-order representation of random matrices  $[\mathbf{H}]$  and  $[\mathbf{X}]$ .* The diffusion-maps vectors  $\mathbf{g}^1, \dots, \mathbf{g}^m \in \mathbb{R}^{N_d}$  span a subspace of  $\mathbb{R}^{N_d}$  that characterizes, for the optimal values  $m_{\text{opt}}$  and  $\varepsilon_{\text{opt}}$  of  $m$  and  $\varepsilon_{\text{DM}}$ , the local geometry structure of dataset  $\{\boldsymbol{\eta}_d^j, j = 1, \dots, N_d\}$ . So the PLoM method introduces the  $\mathbb{M}_{\nu, N_d}$ -valued random matrix  $[\mathbf{H}_m] = [\mathbf{Z}_m] [g_m]^T$  with  $m \leq N_d$ , corresponding to a data-reduction representation of random matrix  $[\mathbf{H}]$ , in which  $[\mathbf{Z}_m]$  is a  $\mathbb{M}_{\nu, m}$ -valued random matrix. The MCMC generator of random matrix  $[\mathbf{Z}_m]$  belongs to the class of Hamiltonian Monte Carlo methods, is explicitly described in [1], and is mathematically detailed in Theorem 6.3 of [2]. For generating the learned dataset, the best probability measure of  $[\mathbf{H}_m]$  is obtained for  $m = m_{\text{opt}}$  and using the previously defined  $[g_{m_{\text{opt}}}]$ . For these optimal quantities  $m_{\text{opt}}$  and  $[g_{m_{\text{opt}}}]$ , the generator allows for computing  $n_{\text{MC}}$  realizations  $\{[\mathbf{z}_{\text{ar}}^\ell], \ell = 1, \dots, n_{\text{MC}}\}$  of  $[\mathbf{Z}_{m_{\text{opt}}}]$  and therefore, for deducing the  $n_{\text{MC}}$  realizations  $\{[\boldsymbol{\eta}_{\text{ar}}^\ell], \ell = 1, \dots, n_{\text{MC}}\}$  of  $[\mathbf{H}_{m_{\text{opt}}}]$ . The reshaping of matrix  $[\boldsymbol{\eta}_{\text{ar}}^\ell] \in \mathbb{M}_{\nu, N_d}$  allows for obtaining  $N_{\text{ar}} = n_{\text{MC}} \times N_d$  learned realizations  $\{\boldsymbol{\eta}_{\text{ar}}^{\ell'}, \ell' = 1, \dots, N_{\text{ar}}\}$  of  $\mathbf{H}$ . These learned realizations allow for estimating converged statistics on  $\mathbf{H}$  and then on  $\mathbf{X}$ , such as pdf, moments, or conditional expectation of the type  $E\{\boldsymbol{\xi}(\mathbf{Q}) | \mathbf{W} = \mathbf{w}\}$  for  $\mathbf{w}$  given in  $\mathbb{R}^{n_w}$  and for any given vector-valued function  $\boldsymbol{\xi}$  defined on  $\mathbb{R}^{n_q}$ .

(A.4) *Criterion for quantifying the concentration of the probability measure of random matrix  $[\mathbf{H}_{m_{\text{opt}}}]$ .* For  $m \leq N_d$ , the concentration of the probability measure of random matrix  $[\mathbf{H}_m]$  is defined (see [2]) by

$$d_{N_d}^2(m) = E\{\|[\mathbf{H}_m] - [\eta_d]\|^2\} / \|[\eta_d]\|^2. \quad (5)$$

Let  $\mathcal{M}_{\text{opt}} = \{m_{\text{opt}}, m_{\text{opt}} + 1, \dots, N_d\}$  in which  $m_{\text{opt}}$  is the optimal value of  $m$  previously defined. Theorem 7.8 of [2] shows that  $\min_{m \in \mathcal{M}_{\text{opt}}} d_{N_d}^2(m) \leq 1 + m_{\text{opt}} / (N_d - 1) < d_{N_d}^2(N_d)$ , which means that the PLoM method, for  $m = m_{\text{opt}}$  and  $[g_{m_{\text{opt}}}]$  is a better method than the usual one corresponding to  $d_{N_d}^2(N_d) = 1 + N_d / (N_d - 1) \simeq 2$ . Using the  $n_{\text{MC}}$  realizations  $\{[\boldsymbol{\eta}_{\text{ar}}^\ell], \ell = 1, \dots, n_{\text{MC}}\}$  of  $[\mathbf{H}_{m_{\text{opt}}}]$ , we have the estimate  $d_{N_d}^2(m_{\text{opt}}) \simeq (1/n_{\text{MC}}) \sum_{\ell=1}^{n_{\text{MC}}} \{\|[\boldsymbol{\eta}_{\text{ar}}^\ell] - [\eta_d]\|^2\} / \|[\eta_d]\|^2$ .



