Machine learning for detecting structural changes from dynamic monitoring using the probabilistic learning on manifolds

Christian Soize, André Orcesi

Machine learning for detecting structural changes from dynamic monitoring using the probabilistic learning on manifolds

Christian Soize[a] and André Orcesi[b]

[a]Université Gustave Eiffel, MSME UMR 8208, 5 Boulevard Descartes, F-77454 Marne-la-Vallée, France
[b]Université Gustave Eiffel, MAST-EMGCU, Cité Descartes, Boulevard Newton, F-77447 Marne-la-Vallée, France

ARTICLE HISTORY
Compiled July 29, 2020

ABSTRACT
This paper presents a machine learning approach for detecting structural stiffness changes of civil engineering structures considered as dynamical systems, using only an experimental database constituted of a small number of records related to the experimental first eigenfrequency of the structure and a set of measured temperatures. Since the number of records in the experimental database is assumed to be small, the "small data" case must be is considered and consequently, the most classical methods of machine learning, which require "big data", cannot be used. The method of the probabilistic learning on manifolds recently introduced for analyzing small data is thus used. The validation of this method is performed on a box-girder bridge for which its dynamic monitoring has generated an experimental database. The proposed approach can be used for other similar problems.

KEYWORDS
bridge; detection; dynamic monitoring; structural change; probabilistic learning; machine learning; small data; identification; detection of disorders;

1. Introduction

The detection of structural changes of civil engineering structures is often analyzed in the framework of the structural health monitoring and of the detection of disorders, which have received a great attention in the recent past (see for instance Brownjohn (2007); Brownjohn, De Stefano, Xu, Wenzel, and Aktan (2011); Farrar and Worden (2007); Patcha and Park (2007); Worden, Farrar, Manson, and Park (2007)). Aspects related to the design, the performance, the maintenance, the deterioration, and the damage detection have been considered by Frangopol (2011); Frangopol and Liu (2007); Ko and Ni (2005); Pandey, Yuan, and Van Noortwijk (2009), Kim and Frangopol (2011); Okasha, Frangopol, and Orcesi (2012); Orcesi and Frangopol (2010); Orcesi, Frangopol, and Kim (2010), J. P. Santos, Crémona, Orcesi, and Silveira (2013); J. P. Santos, Orcesi, Crémona, and Silveira (2015); Wong (2007). The difficulties encountered for this type of problem have led to the use of probabilistic and statistics methods, artificial intelligence methods, in particular the use of neural networks in machine learning, which has also extensively been addressed for improving all these aspects, as it can be seen in Adeli (2001); Farrar and Worden (2012); Salehi and Burgueno

Corresponding Author C. Soize. Email: christian.soize@univ-eiffel.fr

(2018), in particular, for the modeling, the monitoring, and the identification (see for instance Amezquita-Sanchez and Adeli (2015); Arangio and Bontempi (2015); Dong, Celik, Catbas, O Brien, and Taylor (2020); Liu and Zhang (2020); Xu, Wu, Chen, and Yokoyama (2004)), and for the detection of structural damage, degradation, and structural disorders as analyzed by Cha, Choi, and Büyüköztürk (2017); Erfani, Rajasegarar, Karunasekera, and Leckie (2016); Feng, Liu, Kao, and Lee (2017); Gui, Pan, Lin, Li, and Yuan (2017); Hewayde, Nehdi, Allouche, and Nakhla (2007); Lin, Nie, and Ma (2017); Rafiei and Adeli (2017); A. Santos, Figueiredo, Silva, Sales, and Costa (2016); J. P. Santos, Crémona, Calado, Silveira, and Orcesi (2016); Schoefs, Yáñez-Godoy, and Lanata (2011); Strauss (2016); Tan, Thambiratnam, Chan, Gordan, and Abdul Razak (2019).

This paper deals with the detection of structural stiffness changes of civil engineering structures considered as dynamical systems, using an experimental database constituted of a small number of records related to the experimental first eigenfrequency of the structure and a set of measured temperatures. A probabilistic criterion is proposed for the detection, which is coupled with the use of a probabilistic learning method that has recently be introduced for processing small datasets (in opposite to the cases for which big datasets are available). In order to explain the difficulties of the problem treated in this paper and to explain the proposed detection method, a very brief summary of the description of the experimental database is given and will be detailed in Section 3.1. This database will be used in Sections 3.2 to 3.5 for testing and validating the method proposed. This database is related to experimental measurements of a box-girder bridge (Pont de l'Oise in France) subjected to the dynamic traffic loads and to the effects of the environment, in particular the variations of temperatures. The measures are the external and internal temperatures and the accelerations due to the dynamical responses of the structure as a function of time. The first eigenfrequency, which depends on the temperature, is obtained by experimental modal analysis of the dynamical responses induced by the traffic. A structural modification of the bridge was carried out consisting in strengthening the bridge by installing additional prestressing cables. The measurements were carried out over a period before the structural modification and also over a period after the structural modification. The outside and inside temperatures are measured by 7 sensors and noted T_1 to T_7 (see Section 3.1). The experimental database consists of $N_d = 2811$ records of the first eigenfrequency of the bridge (obtained by experimental modal analysis), noted Q, and of 9 parameters related to the temperature, noted $\mathbf{W} = (W_1, \ldots, W_9)$, the components of which being T_1 to T_7, $T_3 - T_2$, and $T_5 - T_4$. Among these 2811 records, 744 records concern the period before the structural modification and 1887 the period after the modification. Obviously, the first eigenfrequency Q depends on the temperature. During the total measurements period (including the two periods, before and after structural modification), the interval of the temperature T_6 measured by sensor number 6 located inside the box girder is $[-7.7\,, 31.4]\,^oC$ and the first eigenfrequency Q varies from 2.20 to $2.39\,Hz$.

The objective of this paper is to use this small experimental database for proposing and validating a methodology that allows for detecting the structural change using only the total experimental database, without using separately the two parts of the database associated with the two periods, before and after the structural modification. However, these two parts of the database will still be used separately, but only for the purpose of producing validation of the proposed method. No computational model of the bridge in its environment is available and the traffic is not measured. It should be noted that the objective of this paper is not centered on retrofitting techniques for post tensioned bridges. This paper proposes a novel identification method for detecting structural changes for civil engineering structures for which a small

experimental database is available. The approach that is proposed can obviously be used for other problems. In this context, the problem posed is difficult enough for two main reasons.
- The first is due to the fact that the variations of temperatures induce amplitudes of variations of the first eigenfrequency, which are of the same order of magnitude as the variation of this first eigenfrequency induced by the structural change. In addition there are other parameters of the dynamical system that are not measured and that influence the values of the first eigenfrequency. This case is thus particularly difficult for the clustering methods, the statistical methods, and the machine learning approaches based on the use of big data.
- The second reason is related to the small size of the database, which does not allow the use of classical machine learning approaches because the database is not sufficiently big. The probabilistic learning methods adapted to small data are thus candidates for solving such a problem and it is interesting to test the novel approaches.

As the database is not sufficiently big to detect the structural change and the convergence of the statistical quantities, an alternative approach is proposed consisting in using the probabilistic learning on manifolds (PLoM) that has recently been introduced by Soize and Ghanem (2016, 2020); Soize et al. (2019) in the context of computational sciences and engineering. The first paper of 2016 presents this novel PLoM method, which has specifically been developed for the small data case in contrast to the big data case. The second paper of 2019 presents a complement to the first one, which is necessary for finding the optimal value of the dimension of the diffusion-maps basis. Finally, the last one of 2020, is a complete mathematical development for the mathematical validation. These three theoretical papers proposed a novel tool in mathematical statistics for the small data case and allow for preserving the concentration of the probability measure. Extensions of the PLoM method have been proposed in Soize and Ghanem (2020a) and many developments and applications can be found in Farhat, Tezaur, Chapman, Avery, and Soize (2019); Ghanem and Soize (2018); Ghanem et al. (2019); Guilleminot and Dolbow (2020). Nevertheless, before using this learning method, it is necessary to introduce an adapted probabilistic criterion for detecting the occurrence of a structural change.

The paper is organized as follows. The approach proposed for detecting the structural change of the dynamical system is defined in Section 2: the quantity of interest (QoI) is defined as well as the parameters which control the dynamical system, the initial dataset constituted of the experimental database, and the probabilistic criterion for the detection. In this section, an estimate of the conditional probability density function is given, based on the use of the initial dataset, and it is shown how the probabilistic learning on manifolds allows for improving the initial information and for obtaining the statistical convergence. Section 3 deals with the application to the detection of structural change of a civil engineering structure for which the description of an experimental database is given. The predictions using the probabilistic learning on manifolds are presented as well as a convergence analysis. This section ends by presenting a validation of the probabilistic criterion coupled with the PLoM approach. A summary of the probabilistic learning on manifolds (PLoM) is given in Appendix A.

2. Approach for detecting the structural change of the dynamical system

2.1. Definition of the QoI, the parameters, and the initial dataset

A dynamical system is considered. It will be the box-girder bridge subjected to the traffic and to the environment loads, in particular the temperature. The QoI is the real-valued random

variable Q defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$ (it will be the first eigenfrequency of the dynamical system). The QoI depends on a random vector $\mathbf{W} = (W_1, \ldots, W_{n_w})$ that is made up of the part of the parameters that control the dynamical system (the components correspond to the temperatures introduced in Section 1), while random vector $\mathbf{U} = (U_1, \ldots, U_{n_u})$ is made up of the other part of the parameters, which are not used for controlling the system (here, $\mathbf{U}$ will not be described and is related to all the parameters of the dynamical system that influence Q). Consequently, there exits a deterministic mapping $(\mathbf{w}, \mathbf{u}) \mapsto f(\mathbf{w}, \mathbf{u})$ on $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$ with values in $\mathbb{R}$ representing the first eigenfrequency of the dynamical system. This mapping is unknown and we have,

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

The random variables $\mathbf{W}$ and $\mathbf{U}$ are defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$, are with values in $\mathbb{R}^{n_w}$ and $\mathbb{R}^{n_u}$, and are assumed to be statistically independent and non-Gaussian. The probability distributions $P_{\mathbf{W}}(d\mathbf{w})$ on $\mathbb{R}^{n_w}$ and $P_{\mathbf{U}}(d\mathbf{u})$ on $\mathbb{R}^{n_u}$ are unknown. Due to Equation (1), the probability measure $P_{Q,\mathbf{W}}(dq, d\mathbf{w})$ is concentrated in the neighborhood of the stochastic manifold defined by the random graph $\{(f(\mathbf{w}, \mathbf{U}), \mathbf{w}), \mathbf{w} \in \mathcal{S}_w\}$ in which $\mathcal{S}_w \subset \mathbb{R}^{n_w}$ is the unknown support of the probability measure $P_{\mathbf{W}}(d\mathbf{w})$. Subset $\mathcal{S}_w$ of $\mathbb{R}^{n_w}$ can be viewed as the admissible set for the values $\mathbf{w}$ of random variable $\mathbf{W}$.

