

STATISTICAL INVERSE PROBLEMS FOR NON-GAUSSIAN NON-STATIONARY STOCHASTIC PROCESSES DEFINED BY A SET OF REALIZATIONS

G. Perrin¹ and C. Soize²

¹Commissariat à l’Energie Atomique et aux Energies Alternatives
CEA/DAM/DIF, F-91297, Arpajon, France
e-mail: guillaume.perrin2@cea.fr

² Université Paris-Est
MSME UMR 8208 CNRS, Marne-la-Vallée, France
e-mail: christian.soize@univ-paris-est.fr

Keywords: Stochastic processes, inverse problems, Karhunen-Loève expansion.

Abstract.

This paper presents a method to analyze the transitory response of complex and nonlinear systems, which are excited by non-Gaussian and non-stationary random fields, by solving of a statistical inverse problem with experimental measurements. Based on a double expansion, it is particularly adapted to the modeling of stochastic processes that are only characterized by a relatively small set of independent realizations. First, an adaptation of the classical Karhunen-Loève expansion is presented. Indeed, for the past fifty years, the use of reduced basis has spread to many scientific fields to condense the statistical properties of stochastic processes, and among these bases, the Karhunen-Loève basis plays a major role as it allows the minimization of the total mean square error. Secondly, the random vector, which gathers the projection coefficients of the stochastic process on this basis, is characterized using a polynomial chaos expansion approach. The dimension of this random vector being very high (around several hundreds), advanced identification techniques are introduced to allow performing relevant convergence analyses and identifications. The non-Gaussian non-stationary stochastic process is identified using the experimental measurements and consequently, constitutes a realistic stochastic modeling. The proposed method is then applied to the modeling of seismic accelerations from a measured data set.

1 INTRODUCTION

For the last decades, the use of simulation has kept increasing in the analysis of complex and non-linear mechanical systems that are excited by stochastic processes (wind, seism, wave). It is indeed a very interesting tool for robust and optimized conception, better knowledge of the critical situations, and more precise reliability analyses. In such analyses, much attention has therefore to be paid to the characterization of the input variability, as it completely drives the relevance of the results. However, the knowledge of these input stochastic processes is generally limited to a finite set of independent realizations that stem from experimental measurements. When such stochastic processes are indexed by parameters that are defined on a compact domain, methods based on a two-steps approach have given very interesting results to identify in inverse the distribution of *a priori* non-Gaussian and non-stationary stochastic processes. The first step of these methods is generally the approximation of the stochastic process, X , by its projection on a finite set of deterministic functions. Following on the work achieved in [1], this paper proposes thus an innovative method, which is particularly adapted to the seismic case, to construct optimized basis from a relatively small set of independent realizations. By introducing an original adaptation of the Karhunen-Loève expansion [2], such a method allows us to maximize the representativeness of the projection basis with respect to the available information.

The second step is then the identification of the multidimensional distribution of the vector gathering the projection coefficients associated with the former step. In that context, it will be shown to what extent the polynomial chaos expansion (PCE) method (see [3, 4, 5]) can be used to identify such a distribution in very high dimension, even if the number of available realizations is very small.

Section 2 introduces the method we propose to identify the statistical distribution of non-Gaussian and non-stationary stochastic processes from a set of independent realizations. Section 3 illustrates then the possibilities of such a two-steps method on the generation of seismic accelerations that are realistic and representative of an experimental data set.

2 STATISTICAL IDENTIFICATION

2.1 Theoretical frame

Let (Θ, \mathcal{C}, P) be a probability space. Let $\mathcal{H} = L^2(\Theta, \mathbb{R})$ be the space of all the second-order random vectors defined on (Θ, \mathcal{C}, P) with values in \mathbb{R} , equipped with the inner product $\langle \cdot, \cdot \rangle$, such that for all U and V in \mathcal{H} ,

$$\langle U, V \rangle = \int_{\Theta} U(\theta)V(\theta)dP(\theta) = \mathbb{E}[UV], \quad (1)$$

where $\mathbb{E}[\cdot]$ is the mathematical expectation. Let $\mathcal{P}(\Omega)$ be the space of all the second-order \mathbb{R} -valued stochastic processes, indexed by the compact interval $\Omega = [0, T]$, where $T < +\infty$. Let $\mathbb{H} = L^2(\Omega, \mathbb{R})$ be the space of square integrable functions on Ω , with values in \mathbb{R} , equipped with the inner product (\cdot, \cdot) , such that for all u and v in \mathbb{H} ,