(A.5) *Generation of learned realizations*  $\{\boldsymbol{\eta}_{\text{ar}}^{\ell'}, \ell' = 1, \dots, N_{\text{ar}}\}$  of random vector  $\mathbf{H}$ . The MCMC generator is detailed in [1]. Let  $\{([\mathbf{Z}(t)], [\boldsymbol{\mathcal{Y}}(t)]), t \in \mathbb{R}^+\}$  be the unique asymptotic (for  $t \rightarrow +\infty$ ) stationary diffusion stochastic process with values in  $\mathbb{M}_{v, m_{\text{opt}}} \times \mathbb{M}_{v, m_{\text{opt}}}$ , of the following reduced-order ISDE (stochastic nonlinear second-order dissipative Hamiltonian dynamic system), for  $t > 0$ ,

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

with  $[\mathbf{Z}(0)] = [\eta_d] [a]$  and  $[\boldsymbol{\mathcal{Y}}(0)] = [\mathcal{N}] [a]$ , in which

$$[a] = [g_{m_{\text{opt}}}] ([g_{m_{\text{opt}}}]^T [g_{m_{\text{opt}}}])^{-1} \in \mathbb{M}_{N_d, m_{\text{opt}}}.$$

(1)  $[\mathcal{L}([\mathbf{Z}(t)])] = [L([\mathbf{Z}(t)] [g_{m_{\text{opt}}}]^T)] [a]$  is a random matrix with values in  $\mathbb{M}_{v, m_{\text{opt}}}$ . For all  $[u] = [\mathbf{u}^1 \dots \mathbf{u}^{N_d}]$  in  $\mathbb{M}_{v, N_d}$  with  $\mathbf{u}^j = (u_1^j, \dots, u_v^j)$  in  $\mathbb{R}^v$ , the matrix  $[L([u])]$  in  $\mathbb{M}_{v, N_d}$  is defined, for all  $k = 1, \dots, v$  and for all  $j = 1, \dots, N_d$ , by

$$\begin{aligned} [L([u])]_{kj} &= \frac{1}{p(\mathbf{u}^j)} \{\nabla_{\mathbf{u}^j} p(\mathbf{u}^j)\}_k, \\ p(\mathbf{u}^j) &= \frac{1}{N_d} \sum_{j'=1}^{N_d} \exp\left\{-\frac{1}{2\widehat{s}_v^2} \left\|\frac{\widehat{s}_v}{s_v} \boldsymbol{\eta}^{j'} - \mathbf{u}^j\right\|^2\right\}, \\ \nabla_{\mathbf{u}^j} p(\mathbf{u}^j) &= \frac{1}{\widehat{s}_v^2 N_d} \sum_{j'=1}^{N_d} \left(\frac{\widehat{s}_v}{s_v} \boldsymbol{\eta}^{j'} - \mathbf{u}^j\right) \exp\left\{-\frac{1}{2\widehat{s}_v^2} \left\|\frac{\widehat{s}_v}{s_v} \boldsymbol{\eta}^{j'} - \mathbf{u}^j\right\|^2\right\}, \end{aligned} \tag{6}$$

in which  $\widehat{s}_v$  is the modified Silverman bandwidth  $s_v$ , which has been introduced in [7],

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

(2)  $[\mathbf{W}^{\text{wien}}(t)] = [\mathbb{W}^{\text{wien}}(t)] [a]$  where  $\{[\mathbb{W}^{\text{wien}}(t)], t \in \mathbb{R}^+\}$  is the  $\mathbb{M}_{v, N_d}$ -valued normalized Wiener process.

(3)  $[\mathcal{N}]$  is the  $\mathbb{M}_{v, N_d}$ -valued normalized Gaussian random matrix that is independent of process  $[\mathbb{W}^{\text{wien}}]$ .

(4) The free parameter  $f_0$ , such that  $0 < f_0 < 4/\widehat{s}_v$ , allows the dissipation term of the nonlinear second-order dynamic system (dissipative Hamiltonian system) to be controlled in order to kill the transient part induced by the initial conditions. A common value is  $f_0 = 4$  (note that  $\widehat{s}_v < 1$ ).