Due to the presence of $\mathbf{U}$, it is consistent to assume that the joint probability distribution $P_{Q,\mathbf{W}}(dq, d\mathbf{w})$ admits a joint probability density function (pdf) $(q, \mathbf{w}) \mapsto p_{Q,\mathbf{W}}(q, \mathbf{w})$ on $\mathbb{R} \times \mathbb{R}^{n_w}$ with respect to $dq\, d\mathbf{w}$. Assuming also that $P_{\mathbf{W}}(d\mathbf{w})$ admits a pdf $\mathbf{w} \mapsto p_{\mathbf{W}}(\mathbf{w})$ on $\mathbb{R}^{n_w}$ with respect to $d\mathbf{w}$, it can be conclude that $P_Q(dq)$ admits a pdf $q \mapsto p_Q(q)$ on $\mathbb{R}$ with respect to dq,

The joint pdf $(q, \mathbf{w}) \mapsto p_{Q,\mathbf{W}}(q, \mathbf{w})$ and the mapping f are thus unknown. The only information available is a database, called the initial dataset and denoted by $\mathcal{D}_d(N_d)$, which consists of N_d independent realizations (for the bridge database, $N_d = 2811$) of the random vector $(Q, \mathbf{W})$ with values in $\mathbb{R} \times \mathbb{R}^{n_w}$,

$$\mathcal{D}_d(N_d) = \{(q_d^j, \mathbf{w}_d^j) \,,\, j = 1, \ldots, N_d\} \quad , \quad q_d^j \in \mathbb{R} \,,\, \mathbf{w}_d^j = (w_{d,1}^j, \ldots, w_{d,n_w}^j) \in \mathbb{R}^{n_w} \,. \tag{2}$$

2.2. Construction of a probabilistic criterion for the detection

In the context of detecting the structural change of the dynamical system (the box-girder bridge), the idea of constructing the detection criterion is as follows. If the first eigenfrequency Q did not depend on the temperature, then the structural change would induce a bimodality of the probability density function of random variable Q. However, as Q depends on the temperature and taking into account the characteristics of the database described in Section 1, we will proceed by conditioning Q with respect to temperatures. As it will be seen in Section 2.4, the probabilistic learning on manifolds will be performed for the random vector $(Q, \mathbf{W})$ in order to use all the available information existing in initial dataset $\mathcal{D}_d(N_d)$ in order to enrich the initial information. Nevertheless, for defining the criterion that will allow for detecting the structural change, it is not necessary to keep all the information related to the temperature sensors. Only a part of them for which Q is very sensitive are taken into account. Sensor number 6, which measures the internal temperature T_6 inside the box girder in which the prestressing cables are installed, will be chosen. In this context, one chooses k in $\{1, \ldots, n_w\}$ as the index of the component of $\mathbf{W}$ used for the detection criterion (for the application to the database, which will be carried out in Section 3, k will be chosen to 6). The conditioning of Q will then be taken with respect to component W_k of $\mathbf{W}$.

Let $\mathcal{S}_k \subset \mathbb{R}$ be the support of the pdf $w_k \mapsto p_{W_k}(w_k) = \int_{\mathbb{R}^{n_w-1}} p_{\mathbf{W}}(\mathbf{w})\, d\mathbf{w}_{-k}$ of the real-valued random variable W_k, in which the notation $d\mathbf{w}_{-k} = \otimes_{\kappa \neq k} dw_\kappa = dw_1 \dots dw_{k-1}\, dw_{k+1} \dots dw_{n_w}$ has been used. Let $q \mapsto p_{Q|W_k}(q|w_k)$ be the conditional probability density function on $\mathbb{R}$ of random variable Q given $W_k = w_k \in \mathcal{S}_k$, such that

$$p_{Q|W_k}(q|w_k) = \frac{p_{Q,W_k}(q, w_k)}{p_{W_k}(w_k)} \quad , \quad \forall\, w_k \in \mathcal{S}_k \subset \mathbb{R}\,, \tag{3}$$

in which $p_{Q,W_k}(q, w_k) = \int_{\mathbb{R}^{n_w-1}} p_{Q,\mathbf{W}}(q, \mathbf{w})\, d\mathbf{w}_{-k}$. For all w_k in $\mathcal{S}_k$, it is assumed that $q \mapsto p_{Q|W_k}(q|w_k)$ admits a unique maximum $q^{\text{LM}}(w_k)$, which is a local maximum. This hypothesis means that, for all w_k in $\mathcal{S}_k$, the subset $\mathcal{E}_k = \{q \in \mathbb{R} : q = \arg\max_{q'} p_{Q|W_k}(q'\,|w_k)\}$ is reduced to a point in $\mathbb{R}$. We then have,

$$q^{\text{LM}}(w_k) = \arg\max_{q'} p_{Q|W_k}(q'\,|w_k) \quad , \quad \forall\, w_k \in \mathcal{S}_k \subset \mathbb{R}\,. \tag{4}$$

Finally, assuming that $q^{\text{LM}}(W_k)$ is a real-valued random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$, the real-valued random variable Q^{LM} is defined by

$$Q^{\text{LM}} = q^{\text{LM}}(W_k)\,. \tag{5}$$

The N_d independent realizations $\{q^{\text{LM},1}, \dots, q^{\text{LM},N_d}\}$ of random variable Q^{LM} are then given by $q^{\text{LM},j} = q^{\text{LM}}(w^j_{d,k})$, that is,

$$q^{\text{LM},j} = \arg\max_{q'} p_{Q|W_k}(q'\,|w^j_{d,k})\ ,\ j = 1, \dots, N_d\,. \tag{6}$$

Let $q \mapsto p_{Q^{\text{LM}}}(q)$ be the estimates of the probability density function of random variable Q^{LM} carried out using the Gaussian kernel-density estimation method with the realizations $\{q^{\text{LM},1}, \dots, q^{\text{LM},N_d}\}$ (see for instance, Bowman and Azzalini (1997)).

As previously explained, the detection of a structural change of the dynamical system consists in analyzing the bimodality character of the pdf $q \mapsto p_{Q^{\text{LM}}}(q)$. Assuming that pdf $p_{Q^{\text{LM}}}$ is bimodal, $q^{\text{LM}}_{\text{max},1}$ and $q^{\text{LM}}_{\text{max},2}$ denote the two values of q for which $q \mapsto p_{Q^{\text{LM}}}(q)$ reaches its two local maxima.

2.3. Estimation of the conditional probability density function using the initial dataset

The conditional probability density function, $q \mapsto p_{Q|W_k}(q|w_k)$ for given w_k in $\mathbb{R}$, has to be estimated using computational statistics. In this section, we construct the formula based on the use of the multidimensional Gaussian kernel-density estimation method and of the realizations of the initial dataset $\mathcal{D}_d(N_d)$ defined by Eq. (2). Consequently, this conditional pdf is rewritten as $q \mapsto p_{Q|W_k}(q|w_k; N_d)$. As it has been explained in Section 1, N_d is too small for obtaining the convergence of $q \mapsto p_{Q|W_k}(q|w_k; N_d)$ with respect to N_d. Consequently, in Section 2.4, the probabilistic learning on manifolds will be used for constructing the learned dataset $\mathcal{D}_{\text{ar}}(\nu_{\text{ar}})$ with $\nu_{\text{ar}} \gg N_d$ and then, we will substitute $\mathcal{D}_d(N_d)$ by $\mathcal{D}_{\text{ar}}(\nu_{\text{ar}})$ for estimating the conditional probability density function, which will then be written as $q \mapsto p_{Q|W_k}(q|w_k; \nu_{\text{ar}})$. Throughout the paper, the subscript "ar" refers to the

additional realizations generated by the probabilistic learning on manifolds.

Let $\underline{q}_d$ and σ_d be the empirical mean value and standard deviation of random variable Q computed using the realizations $\{q_d^j\,,\, j=1,\ldots,N_d\}$. Similarly, let $\underline{w}_{d,k}$ and $\sigma_{d,k}$ be the empirical mean value and standard deviation of random variable W_k computed using the realizations $\{w_{d,k}^j\,,\, j=1,\ldots,N_d\}$. We introduce the normalized random variables $\widetilde{Q}$ and $\widetilde{W}_k$ such that

$$Q = \underline{q}_d + \sigma_d\,\widetilde{Q} \quad , \quad W_k = \underline{w}_{d,k} + \sigma_{d,k}\,\widetilde{W}_k\,. \tag{7}$$

The N_d realizations of $(\widetilde{Q},\widetilde{W}_k)$ are $\{(\widetilde{q}_d^j,\widetilde{w}_{d,k}^j)\,,\, j=1,\ldots,N_d\}$ such that

$$\widetilde{q}_d^j = (q_d^j - \underline{q}_d)/\sigma_d \quad , \quad \widetilde{w}_{d,k}^j = (w_{d,k}^j - \underline{w}_{d,k})/\sigma_{d,k}\,. \tag{8}$$

Consequently, for all real q, $\widetilde{q}$, w_k, and $\widetilde{w}_k$ such that,

$$q = \underline{q}_d + \sigma_d\,\widetilde{q} \quad , \quad w_k = \underline{w}_{d,k} + \sigma_{d,k}\,\widetilde{w}_k\,, \tag{9}$$

that value $p_{Q|W_k}(q|w_k)$ of the conditional pdf of Q given $W_k = w_k$ is written as

$$p_{Q|W_k}(q|w_k;N_d) = \frac{1}{\sigma_d}\,\frac{p_{\widetilde{Q},\widetilde{W}_k}(\widetilde{q},\widetilde{w}_k;N_d)}{p_{\widetilde{W}_k}(\widetilde{w}_k;N_d)}\,, \tag{10}$$

in which $p_{\widetilde{Q},\widetilde{W}_k}$ is the joint pdf of normalized random variables $\widetilde{Q}$ and $\widetilde{W}_k$ with respect to $d\widetilde{q}\,d\widetilde{w}_k$, and $p_{\widetilde{W}_k}$ is the pdf of normalized random variable $\widetilde{W}_k$ with respect to $d\widetilde{w}_k$. Using Equation (10) and the Gaussian kernel-density estimation method, the estimate $p_{Q|W_k}(q|w_k;N_d)$ of the conditional pdf is written as

$$p_{Q|W_k}(q|w_k;N_d) = \frac{1}{\sigma_d\sqrt{2\pi}s_{_{N_d}}}\,\frac{\sum_{j=1}^{N_d}\exp\{-\frac{1}{2s_{_{N_d}}^2}\{(\widetilde{q}-\widetilde{q}_d^j)^2+(\widetilde{w}_k-\widetilde{w}_{d,k}^j)^2\}\}}{\sum_{j=1}^{N_d}\exp\{-\frac{1}{2s_{_{N_d}}^2}(\widetilde{w}_k-\widetilde{w}_{d,k}^j)^2\}}\,, \tag{11}$$

in which $s_{_{N_d}}$ is the Silverman bandwidth that is written as,

$$s_{_{N_d}} = \left\{\frac{4}{N_d(2+2)}\right\}^{1/(2+4)} = N_d^{-1/6}\,. \tag{12}$$

Concerning the bimodality of the pdf $q \mapsto p_{Q^{\text{LM}}}(q;N_d)$ that is estimated by using Section 2.2 with Equation (11), we will denote by $q_{\text{max},1}^{\text{LM}}(N_d)$ and $q_{\text{max},2}^{\text{LM}}(N_d)$ the two values of q for which $q \mapsto p_{Q^{\text{LM}}}(q;N_d)$ reaches its two local maxima (if a second local maximum exists).

2.4. Probabilistic learning on manifolds

The probabilistic learning on manifolds (PLoM), summarized in Appendix A, is used for two reasons.
- The first is the enrichment of the available information represented by the initial set $\mathcal{D}_d(N_d)$ by including in the MCMC method complementary information related to the geometry of

the support of the probability measure $p_{Q,\mathbf{W}}(q,\mathbf{w})\,dq\,d\mathbf{w}$, which consists in projecting the MCMC generator on a diffusion-maps basis (see Appendix A).
- The second is the construction of the learned dataset $\mathcal{D}_{\text{ar}}(\nu_{\text{ar}})$ with an arbitrarily large number $\nu_{\text{ar}} \gg N_d$ of realizations, which allows converged statistics to be obtained.