$$(u, v) = \int_{\Omega} u(s)v(s)ds. \quad (2)$$

Let X be an element of $\mathcal{P}(\Omega)$. Without loss of generality, it is supposed that the mean value

of X is equal to zero:

$$\mathbb{E}[X(\cdot, t)] = 0, \forall t \in \Omega, \quad (3)$$

and we denote by R_X its covariance function, such that for all t, t' in Ω ,

$$R_X(t, t') = \mathbb{E}[X(\cdot, t)X(\cdot, t')]. \quad (4)$$

It is assumed that the maximum available information about X is given by ν independent realizations, $\{X(\theta_1, \cdot), \dots, X(\theta_\nu, \cdot)\}$. For any $M \geq 1$, based on this available set, the idea of the first step is to identify the M -dimension projection family $\mathcal{F}_{opt}^{(M)} = \{f_m, 1 \leq k \leq M\}$ that minimizes, among all the families $\mathcal{F}^{(M)}$ in \mathbb{H}^M , the amplitude of the truncation residue:

$$\mathcal{F}_{opt}^{(M)} = \arg \min_{\mathcal{F}^{(M)} \in \mathbb{H}^M} \|X - \hat{X}^{(M)}(\mathcal{F}^{(M)})\|_2^2, \quad (5)$$

$$\|X - \hat{X}^{(M)}(\mathcal{F}^{(M)})\|_2^2 \stackrel{\text{def}}{=} \mathbb{E} \left[\left(X - \hat{X}^{(M)}(\mathcal{F}^{(M)}), X - \hat{X}^{(M)}(\mathcal{F}^{(M)}) \right) \right], \quad (6)$$

where, for any family $\mathcal{F}^{(M)}$, $\hat{X}(\mathcal{F}^{(M)})$ is the projection of X on $\mathcal{F}^{(M)}$.

2.2 Optimality of the Karhunen-Loève decomposition

The classical Karhunen-Loève (KL) basis associated with X , $\mathcal{K} = \{k_m, 1 \leq m\}$, corresponds to the Hilbertian basis that is constructed as the eigenfunctions of covariance function R_X , such that:

$$\int_{\Omega} R_X(t, t') k_m(t') dt' = \lambda_m k_m(t), \quad (7)$$

$$(k_m, k_\ell) = \delta_{m\ell}, \quad \lambda_1 \geq \lambda_2 \geq \dots \rightarrow 0, \quad (8)$$

with $\delta_{m\ell}$ the kronecker symbol, which is equal to one if $m = \ell$ and zero otherwise. Equation (7) is generally called Fredholm problem (see [6] for further details about its solving). Such a KL basis keeps playing an important role for the definition of reduced basis as it can be shown that projection family $\mathcal{K}^{(M)} = \{k_m, 1 \leq m \leq M\}$ is optimal in the sense that, for all family $\mathcal{F}^{(M)}$ in \mathbb{H}^M and all $M \geq 1$:

$$\|X - \hat{X}(\mathcal{K}^{(M)})\|_2^2 \leq \|X - \hat{X}(\mathcal{F}^{(M)})\|_2^2. \quad (9)$$

2.3 Enrichment of the KL decomposition

As presented in Section 2.2, the KL basis is based on the knowledge of the covariance function, R_X . In cases when the maximal available information is characterized by a set of independent realizations, $\{X(\theta_1, \cdot), \dots, X(\theta_\nu, \cdot)\}$, such a covariance function is however not exactly known, but can be assessed by its empirical estimator, \hat{R}_X , such that:

$$R_X(t, t') \approx \hat{R}_X(t, t') \stackrel{\text{def}}{=} \frac{1}{\nu} \sum_{n=1}^{\nu} X(\theta_n, t) X(\theta_n, t'). \quad (10)$$

However, there is no reason for the projection basis computed from the solving of the Fredholm problem associated with \widehat{R}_X , $\widehat{\mathcal{K}} = \{\widehat{k}_m, 1 \leq m\}$, with:

$$\int_{\Omega} \widehat{R}_X(t, t') \widehat{k}_m(t') dt' = \widehat{\lambda}_m \widehat{k}_m(t), \quad \widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \rightarrow 0, \quad (11)$$

to be still optimal with respect to the minimization of error

$$\varepsilon^2(\mathcal{F}^{(M)}) \stackrel{\text{def}}{=} \left\| X - \widehat{X}(\mathcal{F}^{(M)}) \right\|_2^2, \quad \mathcal{F}^{(M)} \in \mathbb{H}^M. \quad (12)$$