(5) We then have  $[\mathbf{Z}_{m_{\text{opt}}}] = \lim_{t \rightarrow +\infty} [\mathbf{Z}(t)]$  in probability distribution. The Störmer-Verlet scheme is used for solving the reduced-order ISDE, which allows for generating the learned realizations,  $[z_{\text{ar}}^1], \dots, [z_{\text{ar}}^{n_{\text{MC}}}]$ , and then, generating the learned realizations  $[\eta_{\text{ar}}^1], \dots, [\eta_{\text{ar}}^{n_{\text{MC}}}]$  such that  $[\eta_{\text{ar}}^{\ell'}] = [z_{\text{ar}}^{\ell'}] [g_{m_{\text{opt}}}]^T$ .

(6) The learned realizations  $\{\mathbf{x}_{\text{ar}}^{\ell'}, \ell' = 1, \dots, N_{\text{ar}}\}$  of random vector  $\mathbf{X}$  are then calculated (see Eq. (1)) by  $\mathbf{x}_{\text{ar}}^{\ell'} = \underline{\mathbf{x}} + [\varphi] [\mu]^{1/2} \boldsymbol{\eta}_{\text{ar}}^{\ell'}$ .

(A.6) *Constraints on the second-order moments of the components of  $\mathbf{H}$ .* In general, the mean value of  $\mathbf{H}$  estimated using the  $N_{\text{ar}}$  learned realizations  $\{\boldsymbol{\eta}_{\text{ar}}^{\ell'}, \ell' = 1, \dots, N_{\text{ar}}\}$ , is sufficiently close to zero. Likewise, the estimate of the covariance matrix of  $\mathbf{H}$ , which must be the identity matrix, is sufficiently close to a diagonal matrix. However, sometimes the diagonal entries of the estimated covariance matrix can be lower than 1. Normalization can be recovered by imposing constraints

$$\{E\{(H_k)^2\} = 1, k = 1, \dots, \nu\},$$

in the algorithm presented in paragraph (v). For that, we use the method and the iterative algorithm presented in [6] (that is based on Sections 5.5 and 5.6 of [3]). The constraints are imposed by using the Kullback-Leibler minimum cross-entropy principle. The resulting optimization problem is formulated using a Lagrange multiplier  $\mathbf{v} = (v_1, \dots, v_\nu)$  associated with the constraints. The optimal solution of the Lagrange multiplier is computed using an efficient iterative algorithm. At each iteration, the MCMC generator detailed in paragraph A.5 is used. The constraints are rewritten as

$$E\{\mathbf{h}(\mathbf{H})\} = \mathbf{b},$$

in which the function  $\mathbf{h} = (h_1, \dots, h_\nu)$  and the vector  $\mathbf{b} = (b_1, \dots, b_\nu)$  are such that  $h_k(\mathbf{H}) = (H_k)^2$  and  $b_k = 1$  for  $k$  in  $\{1, \dots, \nu\}$ . To take into account the constraints in the algorithm of paragraph A.5, Eq. (6) is replaced by the following one,

$$[L_\lambda([u])]_{kj} = \frac{1}{p(\mathbf{u}^j)} \{\nabla_{\mathbf{u}^j} p(\mathbf{u}^j)\}_k - 2\lambda_k u_k^j.$$

The iteration algorithm for computing  $\boldsymbol{\lambda}^{i+1}$  as a function of  $\boldsymbol{\lambda}^i$  is the following,

$$\begin{aligned} \boldsymbol{\lambda}^{i+1} &= \boldsymbol{\lambda}^i - \alpha_i [\Gamma''(\boldsymbol{\lambda}^i)]^{-1} \boldsymbol{\Gamma}'(\boldsymbol{\lambda}^i) \quad , \quad i \geq 0, \\ \boldsymbol{\lambda}^0 &= \mathbf{0}_\nu, \end{aligned} \tag{7}$$

in which  $\boldsymbol{\Gamma}'(\boldsymbol{\lambda}^i) = \mathbf{b} - E\{\mathbf{h}(\mathbf{H}_{\boldsymbol{\lambda}^i})\}$  and  $[\Gamma''(\boldsymbol{\lambda}^i)] = [\text{cov}\{\mathbf{h}(\mathbf{H}_{\boldsymbol{\lambda}^i})\}]$  (the covariance matrix), and where  $\alpha_i$  is a relaxation function (less than 1) that is introduced for controlling the convergence as a function of iteration number  $i$ . For given  $i_2 \geq 2$ , for given  $\beta_1$  and  $\beta_2$  such that  $0 < \beta_1 < \beta_2 \leq 1$ ,  $\alpha_i$  can be defined by:

- for  $i \leq i_2$ ,  $\alpha_i = \beta_1 + (\beta_2 - \beta_1)(i - 1)/(i_2 - 1)$ ;
- for  $i > i_2$ ,  $\alpha_i = \beta_2$ .