The methodology consists in computing $\mathcal{D}_{\text{ar}}(\nu_{\text{ar}})$ using the algorithm summarized in Appendix A ,

$$\mathcal{D}_{\text{ar}}(\nu_{\text{ar}}) = \{(q_{\text{ar}}^{\ell}, \mathbf{w}_{\text{ar}}^{\ell})\,,\ \ell = 1,\ldots,\nu_{\text{ar}}\}\ ,\ q_{\text{ar}}^{\ell} \in \mathbb{R}\ ,\ \mathbf{w}_{\text{ar}}^{\ell} = (w_{\text{ar},1}^{\ell},\ldots,w_{\text{ar},n_w}^{\ell}) \in \mathbb{R}^{n_w}\,. \quad (13)$$

Then, the conditional probability density function $q \mapsto p_{Q|W_k}(q|w_k;\nu_{\text{ar}})$ is estimated using Eqs. (9) and (11) in which $\mathcal{D}_d(N_d)$ is replaced by $\mathcal{D}_{\text{ar}}(\nu_{\text{ar}})$, that is,

$$p_{Q|W_k}(q|w_k;\nu_{\text{ar}}) = \frac{1}{\sigma_{\text{ar}}\sqrt{2\pi}s_{\nu_{\text{ar}}}} \frac{\sum_{\ell=1}^{\nu_{\text{ar}}} \exp\{-\frac{1}{2s_{\nu_{\text{ar}}}^2}\{(\widetilde{q}-\widetilde{q}_{\text{ar}}^{\ell})^2 + (\widetilde{w}_k - \widetilde{w}_{\text{ar},k}^{\ell})^2\}\}}{\sum_{\ell=1}^{\nu_{\text{ar}}} \exp\{-\frac{1}{2s_{\nu_{\text{ar}}}^2}(\widetilde{w}_k - \widetilde{w}_{\text{ar},k}^{\ell})^2\}}\,, \quad (14)$$

in which $s_{\nu_{\text{ar}}}$ is given by Equation (12) by substituting N_d by ν_{ar} and where

$$q = \underline{q}_{\text{ar}} + \sigma_{\text{ar}}\,\widetilde{q} \quad , \quad w_k = \underline{w}_{\text{ar},k} + \sigma_{\text{ar},k}\,\widetilde{w}_k\,. \quad (15)$$

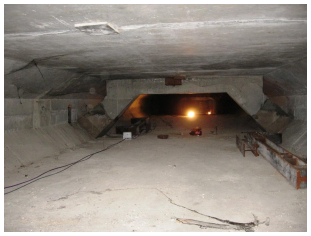
The empirical mean value $\underline{q}_{\text{ar}}$ and the standard deviation σ_{ar} of random variable Q are computed using the realizations $\{q_{\text{ar}}^{\ell}\,,\ \ell = 1,\ldots,\nu_{\text{ar}}\}$, while the empirical mean value $\underline{w}_{\text{ar},k}$ and the standard deviation $\sigma_{\text{ar},k}$ of random variable W_k are computed using the realizations $\{w_{\text{ar},k}^{\ell}\,,\ \ell = 1,\ldots,\nu_{\text{ar}}\}$.

Concerning the bimodality of the pdf $q \mapsto p_{Q^{\text{LM}}}(q;\nu_{\text{ar}})$ that is estimated by using Section 2.2 with Equation (14), we will denote by $q_{\text{max},1}^{\text{LM}}(\nu_{\text{ar}})$ and $q_{\text{max},2}^{\text{LM}}(\nu_{\text{ar}})$ the two values of q for which $q \mapsto p_{Q^{\text{LM}}}(q;\nu_{\text{ar}})$ reaches its two local maxima.

2.5. Convergence analyses

The convergence analyses are performed for the estimates of quantities $q_{\text{max},1}^{\text{LM}}(\nu_{\text{ar}})$ and $q_{\text{max},2}^{\text{LM}}(\nu_{\text{ar}})$ that characterize the bimodality of the pdf $q \mapsto p_{Q^{\text{LM}}}(q)$. Two types of convergence must be studied.
- For the maximum value $N_{d,i_{\text{max}}}$ of the number N_d of realizations in the initial dataset, the convergence is performed with respect to the number ν_{ar} of additional realizations that constitute the learned dataset, $\mathcal{D}_{\text{ar}}(\nu_{\text{ar}})$, which is constructed as explained in Section 2.4. This convergence analysis allows for identifying the optimal value $\nu_{\text{ar}}^{\text{opt}}$ of ν_{ar}.
- For ν_{ar} fixed to its optimal value $\nu_{\text{ar}}^{\text{opt}}$, the convergence is performed with respect to the number N_d of realizations in the initial dataset. We consider a sampling of integers such that $0 < N_{d,1} < N_{d,2} < \ldots < N_{d,i_{\text{max}}}$. For each i in $\{1,\ldots,i_{\text{max}}\}$, the probabilistic learning is performed as explained in Section 2.4 with the initial dataset made up of $N_{d,i}$ realizations. If the convergence is obtained for a value of N_d less than or equal to $N_{d,i_{\text{max}}}$, this means that the initial dataset contained a sufficient information for performing the learning and the learning process is converged. If not, then $N_{d,i_{\text{max}}}$ should be increased.

Figure 1. View of PI-57 Bridge.



(a) View inside box before strengthening.

(b) View after strengthening external with prestressing cables installed in the bridge deck.

Figure 2. View inside the bridge deck (a) before and (b) after strenghtening.

3. Application to the detection of structural changes of a civil engineering structure

3.1. Description of the experimental database

The PI-57 Bridge, used as illustration in this paper, is a $120\,m$ prestressed concrete box girder bridge which was built in France in the 1960s to carry the A1 motorway across the Oise River and connect Paris to Lille (Figure 1). Such box girder bridge as several others built in France by balanced cantilevering before 1975 suffered from insufficient bending strength, which led to local cracking and increase of mid-span deflection. Major reasons for these issues are failure to take thermal gradients into account, and underestimation of the redistribution of forces through the effect of creep and shrinkage of materials. Additional prestressing was applied in 2009 to strengthen the bridge (Figure 2). In parallel, a dynamic monitoring program was considered before (from November 21, 2008 to April 3, 2009) and after strengthening (from November 21, 2009 to April, 2010). Dynamic tests were performed using the traffic as source of excitation (see Cury, Cremona, and Dumoulin (2012) and Alves, Cury, Roitman, Magluta, and Cremona (2015) for further details on the monitoring program). Some accelerometers were placed inside the bridge cross-section, with sampling set to $0.004\,s$ for 5 minutes records every 3 hours over a 24-hour time period. The instrumentation scheme was composed of nineteen sensors (sixteen measuring vertical accelerations and three measuring longitudinal accelerations) instrumented over nearly $80\,m$ length span of the bridge (Figure 3). Temperature was measured at seven different locations across the bridge (see Figure 4), for the two periods before (Figure 5-(a)) and after (Figure 5-(b)) strengthening. For the first and second campaigns, a total of 744 and $1\,887$ tests were registered, respectively. In particular, Figure 6-(a) illustrates the evolution of the first eigenfrequency Q with time for the periods before and after strengthening and Figure 6-(b) shows the relation between Q and temperature T6.

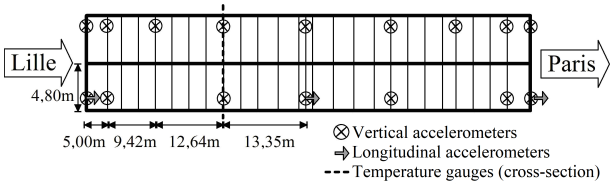


Figure 3. Location of the dynamic monitoring system.

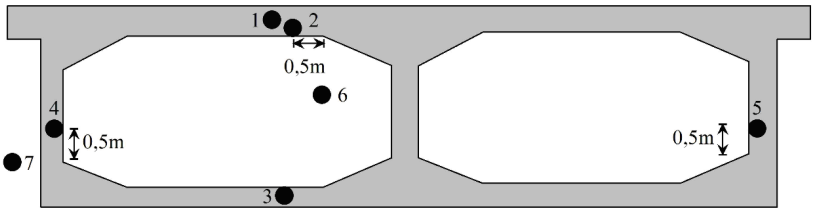
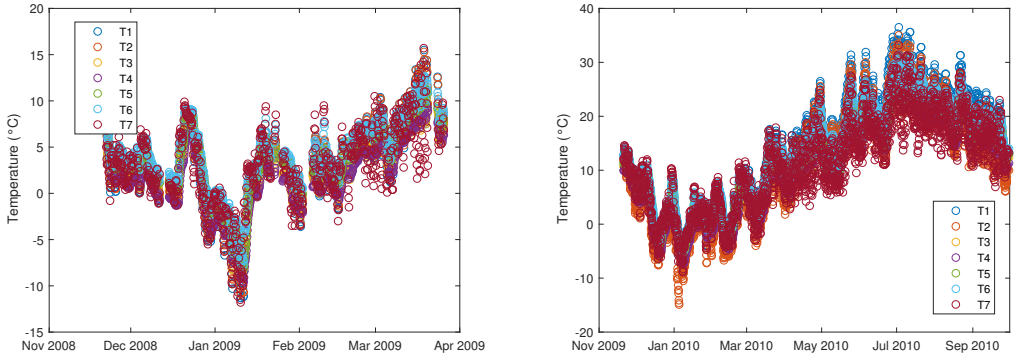


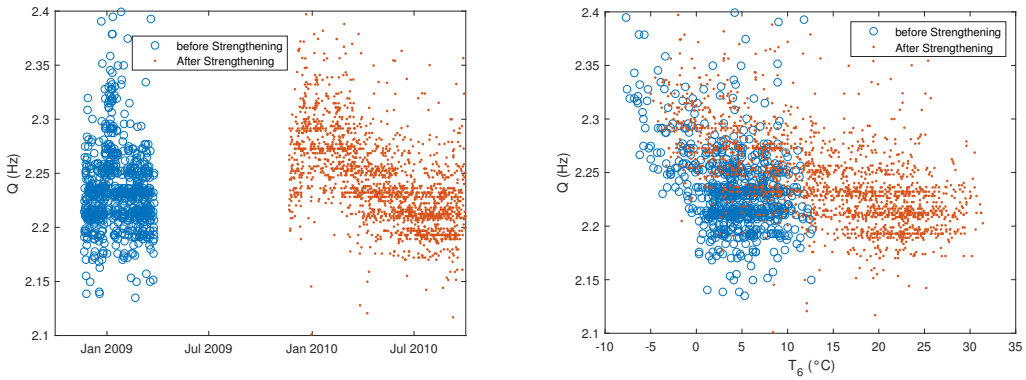
Figure 4. Location of temperature gauges.



(a) Before strengthening.

(b) After strengthening.