More precisely, it can be noticed that, by construction of function \widehat{R} , only the ν first element of $\widehat{\mathcal{K}}$ are adapted to X , whereas the other elements are orthogonal to the available realizations:

$$(k_m, X(\theta_n, \cdot)) = 0, \quad m > \nu, \quad 1 \leq n \leq \nu. \quad (13)$$

When interested in the identification of projection basis whose dimension M is higher than ν , it has been shown in [1] that very promising results could be obtained from the solving of an optimization problem on the choice of the kernel on which the Fredholm problem is based. In particular, if we introduce function \widetilde{R} such that:

$$\widetilde{R}(t, t') = \begin{cases} \frac{1}{T+t'-t} \int_0^{T+t'-t} \widehat{R}(x+t-t', x) dx & \text{if } 0 \leq t-t' \leq T, \\ \frac{1}{T+t-t'} \int_0^{T+t-t'} \widehat{R}(x, x+t'-t) dx & \text{if } 0 \leq t'-t \leq T, \\ \widehat{R}(t, t') & \text{otherwise.} \end{cases} \quad (14)$$

solving the following optimization problem:

$$\alpha^* = \arg \min_{\alpha \in [0,1]} \varepsilon^2(\mathcal{B}^{(M)}(\alpha)), \quad (15)$$

where $\mathcal{B}^{(M)}(\alpha) = \{k_m^\alpha, 1 \leq m \leq M\}$ gathers the M solutions of the Fredholm problem associated with the kernel $A(\alpha, t, t') = \alpha \widetilde{R}(t, t') + (1 - \alpha) \widehat{R}(t, t')$ of highest eigenvalues, could allow us to identify very relevant projection families for X .

When confronted to very non-stationary stochastic processes, such an enrichment of the kernel can still be improved. In this prospect, for N_c an integer greater than 1, let $K(N_c, t, t')$ be the function defined on $[0, T]$ such that, for all $0 \leq p, q \leq N_c - 1$ and for all (t, t') in $[qT/N_c, (q+1)T/N_c] \times [pT/N_c, (p+1)T/N_c]$:

$$\widetilde{K}(N_c, t, t') = \begin{cases} \frac{1}{\frac{T}{N_c} + u' - u} \int_0^{\frac{T}{N_c} + u' - u} \widehat{R}(x + u - u', x) dx & \text{if } u' = t' - \frac{pT}{N_c} \leq u = t - \frac{qT}{N_c}, \\ \frac{1}{\frac{T}{N_c} + u - u'} \int_0^{\frac{T}{N_c} + u - u'} \widehat{R}(x, x + u - u') dx & \text{if } u = t - \frac{qT}{N_c} \leq u' = t' - \frac{pT}{N_c}, \\ \widehat{R}(t, t') & \text{otherwise.} \end{cases} \quad (16)$$

Indeed, whereas \widetilde{R} corresponds to a global enrichment, $\widetilde{K}(N_c, \cdot, \cdot)$ allows a more local enrichment by applying the transformation from \widehat{R} to \widetilde{R} not on the whole domain $[0, T] \times [0, T]$, but on a series of subdomains $[qT/N_c, (q+1)T/N_c] \times [pT/N_c, (p+1)T/N_c]$, $1 \leq p, q \leq N_c$.

Therefore, by solving the following updated optimization problem:

$$(\alpha^*, N_c^*) = \arg \min_{\alpha \in [0,1], N_c \geq 1} \varepsilon^2(\mathcal{B}^{(M)}(\alpha, N_c)), \quad (17)$$

with $\mathcal{B}^{(M)}(\alpha, N_c) = \{k_m^{\alpha, N_c}, 1 \leq m \leq M\}$ the set of solutions of the Fredholm problem associated with $A(\alpha, N_c, t, t') = \alpha \hat{R}(t, t') + (1 - \alpha) \tilde{K}(N_c, t, t')$ of highest eigenvalues, still lower values of M can be chosen for a required projection error.

2.4 Inverse polynomial chaos identification

Once stochastic process X has been projected on the reduced basis,

$$\mathcal{B}^{(M)}(\alpha^*, N_c^*) = \{b_1^{\alpha^*, N_c^*}, \dots, b_M^{\alpha^*, N_c^*}\}, \quad (18)$$

such that:

$$X \approx \hat{X}^* = \sum_{m=1}^M C_m^* k_m^{\alpha^*, N_c^*}, \quad (19)$$

where M has been chosen to guarantee that

$$\|X - \hat{X}^*\|_2^2 \leq e, \quad (20)$$

with $e > 0$ a required precision, the multidimensional distribution of the projection random vector, $\mathbf{C}^* = (C_1^*, \dots, C_N^*)$, has to be identified from the available set of ν independent realizations of \mathbf{C}^* , $\{\mathbf{C}^*(\theta_1), \dots, \mathbf{C}^*(\theta_\nu)\}$, with:

$$C_m^*(\theta_n) = (X(\theta_n, \cdot), k_m^{\alpha^*, N_c^*}), \quad 1 \leq n \leq \nu, \quad 1 \leq m \leq M. \quad (21)$$