The convergence of the iteration algorithm is controlled by the error function  $i \mapsto \text{err}(i)$  defined by

$$\text{err}(i) = \|\mathbf{b} - E\{\mathbf{h}(\mathbf{H}_{\boldsymbol{\lambda}^i})\}\| / \|\mathbf{b}\|. \tag{8}$$

At each iteration  $i$ ,  $E\{\mathbf{h}(\mathbf{H}_{\boldsymbol{\lambda}^i})\}$  and  $[\text{cov}\{\mathbf{h}(\mathbf{H}_{\boldsymbol{\lambda}^i})\}]$  are estimated by using the  $N_{\text{ar}}$  learned realizations of  $\mathbf{H}_{m_{\text{opt}}}(\boldsymbol{\lambda}^i)$  obtained by reshaping the learned realizations.

### III. Appendix B: Algebraic expressions of the conditional statistics

The objective is to give the algebraic expressions of the conditional expectation,

$$E\{\mathbf{Q}|\mathbf{W} = \mathbf{w}_o\} = \int_{\mathbb{R}^{n_q}} \mathbf{q} p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w}_o) d\mathbf{q}, \quad (9)$$

and of the conditional cumulative distribution function,

$$\text{Proba}\{Q_i \leq q_i^* | \mathbf{W} = \mathbf{w}_o\} = \int_{-\infty}^{q_i^*} p_{Q_i|\mathbf{W}}(q_i^* | \mathbf{w}_o) dq_i, \quad (10)$$

which allows to estimate the confidence region. Let  $\tilde{\mathbf{Q}} = (\tilde{Q}_1, \dots, \tilde{Q}_{n_q})$  and  $\tilde{\mathbf{W}} = (\tilde{W}_1, \dots, \tilde{W}_{n_w})$  be the normalized random vector whose components are defined by

$$\tilde{Q}_i = (Q_i - \underline{q}_i)/\sigma_{Q_i}, \quad \tilde{W}_k = (W_k - \underline{w}_k)/\sigma_{W_k}, \quad (11)$$

in which  $\underline{q}_i$ ,  $\underline{w}_k$ , and  $\sigma_{Q_i}$ ,  $\sigma_{W_k}$  are the mean values and the standard deviations of the random variables  $Q_i$  and  $W_k$ , which are estimated with the empirical statistical estimators using the learned realizations  $\{(\mathbf{q}_{\text{ar}}^\ell, \mathbf{w}_{\text{ar}}^\ell), \ell = 1, \dots, N_{\text{ar}}\}$ . The Gaussian KDE estimation of the joint probability distribution of  $\tilde{\mathbf{Q}}$  and  $\tilde{\mathbf{W}}$  is written as,

$$p_{\tilde{\mathbf{Q}}, \tilde{\mathbf{W}}}(\tilde{\mathbf{q}}, \tilde{\mathbf{w}}) = \frac{1}{N_{\text{ar}}} \sum_{\ell=1}^{N_{\text{ar}}} \frac{1}{(\sqrt{2\pi}s)^{n_q}} \exp\left(-\frac{1}{2s^2} \|\tilde{\mathbf{q}} - \tilde{\mathbf{q}}_{\text{ar}}^\ell\|^2\right) \frac{1}{(\sqrt{2\pi}s)^{n_w}} \exp\left(-\frac{1}{2s^2} \|\tilde{\mathbf{w}} - \tilde{\mathbf{w}}_{\text{ar}}^\ell\|^2\right), \quad (12)$$

in which  $s$  is the Silverman bandwidth given by

$$s = \left\{ \frac{4}{N_{\text{ar}}(2+n)} \right\}^{1/(n+4)}, \quad n = n_q + n_w. \quad (13)$$