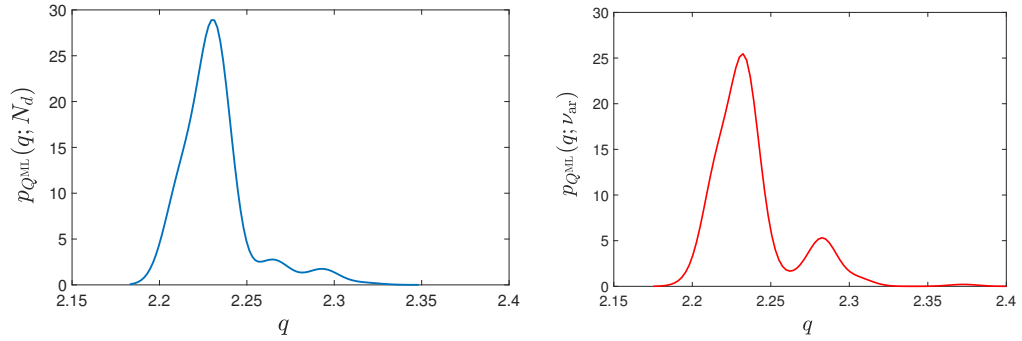
Figure 5. Time history of temperatures before and after strengthening.



(a) Profile of first eigenfrequency Q as a function of time, before and after strengthening.

(b) Profile of first eigenfrequency Q as a function of temperature T_6, before and after strengthening.

Figure 6. Profile of first eigenfrequency Q.



(a) pdf $q \mapsto p_{Q^{\mathrm{LM}}}(q;N_d)$ estimated using the initial dataset ($N_d = 2811$), without probabilistic learning.

(b) pdf $q \mapsto p_{Q^{\mathrm{LM}}}(q;\nu_{\mathrm{ar}})$ estimated using the probabilistic learning from the initial dataset ($N_d = 2811$) and with $\nu_{\mathrm{ar}} = 1\,011\,960$ additional realizations.

Figure 7. Detection of the structural change.

3.2. Predictions with the probabilistic learning on manifolds

The pdf $q \mapsto p_{Q^{\mathrm{LM}}}(q)$ defined in Section 2.2 is estimated using the additional realizations generated by using the probabilistic learning on manifolds whose algorithm is summarized in Appendix A.

(i) A first analysis is performed using the total database that constitutes the initial dataset with $N_d = 2811$ experimental measurements. Figure 7-(a) displays the graph of function $q \mapsto p_{Q^{\mathrm{LM}}}(q;N_d)$. This pdf is unimodal and the peak is obtained for the value $q^{\mathrm{LM}}_{\mathrm{max},1} = 2.23\,Hz$ of the first eigenfrequency. Consequently, the initial dataset does not allow the structural change to be detected.

(ii) A second analysis is carried out using the probabilistic learning with $\nu_{\mathrm{ar}} = 1\,011\,960$ additional realizations for which the initial dataset is the total database for which $N_d = 2811$. The values of the parameters of the PLoM defined in Appendix A are the following: $n_q = 1$; $n_w = 9$; $n = 10$; $N_d = 2811$; relative error for the PCA truncation: 10^{-6}; $\nu = 8$; $\varepsilon^{\mathrm{opt}}_{\mathrm{diff}} = 17.0$; $m = m^{\mathrm{opt}} = 9$ with $\Lambda_1 = 1$, $\Lambda_2 = 0.0297$, $\Lambda_9 = 0.0268$, and $\Lambda_{10} = 0.00296$; $n_{\mathrm{MC}} = 360$; $\nu_a r = n_{\mathrm{MC}} \times N_d = 1\,011\,960$; $s_\nu = 0.478$; $\widehat{s}_\nu = 0.431$; $f_0 = 1.5$; $M_0 = 10$; $M = 3600$; $\Delta r = 0.135506$. Figure 7-(b) displays the graph of function $q \mapsto p_{Q^{\mathrm{LM}}}(q;\nu_{\mathrm{ar}})$. This pdf is bimodal. The first peak is obtained for the value $q^{\mathrm{LM}}_{\mathrm{max},1} = 2.23\,Hz$ of the first eigenfrequency while the second peak, which allows for detecting the structural change, occurs for the value $q^{\mathrm{LM}}_{\mathrm{max},2} = 2.28\,Hz$ of the first eigenfrequency.

(iii) The convergence of the family of probability density functions $q \mapsto p_{Q^{\mathrm{LM}}}(q;\nu_{\mathrm{ar}})$ of random variable Q^{LM} with respect to the number ν_{ar} of additional realizations computed using the probabilistic learning on manifolds is shown in Figure 8. In this figure it can also be seen the graph of $q \mapsto p_{Q^{\mathrm{LM}}}(q;N_d)$ estimated with the initial dataset ($N_d = 2811$) without using the probabilistic learning (it is the graph shown in Figure 7-(a)). For $\nu_{\mathrm{ar}} = 1\,011\,960$ the convergence of the probabilistic learning is reached (it is the graph shown in Figure 7-(b)).

3.3. Convergence of the estimates with respect to the size of the learned dataset

For the maximum size $N_d = 2811$ of the initial dataset, the convergence of $q^{\mathrm{LM}}_{\mathrm{max},1}$ and $q^{\mathrm{LM}}_{\mathrm{max},2}$ is performed with respect to the size of the learned dataset, that is, with respect to the number

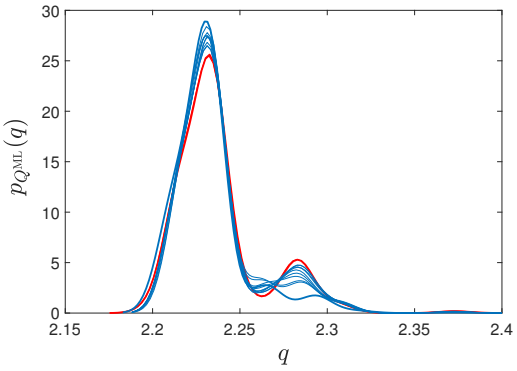
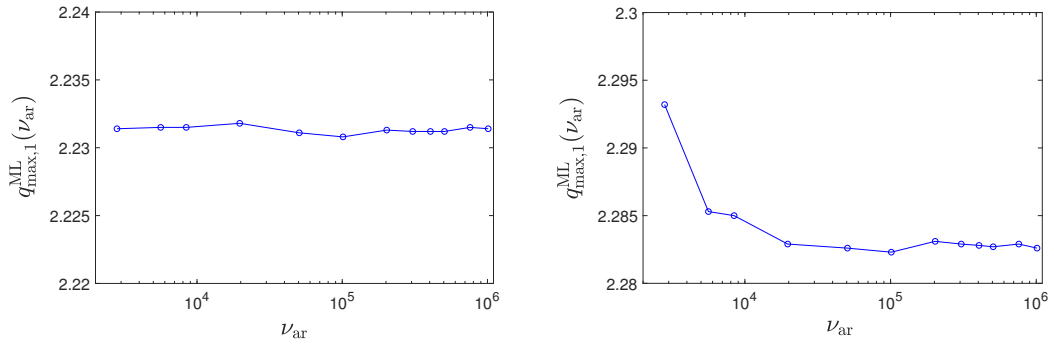


Figure 8. Convergence analysis of the pdf of Q^{LM}: graph of $q \mapsto p_{Q^{\mathrm{LM}}}(q; N_d)$ estimated without using the probabilistic learning (thick blue line) and graphs of $q \mapsto p_{Q^{\mathrm{LM}}}(q; \nu_{\mathrm{ar}})$ estimated using the probabilistic learning (thin blue line) for $\nu_{\mathrm{ar}} \in \{5\,622, \ldots, 1\,011\,960\}$; for the last value, $\nu_{\mathrm{ar}} = 1\,011\,960$, the graph is the thick red line.



(a) Convergence of $q^{\mathrm{LM}}_{\mathrm{max},1}(\nu_{\mathrm{ar}})$.

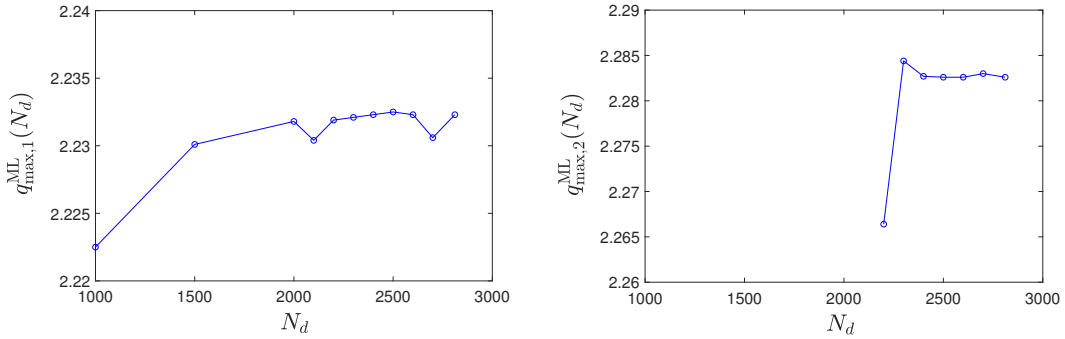
(b) Convergence of $q^{\mathrm{LM}}_{\mathrm{max},2}(\nu_{\mathrm{ar}})$.

Figure 9. Convergence of $q^{\mathrm{LM}}_{\mathrm{max},1}(\nu_{\mathrm{ar}})$ and $q^{\mathrm{LM}}_{\mathrm{max},2}(\nu_{\mathrm{ar}})$ with respect to the number of additional realizations ν_{ar} of the probabilistic learning, for the initial dataset of size $N_d = 2811$.

ν_{ar} of additional realizations computed with the PLoM. Figures 9-(a) and -(b) show the graphs of convergence that is obtained (the convergence) for $\nu_{\mathrm{ar}} = 1\,012\,000$. It can be seen that the estimates are converged.

3.4. Convergence of the learning with respect to the size of the initial dataset

The learned dataset, which is constituted of ν_{ar} additional realizations $\{\mathbf{x}^{\ell}_{\mathrm{ar}}, \ell = 1, \ldots, \nu_{\mathrm{ar}}\}$ computed with the PLoM, is performed from the initial dataset $\{\mathbf{x}^{j}_{d}, j = 1, \ldots, N_d\}$ with $N_d \ll \nu_{\mathrm{ar}}$. It is important to control the convergence of the learning, that is to say to control the convergence of the statistical estimates of the observed quantities, $q^{\mathrm{LM}}_{\mathrm{max},1}$ and $q^{\mathrm{LM}}_{\mathrm{max},2}$, with respect to the size, N_d, of the initial dataset. For that, as explained in Section 2.5, we introduce the set $\mathcal{N}_d = \{1000, 1500, 2000, 2100 \text{ to } 2700 \text{ in steps of } 100, 2811\}$ of values of N_d. Figure 10 shows the convergence of the probabilistic learning with respect to the size $N_d \in \mathcal{N}_d$ of the initial dataset. Let $[x_d] \in \mathbb{M}_{10,2811}$ be the initial dataset of maximum size. Let $[x_{d,\mathrm{perm}}] \in \mathbb{M}_{10,2811}$ be a random permutation of the 2 811 columns of $[x_d]$. For each N_d in $\mathcal{N}_d$, the initial dataset $[x_d(N_d)] \in \mathbb{M}_{10,N_d}$, which is used to perform the probabilistic learning with $\nu_{\mathrm{ar}} = 1\,012\,000$ additional realizations, is made up of the first N_d columns of $[x_{d,\mathrm{perm}}]$. In Figure 10-(b), for N_d equal to 1000 and to 2000, the second peak (local maximum) of the



(a) Convergence of $q^{\mathrm{LM}}_{\mathrm{max},1}(N_d)$.

(b) Convergence of $q^{\mathrm{LM}}_{\mathrm{max},2}(N_d)$.