When analyzing complex stochastic processes, M can be very large, and specific statistical identification methods have to be used. Among these methods, the polynomial chaos expansion (PCE) method [3] has given very promising results, even in cases when $M \gg \nu$ ([5, 4, 7]). This technique is based on a direct projection of \mathbf{C}^* on a polynomial hilbertian basis, $\{\psi_j(\boldsymbol{\xi}), 1 \leq j\}$, of all the second-order random vectors with values in \mathbb{R}^M , such that:

$$\mathbf{C}^* = \sum_{j=1}^{+\infty} \mathbf{y}^{(j)} \psi_j(\boldsymbol{\xi}), \quad (22)$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_M)$ is a random vector, whose distribution is chosen and known. In practice, such an infinite sum has to be truncated. Therefore, two truncation parameters are generally introduced, which are denoted by N and N_g , such that:

$$\mathbf{C}^* \approx \sum_{j=1}^N \mathbf{y}^{(j)} \psi_j(\xi_1, \dots, \xi_{N_g}). \quad (23)$$

For given values of N and N_g , the values of the PCE projection coefficients, $\{\mathbf{y}^{(j)}, 1 \leq j \leq N\}$, have now to be identified from the available independent realizations of \mathbf{C}^* (see [4, 5] for further details).

Finally, once all these parameters M, N, N_g are identified according to convergence analyses, once PCE projection coefficients $\{\mathbf{y}^{(j)}, 1 \leq j \leq N\}$ are computed, the statistical properties of X are completely known:

$$\mathbf{X} \approx \sum_{m=1}^M k_m^{\alpha^*, N_c^*} \sum_{j=1}^N y_m^{(j)} \psi_j(\xi_1, \dots, \xi_{N_g}), \quad (24)$$

and for each realization of $(\xi_1, \dots, \xi_{N_g})$, we have access to a particular realization of X , which is realistic and representative of the available realizations gathered in $\{X(\theta_1, \cdot), \dots, X(\theta_\nu, \cdot)\}$.

3 APPLICATION TO A SEISMIC DATA SET

Let $X = \{X(\theta, t), \theta \in \Theta, t \in \Omega\}$ be the ground acceleration associated with a seism of duration T , at a given spatial point and in a given direction. For confidentiality reasons, only few numerical values are given in the following. This section aims now at showing in what extent the complete parametrization presented in Section 2 can give very relevant results when trying to generate realistic and representative seismic accelerations, even if the available information is characterized by a relatively small set of independent realizations.

In this work, only $\nu = 97$ independent realizations of X are supposed to be known. As an illustration, three independent realizations of X are represented in Figure 1. Stochastic process X is by construction neither stationary nor Gaussian. From these ν independent realizations, it is possible to approximate the covariance function of X from its empirical estimator, which allows us to identify the projection basis $\hat{\mathcal{K}} = \{\hat{k}_m, 1 \leq m\}$. Figure 2 compares therefore the relevance of such a basis to the projection basis $\{\hat{k}_m^{\alpha^*}, 1 \leq m\}$ and $\{\hat{k}_m^{\alpha^*, N_c^*}, 1 \leq m\}$, which stem from the solving of Eqs. (15) and (17) respectively. Hence, to guarantee that the projection error is lower than 0.01, it can be seen that, whereas 1613 elements would have been needed for $\{\hat{k}_m, 1 \leq m\}$, only 614 are needed for $\{\hat{k}_m^{\alpha^*, N_c^*}, 1 \leq m\}$, with $\alpha^* = 0.22$ and $N_c^* = 10$. We then define \mathbf{C}^* as the projection vector of X on the first $M = 614$ elements $\{\hat{k}_m^{\alpha^*, N_c^*}, 1 \leq m \leq M\}$. From the ν independent realizations of X , ν realizations of \mathbf{C}^* are therefore deduced. A polynomial chaos expansion is thus computed to identify the multidimensional distribution of \mathbf{C}^* from this available information. The germ, on which this expansion is based, is chosen as a N_g -dimension random vector, for which components are independent and uniformly distributed between -1 and 1. To maximize the likelihood of the PCE approximation of \mathbf{C}^* at the ν available realizations, the values of N and N_g are chosen equal to 1365 and 4 respectively. The PCE projection coefficients have then been identified according to the advanced algorithms defined in [4, 5].