The derived algebraic expression of conditional mathematical expectation of component  $Q_i$  given  $\mathbf{W} = \mathbf{w}_o$  is written as

$$E\{Q_i|\mathbf{W} = \mathbf{w}_o\} = \underline{q}_i + \sigma_{Q_i} E\{\tilde{Q}_i|\tilde{\mathbf{W}} = \tilde{\mathbf{w}}_o\}, \quad \tilde{w}_{o,k} = (w_{o,k} - \underline{w}_k)/\sigma_{W_k}, \quad (14)$$

$$E\{\tilde{Q}_i|\tilde{\mathbf{W}} = \tilde{\mathbf{w}}_o\} = \frac{\sum_{\ell=1}^{N_{\text{ar}}} \tilde{q}_{\text{ar},i}^\ell \times \exp\left(-\frac{1}{2s^2} \|\tilde{\mathbf{w}}_o - \tilde{\mathbf{w}}_{\text{ar}}^\ell\|^2\right)}{\sum_{\ell=1}^{N_{\text{ar}}} \exp\left(-\frac{1}{2s^2} \|\tilde{\mathbf{w}}_o - \tilde{\mathbf{w}}_{\text{ar}}^\ell\|^2\right)}. \quad (15)$$

The conditional cdf  $F_{Q_i|\mathbf{W}}(q_i^*|\mathbf{w}_o) = \text{Proba}\{Q_i \leq q_i^* | \mathbf{W} = \mathbf{w}_o\}$  is written as,

$$F_{Q_i|\mathbf{W}}(q_i^*|\mathbf{w}_o) = \frac{\sum_{\ell=1}^{N_{\text{ar}}} \tilde{F}_{Q_i}^\ell(\tilde{q}_i^*) \times \exp\left(-\frac{1}{2s^2} \|\tilde{\mathbf{w}}_o - \tilde{\mathbf{w}}_{\text{ar}}^\ell\|^2\right)}{\sum_{\ell=1}^{N_{\text{ar}}} \exp\left(-\frac{1}{2s^2} \|\tilde{\mathbf{w}}_o - \tilde{\mathbf{w}}_{\text{ar}}^\ell\|^2\right)}, \quad \tilde{q}_i^* = (q_i^* - \underline{q}_i)/\sigma_{Q_i}, \quad (16)$$

$$F_{Q_i|\mathbf{W}}(q_i^*|\mathbf{w}_o) = \frac{1}{2} + \frac{1}{2} \text{erf}\left(\frac{1}{\sqrt{2}s} (\tilde{q}_i^* - \tilde{q}_{\text{ar},i}^\ell)\right), \quad \text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2} dt. \quad (17)$$

## References

- [1] Soize, C., and Ghanem, R., “Data-driven probability concentration and sampling on manifold,” *Journal of Computational Physics*, Vol. 321, 2016, p. 242–258. <https://doi.org/10.1016/j.jcp.2016.05.044>.
- [2] Soize, C., and Ghanem, R., “Probabilistic Learning on Manifolds,” *Foundations of Data Science*, Vol. 2, No. 3, 2020, pp. 279–307. <https://doi.org/10.3934/fods.2020013>.
- [3] Soize, C., and Ghanem, R., “Physics-constrained non-Gaussian probabilistic learning on manifolds,” *International Journal for Numerical Methods in Engineering*, Vol. 121, No. 1, 2020, pp. 110–145. <https://doi.org/10.1002/nme.6202>.
- [4] Coifman, R., and Lafon, S., “Diffusion maps,” *Applied and Computational Harmonic Analysis*, Vol. 21, No. 1, 2006, pp. 5–30. <https://doi.org/10.1016/j.acha.2006.04.006>.
- [5] Lafon, S., and Lee, A. B., “Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization,” *IEEE transactions on pattern analysis and machine intelligence*, Vol. 28, No. 9, 2006, pp. 1393–1403. <https://doi.org/10.1109/TPAMI.2006.184>.
- [6] Soize, C., and Ghanem, R., “Probabilistic learning on manifolds (PLOM) with partition,” *International Journal for Numerical Methods in Engineering*, Vol. 123, No. 1, 2022, pp. 268–290. <https://doi.org/https://doi.org/10.1002/nme.6856>.
- [7] Soize, C., “Polynomial chaos expansion of a multimodal random vector,” *SIAM-ASA Journal on Uncertainty Quantification*, Vol. 3, No. 1, 2015, pp. 34–60. <https://doi.org/10.1137/140968495>.