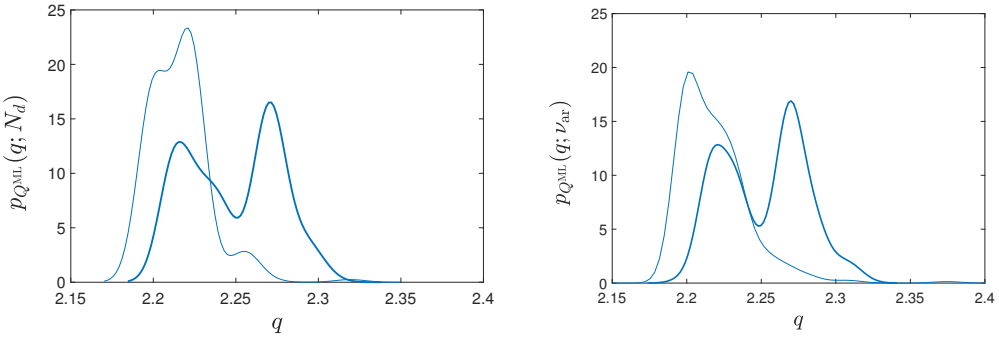
Figure 10. Convergence of $q^{\mathrm{LM}}_{\mathrm{max},1}(N_d)$ and $q^{\mathrm{LM}}_{\mathrm{max},2}(N_d)$ with respect to the size N_d of the initial dataset, for $\nu_{\mathrm{ar}} = 1\,012\,000$ additional realizations of the probabilistic learning.

pdf $q \mapsto p_{Q^{\mathrm{LM}}}(q)$ does not exist (the pdf $p_{Q^{\mathrm{LM}}}$ is unimodal and not bimodal and consequently, the detection of the structural change of the bridge cannot be done). This is the reason why there is no values of $q^{\mathrm{LM}}_{\mathrm{max},2}$ for these two values of N_d. This means that for an initial dataset whose size is less than 2000, the information contained in the database is not sufficiently big for performing the learning. Figures 10-(a) and -(b) show that the convergence of the learning is good. This means that the size $N_d = 2811$ is adapted for performing the probabilistic learning for the two quantities that are observed.

3.5. Validation of the probabilistic criterion coupled with the PLoM

A decision criterion is presented for detecting the occurrence of a structural change of the bridge, based on the analysis of the type of the pdf $q \mapsto p_{Q^{\mathrm{LM}}}(q)$ of the random variable Q^{LM}: if this pdf is bimodal, then there was a structural change. The two local maxima of this bimodal pdf are $q^{\mathrm{LM}}_{\mathrm{max},1}$ and $q^{\mathrm{LM}}_{\mathrm{max},2}$. In addition, in order to obtain a sufficiently good convergence of the estimates of $q^{\mathrm{LM}}_{\mathrm{max},1}$ and $q^{\mathrm{LM}}_{\mathrm{max},2}$, the PLoM algorithm is used in order to enrich the initial information and to increase the number of realizations. In the framework of the use of such a detection algorithm, a small initial database is available (the initial dataset), but the time/moment for which the structural change occurs is unknown and consequently, the analysis cannot be performed using separately the database before the structural modification and using the database after the structural modification. Indeed, it should be noted that if the instant of structural change is known, then the problem of detecting structural change no longer arises. For the application presented, the modification is due to the installation of additional prestressing cables and therefore, this instant of structural change is known. Thanks to this information, we can separate the two data sets, which allows the proposed identification method to be validated (Figure 11). Such a separation cannot be made in the general case, which interests us, for which the identification method is developed. In order to avoid any ambiguity on the role played by Figure 11, we wish to emphasize that this figure is only used to validate the method. It should be noted, as explained in Section 1, that this application was chosen because there was an experimental database, which made it possible to validate the detection methodology, which allows for detecting, from a small database, a possible structural change in a civil engineering structure without knowing the existence or not of a structural modification.

The size N_d of the initial dataset corresponding to the database before structural modifica-



(a) pdf $q \mapsto p_{Q^{\mathrm{LM}}}(q;N_d)$ estimated without using the probabilistic learning.

(b) pdf $q \mapsto p_{Q^{\mathrm{LM}}}(q;\nu_{\mathrm{ar}})$ estimated using the probabilistic learning with $\nu_{\mathrm{ar}} = 1\,012\,000$ additional realizations.

Figure 11. pdf before structural modification with $N_d = 744$ (thin line) and after modification with $N_d = 1887$ (thick line).

tion is 744 while it is 1887 after the structural modification. During all the time considered for the measurements, the interval of the 2811 values of the temperature measured by sensor number 6, located inside the box girder is $[-7.7\,,31.4]\,^oC$ for which the statistical mean value is $11.4\,^oC$. The value of the first eigenfrequency corresponding to $-7.7\,^oC$ is $2.39\,Hz$ while it is $2.20\,Hz$ for $31.4\,^oC$. The measurements performed before the structural modification yields $[-7.7\,,12.7]\,^oC$ with a statistical mean value $4.3\,^oC$ yielding for the corresponding eigenfrequencies $2.39\,Hz$ and $2.20\,Hz$, respectively. Concerning the measurements performed after the structural modification, the interval of temperatures is $[-5.0\,,31.4]\,^oC$ with a statistical mean value $14.4\,^oC$ yielding for the corresponding eigenfrequencies $2.32\,Hz$ and $2.20\,Hz$, respectively. The database shows that there is a large number of temperature in the interval $[-1.0\,,10.0]\,^oC \subset [-5.0\,,12.7]\,^oC$, measured before and after the structural modification.
As the first frequency decreases when the temperature increases (and therefore increases when the temperature decreases), the database shows that there is effectively an overlap of the values of the first eigenfrequency before and after the structural modification. This implies that, the pdf $q \mapsto p_{Q^{\mathrm{LM}}}(q)$ resulting from the analysis of the data after the structural modification must be bimodal with a first peak occurring for a first eigenfrequency close to the peak of the pdf before the structural modification and a second peak for a higher value of the first eigenfrequency, linked to the structural modification and which must not appear in the pdf before the structural modification. Figure 11 shows the pdf $q \mapsto p_{Q^{\mathrm{LM}}}(q;N_d)$ estimated without learning and $q \mapsto p_{Q^{\mathrm{LM}}}(q;\nu_{\mathrm{ar}})$ estimated with the learning, using only the initial dataset before structural modification (Figure 11-(a), $N_d = 744$) and after structural modification (Figure 11-(b), $N_d = 1887$). Figure 11-(a) (without probabilistic learning) confirms the analysis proposed. Figure 11-(b), obtained with the probabilistic learning, gives a result similar to Figure 11-(a) and therefore, confirms the proposed analysis. This set of results validates the predictions presented in Section 3.2 using the global database and the PLoM (in particular, in Figure 7, where it was shown that the detection cannot be made without learning).

4. Conclusions

In this paper, the probabilistic learning on manifolds recently developed for the case of small data has been applied to detect structural changes in civil engineering structures. The available database consists of a small number of experimental records as opposed to the case of big data. Therefore, the very effective methods of machine learning based on the use of

neural networks do not apply. This small database comes from the dynamic monitoring of a box-girder bridge subjected to the dynamic effects of road traffic and to temperature variations. The novelty of this work does not lie in the very recent statistical method, which is used for this application, since it has already been validated for many application fields. The novelty lies in the possibility of using it and applying it to a difficult case of detection in the field of civil engineering structures as we have demonstrated in this paper. In addition, the implementation of this probabilistic learning tool has required the development of a probabilistic detection criterion based on the occurrence of a bimodal character of the probability distribution of the quantity of interest, which is novel.

The probabilistic learning method used has no intrinsic limitation. It allows probabilistic learning from small databases. On the other hand, it can be limited in its learning capacity depending on the applications. If the small database that constitutes the initial set does not contain the information required to discover the probabilistic structure of the data, then the method will not succeed in making the detection. However, the method allows to diagnose whether, for a given application, probabilistic learning has been done correctly or not. This diagnosis consists to do the convergence analysis of learning with respect to the dimension of the initial dataset, as we have done for this application and we have shown that this convergence of the learning has been obtained.

The results obtained show that the initial database does not make it possible to detect the changes in structural stiffness without learning while it is detected with the probabilistic learning on manifolds. The proposed method is applicable to similar structural health monitoring problems.

Acknowledgement(s)

The authors thank SANEF (Société des Autoroutes du Nord et de l'Est de la France), which authorized the diffusion of the experimental data related to the dynamic health monitoring of the Oise bridge, as part of the French national project S3 (Structural Health Monitoring of structures).

Disclosure statement

No potential conflict of interest was reported by the authors.

Nomenclature/Notation

Lower-case letters such as q or η are deterministic real variables.
Boldface lower-case letters such as $\mathbf{q}$ or $\boldsymbol{\eta}$ are deterministic vectors.
Lower-case letters $\mathfrak{q}$, $\mathfrak{w}$, and $\mathfrak{x}$ are deterministic vectors.
Upper-case letters such as X or H are real-valued random variables.
Boldface upper-case letters such as $\mathbf{X}$ or $\mathbf{H}$ are vector-valued random variables.
Upper-case letters $\mathbb{Q}$, $\mathbb{U}$, $\mathbb{W}$, and $\mathbb{X}$ are vector-valued random variables.
Lower-case letters between brackets such as $[x]$ or $[\eta]$ are deterministic matrices.
Boldface upper-case letters between brackets such as $[\mathbf{X}]$ or $[\mathbf{H}]$ are matrix-valued random variables.

n: dimension ($n = n_q + n_w$) of vector $\mathbf{x}$ and $\mathbf{X}$.

n_q : dimension of vectors **q** and **Q**.
n_w : dimension of vectors **w** and **W**.
ν : dimension of **H**.
$\mathbf{q}^j$ or $\mathbf{q}^\ell$: realization of **Q**.
$\mathbf{w}^j$ or $\mathbf{w}^\ell$: realization of **W**.
$\mathbf{x}^j$ or $\mathbf{x}^\ell$: realization of **X**.
ν_{ar} : number of additional realizations generated with the PLoM.
N_d : number of realizations in initial dataset.
Q : random QoI (output).
W : random control parameter (input).
X : random vector $(\mathbf{Q}, \mathbf{W})$.

$[I_n]$: identity matrix in $\mathbb{M}_n$.
$\mathbb{M}_{n,N}$: set of all the $(n \times N)$ real matrices.
$\mathbb{M}_n$: set of all the square $(n \times n)$ real matrices.
$\mathbb{R}$: set of all the real numbers.
$\mathbb{R}^n$: Euclidean vector space on $\mathbb{R}$ of dimension n.
$[x]_{kj}$: entry of matrix $[x]$.
$[x]^T$: transpose of matrix $[x]$.
$\delta_{kk'}$: Kronecker's symbol such that $\delta_{kk'} = 0$ if $k \neq k'$ and $= 1$ if $k = k'$.
$\|\mathbf{x}\|$: usual Euclidean norm in $\mathbb{R}^n$.
$<\mathbf{x}, \mathbf{y}>$: usual Euclidean inner product in $\mathbb{R}^n$.
$\delta_{kk'}$: Kronecker's symbol.
E: mathematical expectation.

pdf: probability density function.
MCMC: Markov Chain Monte Carlo.

5. References

References

Adeli, H. (2001). Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering*, *16*(2), 126–142.

Aggarwal, C. C., & Zhai, C. (2012). *Mining text data*. Springer Science & Business Media.