At last, from any independent realization of ξ , it is possible to generate a realistic and representative realization of X .

4 CONCLUSION

When the available information is a finite set of independent realizations, a method to identify the statistical characteristics of a non-Gaussian and non-stationary stochastic process has been presented in this work. Based on an adaptation of the Karhunen Loève expansion and a polynomial chaos expansion, this methods allows us to generate new realizations of this stochastic process, which are representative of the measured ones, from a frequency and statistical point of view. Such a method has been applied to the modeling of random seismic accelerations, in order to make possible the stochastic analysis of the dynamical response of non-linear structures.

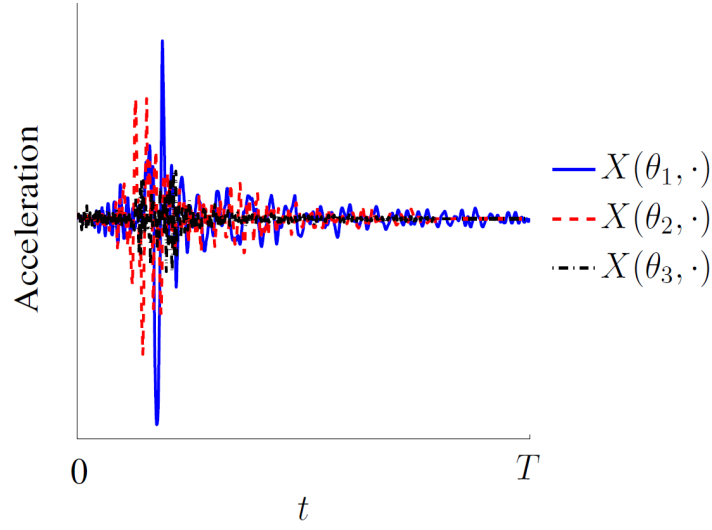
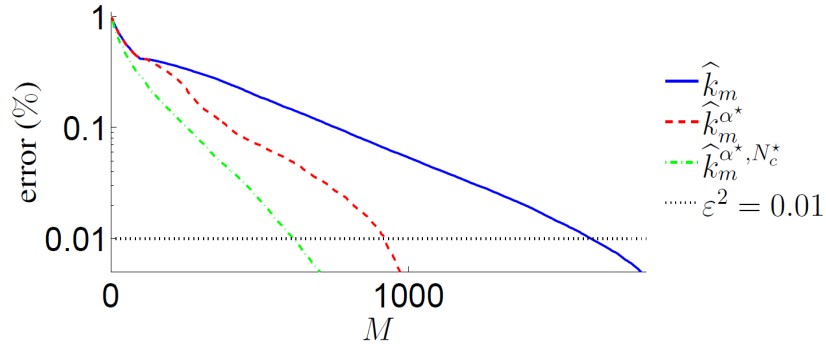
Figure 1: Representation of three independent realizations of X .

Figure 2: Comparison of the error evolutions of projection basis $\{k_m, 1 \leq m\}$, $\{k_m^{\alpha*}, 1 \leq m\}$ and $\{k_m^{\alpha*, N_c^*}, 1 \leq m\}$ with respect to M . These errors are computed from the only ν available realizations from a leave-one-out procedure.

REFERENCES

- [1] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, A Posteriori Error and Optimal Reduced Basis for Stochastic Processes Defined by a Finite Set of Realizations, *SIAM/ASA J. Uncertainty Quantification*, **2**, 745-762, 2014.
- [2] M. Loève, Probability Theory, fourth ed., *Springer-Verlag, Berlin* , 1977.
- [3] R. Ghanem and P.D. Spanos, Polynomial chaos in stochastic finite elements, *Journal of Applied Mechanics*, 197-202, 1990.
- [4] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Identification of polynomial chaos representations in high dimension from a set of realizations, *SIAM J. Sci. Comput.*, **34(6)**, 2917-2945, 2012.
- [5] C. Soize, Identification of high-dimension polynomial chaos expansions with random coefficients for non-Gaussian tensor-valued random fields using partial and limited experimental data, *Computer Methods in Applied Mechanics and Engineering*, **199**, 2150-2164, 2010.
- [6] O. Le Maître, O. Knio, Spectral Methods for Uncertainty Quantification, *Springer*, 2010.
- [7] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Track irregularities stochastic modeling, *Probabilistic Engineering Mechanics*, **34**, 123-130, 2013.