Alves, V., Cury, A., Roitman, N., Magluta, C., & Cremona, C. (2015). Structural modification assessment using supervised learning methods applied to vibration data. *Engineering Structures*, *99*, 439–448.

Amezquita-Sanchez, J., & Adeli, H. (2015). Feature extraction and classification techniques for health monitoring of structures. *Scientia Iranica. Transaction A, Civil Engineering*, *22*(6), 1931.

Arangio, S., & Bontempi, F. (2015). Structural health monitoring of a cable-stayed bridge with bayesian neural networks. *Structure and Infrastructure Engineering*, *11*(4), 575–587.

Balcan, M.-F. F., & Feldman, V. (2013). Statistical active learning algorithms. *Proceedings of Advances in neural information processing systems*, 1295–1303.

Bowman, A., & Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with s-plus illustrations* (Vol. 18). Oxford: Clarendon Press, New

York: Oxford University Press.

Brownjohn, J. M. (2007). Structural health monitoring of civil infrastructure. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *365*(1851), 589–622.

Brownjohn, J. M., De Stefano, A., Xu, Y.-L., Wenzel, H., & Aktan, A. E. (2011). Vibration-based monitoring of civil infrastructure: challenges and successes. *Journal of Civil Structural Health Monitoring*, *1*(3-4), 79–95.

Burrage, K., Lenane, I., & Lythe, G. (2007). Numerical methods for second-order stochastic differential equations. *SIAM Journal on Scientific Computing*, *29*(1), 245–264.

Byrd, R., Chin, G., Neveitt, W., & Nocedal, J. (2011). On the use of stochastic Hessian information in optimization methods for machine learning. *SIAM Journal of Optimization*, *21*(3), 977–995.

Cha, Y.-J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, *32*(5), 361–378.

Coifman, R., Lafon, S., Lee, A., Maggioni, M., Nadler, B., Warner, F., & Zucker, S. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *PNAS*, *102*(21), 7426–7431.

Cury, A., Cremona, C., & Dumoulin, J. (2012). Long-term monitoring of a psc box girder bridge: Operational modal analysis, data normalization and structural modification assessment. *Mechanical Systems and Signal Processing*, *33*, 13–37.

Dalalyan, A. S., & Tsybakov, A. B. (2012). Sparse regression learning by aggregation and langevin monte-carlo. *Journal of Computer and System Sciences*, *78*(5), 1423–1443.

Dong, C.-Z., Celik, O., Catbas, F. N., O Brien, E. J., & Taylor, S. (2020). Structural displacement monitoring using deep learning-based full field optical flow methods. *Structure and Infrastructure Engineering*, *16*(1), 51–71.

Du, X., & Chen, W. (2004). Sequential optimization and reliability assessment method for efficient probabilistic design. *ASME Journal of Mechanical Design*, *126*(2), 225–233.

Eldred, M. (2011). Design under uncertainty employing stochastic expansion methods. *International Journal for Uncertainty Quantification*, *1*(2), 119–146.

Erfani, S. M., Rajasegarar, S., Karunasekera, S., & Leckie, C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. *Pattern Recognition*, *58*, 121–134.

Farhat, C., Tezaur, R., Chapman, T., Avery, P., & Soize, C. (2019). Feasible probabilistic learning method for model-form uncertainty quantification in vibration analysis. *AIAA Journal*, *57*(11), 4978-4991.

Farrar, C. R., & Worden, K. (2007). An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *365*(1851), 303–315.

Farrar, C. R., & Worden, K. (2012). *Structural health monitoring: a machine learning perspective*. John Wiley & Sons.

Feng, C., Liu, M.-Y., Kao, C.-C., & Lee, T.-Y. (2017). Deep active learning for civil infrastructure defect detection and classification. In *Computing in civil engineering 2017* (pp. 298–306).

Frangopol, D. M. (2011). Life-cycle performance, management, and optimisation of structural systems under uncertainty: accomplishments and challenges. *Structure and infrastructure Engineering*, *7*(6), 389–413.

Frangopol, D. M., & Liu, M. (2007). Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost. *Structure and infrastructure engineering*, *3*(1), 29–41.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452.

Ghanem, R., Higdon, D., & Owhadi, H. (2017). *Handbook of uncertainty quantification* (Vol. 1 to 3). Cham, Switzerland: Springer.

Ghanem, R., & Soize, C. (2018). Probabilistic nonconvex constrained optimization with fixed number of function evaluations. *International Journal for Numerical Methods in Engineering*, *113*(4), 719–741.

Ghanem, R., Soize, C., Safta, C., Huan, X., Lacaze, G., Oefelein, J. C., & Najm, H. N. (2019). Design optimization of a scramjet under uncertainty using probabilistic learning on manifolds. *Journal of Computational Physics*, *399*, 108930.

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

Gui, G., Pan, H., Lin, Z., Li, Y., & Yuan, Z. (2017). Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection. *KSCE Journal of Civil Engineering*, *21*(2), 523–534.

Guilleminot, J., & Dolbow, J. E. (2020). Data-driven enhancement of fracture paths in random composites. *Mechanics Research Communications*, *103*, 103443.

Hairer, E., Lubich, C., & Wanner, G. (2006). *Geometric numerical integration. structure-preserving algorithms for ordinary differential equations* (Vol. 31). Springer Science & Business Media.

Hewayde, E., Nehdi, M., Allouche, E., & Nakhla, G. (2007). Neural network prediction of concrete degradation by sulphuric acid attack. *Structure and Infrastructure Engineering*, *3*(1), 17–27.

Homem-de Mello, T., & Bayraksan, G. (2014). Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Researh and Management Science*, *19*(1), 56–85.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jones, D., Schonlau, M., & Welch, W. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, *13*(4), 455–492.

Kim, S., & Frangopol, D. M. (2011). Cost-based optimum scheduling of inspection and monitoring for fatigue-sensitive structures under uncertainty. *Journal of Structural Engineering*, *137*(11), 1319–1331.

Kleijnen, J., van Beers, W., & van Nieuwenhuyse, I. (2010). Constrained optimization in expensive simulation: Novel approach. *European Journal of Operational Research*, *202*(1), 164–174.

Ko, J., & Ni, Y. Q. (2005). Technology developments in structural health monitoring of large-scale bridges. *Engineering structures*, *27*(12), 1715–1725.

Lin, Y.-z., Nie, Z.-h., & Ma, H.-w. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering*, *32*(12), 1025–1046.

Liu, H., & Zhang, Y. (2020). Bridge condition rating data modeling using deep learning algorithm. *Structure and Infrastructure Engineering*, 1–14.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT press.

Okasha, N. M., Frangopol, D. M., & Orcesi, A. D. (2012). Automated finite element updating using strain data for the lifetime reliability assessment of bridges. *Reliability Engineering & System Safety*, *99*, 139–150.

Orcesi, A. D., & Frangopol, D. M. (2010). Inclusion of crawl tests and long-term health monitoring in bridge serviceability analysis. *Journal of Bridge Engineering*, *15*(3),

312–326.

Orcesi, A. D., Frangopol, D. M., & Kim, S. (2010). Optimization of bridge maintenance strategies based on multiple limit states and monitoring. *Engineering Structures*, *32*(3), 627–640.

Pandey, M., Yuan, X.-X., & Van Noortwijk, J. (2009). The influence of temporal uncertainty of deterioration on life-cycle management of structures. *Structure and Infrastructure Engineering*, *5*(2), 145–156.

Patcha, A., & Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer networks*, *51*(12), 3448–3470.

Perrin, G., Soize, C., Marque-Pucheu, S., & Garnier, J. (2017). Nested polynomial trends for the improvement of gaussian process-based predictors. *Journal of Computational Physics*, *346*, 389–402.

Perrin, G., Soize, C., & Ouhbi, N. (2018). Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints. *Computational Statistics & Data Analysis*, *119*, 139–154.

Queipo, N., Haftka, R., Shyy, W., Goel, T., Vaidyanathan, R., & Tucker, K. (2005). Surrogate-based analysis and optimization. *Progress in Aerospace Science*, *41*(1), 1–28.

Rafiei, M. H., & Adeli, H. (2017). A novel machine learning-based algorithm to detect damage in high-rise building structures. *The Structural Design of Tall and Special Buildings*, *26*(18), e1400.

Salehi, H., & Burgueno, R. (2018). Emerging artificial intelligence methods in structural engineering. *Engineering structures*, *171*, 170–189.

Santos, A., Figueiredo, E., Silva, M., Sales, C., & Costa, J. (2016). Machine learning algorithms for damage detection: Kernel-based approaches. *Journal of Sound and Vibration*, *363*, 584–599.

Santos, J. P., Crémona, C., Calado, L., Silveira, P., & Orcesi, A. D. (2016). On-line unsupervised detection of early damage. *Structural Control and Health Monitoring*, *23*(7), 1047–1069.

Santos, J. P., Crémona, C., Orcesi, A. D., & Silveira, P. (2013). Multivariate statistical analysis for early damage detection. *Engineering Structures*, *56*, 273–285.

Santos, J. P., Orcesi, A. D., Crémona, C., & Silveira, P. (2015). Baseline-free real-time assessment of structural changes. *Structure and Infrastructure Engineering*, *11*(2), 145–161.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85–117.

Schoefs, F., Yáñez-Godoy, H., & Lanata, F. (2011). Polynomial chaos representation for identification of mechanical characteristics of instrumented structures. *Computer-Aided Civil and Infrastructure Engineering*, *26*(3), 173–189.

Schölkopf, B., Smola, A., & Müller, K. (1997). Kernel principal component analysis. In W. Gerstner, A. Germond, M. Hasler, & J. Nicoud (Eds.), *Artificial neural networks ICANN'97* (Vol. 1327, p. 583-588). Berlin, Heidelberg: Springer.

Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., ... Adams, R. (2015). Scalable bayesian optimization using deep neural networks. In *International conference on machine learning* (pp. 2171–2180).

Soize, C. (2008). Construction of probability distributions in high dimension using the maximum entropy principle. applications to stochastic processes, random fields and random matrices. *International Journal for Numerical Methods in Engineering*, *76*(10), 1583-1611.

Soize, C., & Ghanem, R. (2016). Data-driven probability concentration and sampling on manifold. *Journal of Computational Physics*, *321*, 242-258.

Soize, C., & Ghanem, R. (2020). Probabilistic learning on manifolds. *submitted, also in arXiv preprint arXiv:2002.12653, 2020, math.ST, 28 February 2020*.

Soize, C., & Ghanem, R. (2020a). Physics-constrained non-gaussian probabilistic learning on manifolds. *International Journal for Numerical Methods in Engineering*, *121*(1), 110-145.

Soize, C., Ghanem, R., Safta, C., Huan, X., Vane, Z. P., Oefelein, J. C., ... Chen, X. (2019). Entropy-based closure for probabilistic learning on manifolds. *Journal of Computational Physics*, *388*, 528-533.

Strauss, A. (2016). Numerical and monitoring based markov chain approaches for the fatigue life prediction of concrete structures. *Engineering Structures*, *112*, 265–273.

Tan, Z. X., Thambiratnam, D. P., Chan, T. H., Gordan, M., & Abdul Razak, H. (2019). Damage detection in steel-concrete composite bridge using vibration characteristics and artificial neural network. *Structure and Infrastructure Engineering*, 1–15.

Taylor, J., & Tibshirani, R. J. (2015). Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, *112*(25), 7629–7634.

Vapnik, V. (2000). *The nature of statistical learning theory*. New York: Springer.

Wang, Z., Zoghi, M., Hutter, F., Matheson, D., & de Freitas, N. (2016). Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, *55*, 361–387.

Wong, K.-Y. (2007). Design of a structural health monitoring system for long-span bridges. *Structure and Infrastructure Engineering*, *3*(2), 169–185.

Worden, K., Farrar, C. R., Manson, G., & Park, G. (2007). The fundamental axioms of structural health monitoring. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *463*(2082), 1639–1664.

Xie, J., Frazier, P., & Chick, S. (2016). Bayesian optimization via simulation with pairwise sampling and correlated pair beliefs. *Operations Research*, *64*(2), 542–559.

Xu, B., Wu, Z., Chen, G., & Yokoyama, K. (2004). Direct identification of structural parameters from dynamic responses with neural networks. *Engineering Applications of Artificial Intelligence*, *17*(8), 931–943.

Yao, W., Chen, X., Luo, W., vanTooren, M., & Guo, J. (2011). Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. *Progress in Aerospace Sciences*, *47*, 450–479.

Appendix A. Summary of the probabilistic learning on manifolds (PLoM)

In this Appendix, we summarize the probabilistic learning algorithm introduced in Soize and Ghanem (2016, 2020); Soize et al. (2019) that we use, for which extensions can be found in Soize and Ghanem (2020a), and for which many developments and applications can be found in Farhat et al. (2019); Ghanem and Soize (2018); Ghanem et al. (2019); Guilleminot and Dolbow (2020). This appendix gives the objective and the framework PLoM, and then the algorithm. Nevertheless, we will begin it giving a very short literature review on this subject.

Machine learning is revolved around empirical models such as kernels or Neural Networks (NN) that require big data and efficient algorithms for their identification and training. If the volume of data is not sufficiently large, it may not generally be possible to train the NN to the desired behavior. In the framework of computational science and engineering, while computationally taxing simulations are typically used to generate big data, the quantities of interest (QoI) from each such simulations are typically much smaller '"small data"). In this

context, probabilistic learning is a way for improving the knowledge that one has from only a small number of expensive evaluations of a computational model in order to be able to solve a problem or from a limited and partial experimental measurements. This is one reason for which statistical and probabilistic learning methods have been extensively developed (see for instance, Aggarwal and Zhai (2012); Balcan and Feldman (2013); Dalalyan and Tsybakov (2012); Ghahramani (2015); James, Witten, Hastie, and Tibshirani (2013); Murphy (2012); Schmidhuber (2015); Schölkopf, Smola, and Müller (1997); Taylor and Tibshirani (2015); Vapnik (2000)) and play an increasingly important role in computational physics and engineering science Ghanem, Higdon, and Owhadi (2017). In large scale model-driven design optimization under uncertainty, and more generally, in artificial intelligence for extracting information from big data, statistical learning methods have been developed in the form of surrogate models that can easily be evaluated Homem-de Mello and Bayraksan (2014); Queipo et al. (2005); Snoek et al. (2015) such as, Gaussian process surrogate models Kleijnen, van Beers, and van Nieuwenhuyse (2010); Perrin, Soize, Marque-Pucheu, and Garnier (2017), Bayesian calibration methods Jones, Schonlau, and Welch (1998); Wang, Zoghi, Hutter, Matheson, and de Freitas (2016); Xie, Frazier, and Chick (2016), active learning Gorissen, Couckuyt, Demeester, Dhaene, and Crombecq (2010); Perrin, Soize, and Ouhbi (2018), which allow for decreasing the numerical cost of the evaluations of expensive functions Byrd, Chin, Neveitt, and Nocedal (2011); Du and Chen (2004); Eldred (2011); Yao, Chen, Luo, vanTooren, and Guo (2011).

A.1. Objective and framework of the PLoM

A typical problem for the use of the PLoM is the following. Let $(\mathbf{w}, \mathbf{u}) \mapsto \mathbf{f}(\mathbf{w}, \mathbf{u})$ be any measurable mapping on $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$ with values in $\mathbb{R}^{n_q}$ representing a system for which a mathematical/computational model is developed or a system for which experimental measurements are done. Let $\mathbf{W}$ and $\mathbf{U}$ be two independent (non-Gaussian) random variables defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^{n_w}$ and $\mathbb{R}^{n_u}$, for which the probability distributions $P_{\mathbf{W}}(d\mathbf{w}) = p_{\mathbf{W}}(\mathbf{w})\, d\mathbf{w}$ and $P_{\mathbf{U}}(d\mathbf{u}) = p_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u}$ are defined by the probability density functions $p_{\mathbf{W}}$ and $p_{\mathbf{U}}$ with respect to the Lebesgue measures $d\mathbf{w}$ and $d\mathbf{u}$ on $\mathbb{R}^{n_w}$ and $\mathbb{R}^{n_u}$. Random vector $\mathbf{W}$ is made up of the part of the random parameters that control the system, while random vector $\mathbf{U}$ is made up of the other part of the random parameters, which are not used for controlling the system. Let $\mathbf{Q}$ be the vector of the quantities of interest (QoI) that is a random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^{n_q}$ such that

$$\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbf{U})\,. \tag{A1}$$

Random QoI, $\mathbf{Q}$, represents the vector of all the observations performed in the system, which is either a function of the solution of the computational model or for which measurements are performed. Let us assume that N_d calculations have been performed with the mathematical/computational model whose solution is represented by Equation (A1) or that N_d experimental measurements have been processed, allowing N_d independent realizations $\{\mathbf{q}^j, j = 1, \ldots, N_d\}$ of $\mathbf{Q}$ to be known such that

$$\mathbf{q}^j = \mathbf{f}(\mathbf{w}^j, \mathbf{u}^j)\,, \tag{A2}$$

in which $\{\mathbf{w}^j, j = 1, \ldots, N_d\}$ and $\{\mathbf{u}^j, j = 1, \ldots, N_d\}$ are N_d independent realizations of $(\mathbf{W}, \mathbf{U})$. We then consider the random variable $\mathbf{X}$ with values in $\mathbb{R}^n$, such that

$$\mathbf{X} = (\mathbf{Q}, \mathbf{W}) \quad , \quad n = n_q + n_w\,. \tag{A3}$$

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

$$\{\mathbf{x}^j, j = 1, \ldots, N_d\} \quad , \quad \mathbf{x}^j = (\mathbf{q}^j, \mathbf{w}^j) \in \mathbb{R}^n. \tag{A4}$$

The PLoM allows for generating additional realizations $\{(\mathbf{q}^\ell_{\text{ar}}, \mathbf{w}^\ell_{\text{ar}}), \ell = 1, \ldots, \nu_{\text{ar}}\}$ for $\nu_{\text{ar}} \gg N_d$ (without using the computational model or without using other experimental measurements), but using only the initial dataset. These additional realizations allow for calculating converged estimates for any statistical quantities related to $\mathbf{Q}$ and $\mathbf{W}$.

A.2. Algorithm of the PLoM

1. Scaling of the initial dataset. In practice, the initial dataset can be made up of heterogeneous numerical values and must be scaled for performing computational statistics. Consequently, quantities $\mathbf{q}^j$, $\mathbf{x}^j$, and $\mathbf{x}^j$ are assumed to be scaled quantities (see Soize and Ghanem (2016) for the scaling).

2. Data normalization. Let $\mathbf{X}$ be the $\mathbb{R}^n$-valued second-order random vector defined by Equation (A3) for which the N_d independent realizations are N_d data points in $\mathbb{R}^n$, represented by the matrix $[x_d] = [\mathbf{x}^1 \ldots \mathbf{x}^{N_d}]$ in $\mathbb{M}_{n,N_d}$. Let $[\mathbf{X}] = [\mathbf{X}^1, \ldots, \mathbf{X}^{N_d}]$ be the random matrix with values in $\mathbb{M}_{n,N_d}$, whose columns are N_d independent copies of random vector $\mathbf{X}$. The normalization of random matrix $[\mathbf{X}]$ is attained with random matrix $[\mathbf{H}] = [\mathbf{H}^1, \ldots, \mathbf{H}^{N_d}]$ with values in $\mathbb{M}_{\nu,N_d}$, whose columns are N_d independent copies of a random vector $\mathbf{H}$, with $\nu \leq n$, obtained by using the principal component analysis that allows us to write $[\mathbf{X}]$ as

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

in which $[\lambda]$ is the $(\nu \times \nu)$ diagonal matrix of the ν positive eigenvalues of the empirical estimate of the covariance matrix of $\mathbf{X}$ (computed using $\mathbf{x}^1, \ldots, \mathbf{x}^{N_d}$), where $[\varphi]$ is the $(n \times \nu)$ matrix of the associated eigenvectors such $[\varphi]^T\,[\varphi] = [I_\nu]$, and where $[\underline{x}]$ is the matrix in $\mathbb{M}_{n,N_d}$ with identical columns, each one being equal to the empirical estimate $\underline{\mathbf{x}} \in \mathbb{R}^n$ of the mean value of random vector $\mathbf{x}$ (computed using $\mathbf{x}^1, \ldots, \mathbf{x}^{N_d}$). The sample $[\eta_d] = [\boldsymbol{\eta}^1 \ldots \boldsymbol{\eta}^{N_d}] \in \mathbb{M}_{\nu,N_d}$ of $[\mathbf{H}]$ (associated with the sample $[x_d]$ of $[\mathbf{X}]$) is computed by

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

When n is small, ν can be chosen as n. If some eigenvalues are zero, they must be eliminated and then $\nu < n$. When n is high, a statistical reduction can be done as usual and therefore $\nu < n$ in such a case.

3. Diffusion-maps basis. For $1 < m \leq N_d$, let $[g] = [\mathbf{g}^1 \ldots \mathbf{g}^m] \in \mathbb{M}_{N_d,m}$ be the "diffusion-maps basis" that is constructed by using the diffusion maps proposed by Coifman et al. (2005). Let $[\mathbb{b}]$ be the positive-definite diagonal real matrix in $\mathbb{M}_{N_d}$ such that $[\mathbb{b}]_{ij} = \delta_{ij}\ \sum_{j'=1}^{N_d}[K]_{ij'}$ in which $[K]_{ij'} = \exp(-\frac{1}{4\varepsilon_{\text{diff}}}\|\boldsymbol{\eta}^i - \boldsymbol{\eta}^{j'}\|^2)$, depending on a real smoothing parameter $\varepsilon_{\text{diff}} > 0$. Let $[\mathbb{P}]$ be the transition matrix in $\mathbb{M}_{N_d}$ such that $[\mathbb{P}] = [\mathbb{b}]^{-1}\,[K]$. Let $\boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^m$ be the right

eigenvectors of $[\mathbb{P}]$, associated with the eigenvalues $1 = \Lambda_1 > \Lambda_2 > \ldots > \Lambda_{m+1}$, such that

$$[\mathbb{P}]\, \boldsymbol{\psi}^\alpha = \Lambda_\alpha\, \boldsymbol{\psi}^\alpha \,.$$

The normalization of the right eigenvectors of $[\mathbb{P}]$ is such that the matrix $[\psi] = [\boldsymbol{\psi}^1 \ldots \boldsymbol{\psi}^m]$ is chosen for that $[\psi]^T\, [\mathbb{b}]\, [\psi] = [I_m]$. The eigenvector $\boldsymbol{\psi}^1$ associated with the largest eigenvalue $\Lambda_1 = 1$ is a constant vector (all its components are equal). We will defined the "diffusion-maps basis" by $[g] = [\mathbf{g}^1 \ldots \mathbf{g}^m] \in \mathbb{M}_{N_d,m}$ such that

$$\mathbf{g}^\alpha = \Lambda_\alpha^\kappa\, \boldsymbol{\psi}^\alpha \in \mathbb{R}^{N_d} \,,$$

in which $\kappa \geq 0$. It is proven in Soize and Ghanem (2020) that, for the PLoM method, $\kappa = 0$ can be used.

4. Reduced-order representation of random matrices $[\boldsymbol{H}]$ *and* $[\boldsymbol{X}]$. The diffusion-maps vectors $\mathbf{g}^1, \ldots, \mathbf{g}^m \in \mathbb{R}^{N_d}$ span a subspace of $\mathbb{R}^{N_d}$ that characterizes the local geometry structure of the dataset concentrated in the neighborhood of a subset of $\mathbb{R}^{N_d}$. The reduced-order representation is obtained in projecting each column of the $\mathbb{M}_{N_d,\nu}$-valued random matrix $[\mathbf{H}]^T$ on the subspace of $\mathbb{R}^{N_d}$, spanned by $\{\mathbf{g}^1 \ldots \mathbf{g}^m\}$. Let $[\mathbf{Z}]$ be the random matrix with values in $\mathbb{M}_{\nu,m}$ such that $[\mathbf{H}] = [\mathbf{Z}]\, [g]^T$. As the matrix $[g]^T\, [g] \in \mathbb{M}_m$ is invertible, the least squares approximation of $\mathbf{Z}$ is written as $[\mathbf{Z}] = [\mathbf{H}]\, [a]$ in which

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

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

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

The representation of random matrix $[\mathbf{X}]$ as function of random matrix $[\mathbf{Z}]$ is then given by

$$[\mathbf{X}] = [\underline{x}] + [\varphi]\, [\lambda]^{1/2}\, [\mathbf{Z}]\, [g]^T \,. \tag{A5}$$

The construction introduces two hyperparameters: the dimension $m \leq N_d$ and the smoothing parameter $\varepsilon_{\text{diff}} > 0$. An algorithm is proposed in Soize et al. (2019) for estimating their values. Most of the time, m and $\varepsilon_{\text{diff}}$ can be chosen as follows. Let $\varepsilon_{\text{diff}} \mapsto \widehat{m}(\varepsilon_{\text{diff}})$ be the function from $]0\,, +\infty[$ into the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ of all the integers such that

$$\widehat{m}(\varepsilon_{\text{diff}}) = \arg \min_{\alpha \,|\, \alpha \geq 3} \left\{ \frac{\Lambda_\alpha(\varepsilon_{\text{diff}})}{\Lambda_2(\varepsilon_{\text{diff}})} < 0.1 \right\} \,. \tag{A6}$$

If function $\widehat{m}$ is a decreasing function of $\varepsilon_{\text{diff}}$ in the broad sense (if not, see Soize et al. (2019)), then the optimal value $\varepsilon_{\text{diff}}^{\text{opt}}$ of $\varepsilon_{\text{diff}}$ can be chosen as the smallest value of the integer $\widehat{m}(\varepsilon_{\text{diff}}^{\text{opt}})$ such that

$$\{\widehat{m}(\varepsilon_{\text{diff}}^{\text{opt}}) < \widehat{m}(\varepsilon_{\text{diff}})\,, \forall \varepsilon_{\text{diff}} \in]0, \varepsilon_{\text{diff}}^{\text{opt}}[\,\} \cap \{\widehat{m}(\varepsilon_{\text{diff}}^{\text{opt}}) = \widehat{m}(\varepsilon_{\text{diff}})\,, \forall \varepsilon_{\text{diff}} \in]\varepsilon_{\text{diff}}^{\text{opt}}, 1.5\, \varepsilon_{\text{diff}}^{\text{opt}}[\,\} \,. \tag{A7}$$

The corresponding optimal value m^{opt} of m is then given by $m^{\text{opt}} = \widehat{m}(\varepsilon_{\text{diff}}^{\text{opt}}) - 1$.

5. Generation of additional realizations $\{\boldsymbol{x}_{\text{ar}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$ *of random vector* $\boldsymbol{X}$. The generation of additional realizations $[z_{\text{ar}}^1], \ldots, [z_{\text{ar}}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{Z}]$ is carried out by using an

unusual MCMC method based on a reduced-order Itô stochastic differential equation (ISDE) that is constructed as the projection on the diffusion-maps basis of the ISDE related to a dissipative Hamiltonian dynamical system Soize (2008); Soize and Ghanem (2016) for which the invariant measure is the pdf of random matrix $[\mathbf{H}]$ constructed with the Gaussian kernel-density estimation method and $[\eta_d]$. This method preserves the concentration of the probability measure and avoids the scatter phenomenon. The constructed reduced-order ISDE is then used for generating n_{MC} additional realizations, $[z_{\text{ar}}^1], \ldots, [z_{\text{ar}}^{n_{\text{MC}}}]$ in $\mathbb{M}_{\nu,m}$, of random matrix $[\mathbf{Z}]$, and therefore, for deducing the n_{MC} additional realizations, $[\eta_{\text{ar}}^1], \ldots, [\eta_{\text{ar}}^{n_{\text{MC}}}]$ in $\mathbb{M}_{\nu,N_d}$ of random matrix $[\mathbf{H}]$, such that $[\eta_{\text{ar}}^\ell] = [z_{\text{ar}}^\ell]\,[g]^T$ for $\ell = 1, \ldots, n_{\text{MC}}$. Let $\{([\boldsymbol{\mathcal{Z}}(r)], [\boldsymbol{\mathcal{Y}}(r)]), r \in \mathbb{R}^+\}$ be the unique asymptotic (for $r \to +\infty$) stationary and ergodic diffusion stochastic process with values in $\mathbb{M}_{\nu,m} \times \mathbb{M}_{\nu,m}$, of the following reduced-order ISDE (stochastic nonlinear second-order dissipative Hamiltonian dynamical system), for $r > 0$,

$$
\begin{aligned}
d[\boldsymbol{\mathcal{Z}}(r)] &= [\boldsymbol{\mathcal{Y}}(r)]\, dr\,, \\
d[\boldsymbol{\mathcal{Y}}(r)] &= [\mathcal{L}([\boldsymbol{\mathcal{Z}}(r)])]\, dr - \frac{1}{2} f_0\, [\boldsymbol{\mathcal{Y}}(r)]\, dr + \sqrt{f_0}\, [d\boldsymbol{\mathcal{W}}(r)]\,,
\end{aligned}
$$

with the initial condition

$$
[\boldsymbol{\mathcal{Z}}(0)] = [z_d] \quad , \quad [\boldsymbol{\mathcal{Y}}(0)] = [\boldsymbol{\mathcal{N}}]\,[a] \quad a.s\,.
$$

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

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

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

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

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

(ii) $[d\,\boldsymbol{\mathcal{W}}(r)] = [d\,\mathbb{W}(r)]\,[a]$ where $[d\,\mathbb{W}(r)]$ is the $\mathbb{M}_{\nu,N_d}$-valued normalized Wiener stochastic process.
(iii) $[\boldsymbol{\mathcal{N}}]$ is the $\mathbb{M}_{\nu,N_d}$-valued normalized Gaussian random matrix.
(iv) The free parameter $f_0 > 0$ allows the dissipation term of the nonlinear second-order dynamical system (dissipative Hamiltonian system) to be controlled in order to kill the transient part induced by the initial conditions.

(v) We then have $[\mathbf{Z}] = \lim_{r \to +\infty} [\boldsymbol{\mathcal{Z}}(r)]$ in probability distribution, which allows for generating the additional realizations, $[z_{\text{ar}}^1], \ldots, [z_{\text{ar}}^{n_{\text{MC}}}]$, and then, generating the additional realizations $[x_{\text{ar}}^1], \ldots, [x_{\text{ar}}^{n_{\text{MC}}}]$ by using Equation (A5), which are reshaped in order to deduce the $\nu_{\text{ar}} = n_{\text{MC}} \times N_d$ additional realizations, $\{\mathbf{x}_{\text{ar}}^\ell, \ell = 1, \ldots, \nu_{\text{ar}}\}$.

6. Algorithm for solving the reduced-order ISDE. The algorithm for solving the reduced-order ISDE is detailed in Soize and Ghanem (2016) and is summarized hereinafter. The Störmer-Verlet scheme Burrage, Lenane, and Lythe (2007); Hairer, Lubich, and Wanner (2006), is used. Let $M = n_{\text{MC}} \times M_0$ be the positive integer in which n_{MC} and M_0 are integers. The reduced-order ISDE is solved on the finite interval $\mathcal{R} = [0\,, M\,\Delta r]$, in which Δr is the sampling step of the continuous index parameter r. The integration scheme is based on the use of the $M+1$ sampling points $r_{\ell'}$ such that $r_{\ell'} = \ell'\,\Delta r$ for $\ell' = 0, \ldots, M$. The following notations are introduced: $[\boldsymbol{\mathcal{Z}}_{\ell'}] = [\boldsymbol{\mathcal{Z}}(r_{\ell'})]$, $[\boldsymbol{\mathcal{Y}}_{\ell'}] = [\boldsymbol{\mathcal{Y}}(r_{\ell'})]$, and $[\boldsymbol{\mathcal{W}}_{\ell'}] = [\boldsymbol{\mathcal{W}}(r_{\ell'})]$, for $\ell' = 0, \ldots, M$, with $[\boldsymbol{\mathcal{Z}}_0] = [z_d]$, $[\boldsymbol{\mathcal{Y}}_0] = [\boldsymbol{\mathcal{N}}]\,[a]$, and $[\boldsymbol{\mathcal{W}}_0] = [0_{\nu,m}]$. For $\ell' = 0, \ldots, M-1$, let $[\Delta\boldsymbol{\mathcal{W}}_{\ell'+1}] = [\Delta\mathbb{W}_{\ell'+1}]\,[a]$ be the sequence of random matrices with values in $\mathbb{M}_{\nu,m}$, in which $[\Delta\mathbb{W}_{\ell'+1}] = [\mathbb{W}_{\ell'+1}] - [\mathbb{W}_{\ell'}]$. The increments $[\Delta\mathbb{W}_1], \ldots, [\Delta\mathbb{W}_M]$ are M independent random matrices with values in $\mathbb{M}_{\nu,N_d}$. For all $k = 1, \ldots, \nu$ and for all $j = 1, \ldots, N_d$, the real-valued random variables $\{[\Delta\mathbb{W}_{\ell'+1}]_{kj}\}_{kj}$ are independent, Gaussian, second-order, and centered random variables such that

$$E\{[\Delta\mathbb{W}_{\ell'+1}]_{kj}[\Delta\mathbb{W}_{\ell'+1}]_{k'j'}\} = \Delta r\,\delta_{kk'}\,\delta_{jj'}\,.$$

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

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

$$[\boldsymbol{\mathcal{Y}}_{\ell'+1}] = \frac{1-b}{1+b}\,[\boldsymbol{\mathcal{Y}}_{\ell'}] + \frac{\Delta r}{1+b}\,[\boldsymbol{\mathcal{L}}_{\ell'+\frac{1}{2}}] + \frac{\sqrt{f_0}}{1+b}\,[\Delta\boldsymbol{\mathcal{W}}_{\ell'+1}]\,,$$

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

with the initial condition defined before, where $b = f_0\,\Delta r\,/4$, and where $[\boldsymbol{\mathcal{L}}_{\ell'+\frac{1}{2}}]$ is the $\mathbb{M}_{\nu,m}$-valued random variable such that

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