

Computation of Sobol indices in global sensitivity analysis from small data sets by probabilistic learning on manifolds

Maarten Arnst¹, Christian Soize², and Kevin Bulthuis¹

¹Université de Liège, Aérospatiale et Mécanique, Quartier Polytech 1, allée de la Découverte 9, 4000 Liège, Belgium (maarten.arnst@uliege.be)

²Université Paris-Est Marne-la-Vallée, Modélisation et Simulation Multi-Echelle, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée, France

Abstract

Global sensitivity analysis provides insight into how sources of uncertainty contribute to uncertainty in predictions of computational models. Global sensitivity indices, also called variance-based sensitivity indices and Sobol indices, are most often computed with Monte Carlo methods. However, when the computational model is computationally expensive and only a small number of samples can be generated, that is, in so-called small-data settings, usual Monte Carlo estimates may lack sufficient accuracy. As a means of improving accuracy in such small-data settings, we explore the use of probabilistic learning. The objective of the probabilistic learning is to learn from the available samples a probabilistic model that can be used to generate additional samples, from which Monte Carlo estimates of the global sensitivity indices are then deduced. We demonstrate the interest of such a probabilistic learning method by applying it in an illustration concerned with forecasting the contribution of the Antarctic ice sheet to sea-level rise.

Keywords: global sensitivity analysis, Sobol index, probabilistic learning on manifolds, small data

1 Introduction

Global sensitivity analysis is concerned with understanding how uncertainties in input variables of a computational model induce uncertainty in a predicted quantity of interest. Global sensitivity analysis is a probabilistic method based on describing the uncertainties in the input variables with a probabilistic representation and then analyzing the induced probabilistic representation of the quantity of interest. More specifically, global sensitivity analysis provides for each uncertain input variable a global sensitivity index, also called variance-based sensitivity index as well as Sobol index, that quantifies the relative contribution of the uncertainty in this input variable to the variance of the quantity of interest. These global sensitivity indices allow the uncertain input variables to be arranged in an order that reflects their significance in inducing uncertainty in the quantity of interest, thus permitting dominant uncertain input variables to be distinguished from less important ones. Gaining such understanding can serve to identify where to direct efforts aimed at reducing uncertainties and hence be useful in robust design optimization under uncertainty, model enrichment and model validation, and many other purposes; see for instance, [Alhossen et al., 2019, Arnst and Ponthot, 2014, Nossent et al., 2011, Rosolem et al., 2012]. For a recent survey of the current state of the art in global and other types of sensitivity analysis, we refer to Part 4 “Introduction to sensitivity analysis” in the recent “Handbook of Uncertainty Quantification” [Ghanem et al., 2017]. Global sensitivity indices are most often defined based on a decomposition of the function that the

computational model establishes between the input variables and the quantity of interest into an expansion of component functions. These component functions are defined as solutions to regression problems in which only individual input variables or only subsets of input variables serve as regressors, and mean-square norms of these component functions are then used to define the global sensitivity indices [Owen, 2013, Sobol, 2001]. Equivalently, the global sensitivity indices can also be defined directly, and without explicit reference to such a decomposition into an expansion of component functions, as variances of conditional expectations, see, for instance, [Oakley and O'Hagan, 2004, Owen, 2013, Saltelli, 2002].

The aforementioned different ways of defining global sensitivity indices are the basis for different ways of numerically computing global sensitivity indices. Monte Carlo methods are most often used to compute global sensitivity indices. Their principle is to approximate the mathematical expectations involved in the expression of the global sensitivity indices as variances of conditional expectations by using Monte Carlo integration. To deal with the nested structure of this expression, such Monte Carlo methods most often take the form of so-called fixing methods, also called pick-freeze methods, which require the computational model to be evaluated for pairs of samples of the input variables that share the same values for some of the input variables but involve different, independently sampled values for the others [Owen, 2013, Saltelli, 2002, Sobol, 2001]. There exist central limit theorems for the Monte Carlo estimators thus obtained, which indicate that accuracy improves with the square root of the number of samples [Janon et al., 2014a]. Especially when the computational model is computationally expensive and only a small number of samples can be generated, this rather slow rate of convergence may lead to such Monte Carlo estimates lacking sufficient accuracy.

To alleviate this issue, numerical methods have been proposed that are able to use available evaluations of the computational model more effectively when special structure is present within the probability distribution of the uncertain input variables and the function that the computational model establishes between the input variables and the quantity of interest. Variants of the aforementioned Monte Carlo methods have been proposed that involve approximating the aforementioned mathematical expectations by using other numerical integration methods, such as Gaussian and other types of deterministic quadrature, quasi Monte Carlo integration, and multilevel and multifidelity Monte Carlo integration, see, for instance, [Lamboni, 2016, Qian et al., 2018]. Further, numerical methods have been proposed that involve approximating the aforementioned regression functions by using nonparametric kernel regression methods [Scott, 2015], see, for instance, [Janon et al., 2014b, Luo et al., 2014]. Finally, numerical methods have been proposed that involve the use of metamodels, including numerical methods that involve the approximation of the function that the computational model establishes between the input variables and the quantity of interest with a Gaussian process [Oakley and O'Hagan, 2004] and numerical methods that involve the approximation of the aforementioned expansion of component functions by using a polynomial chaos expansion, which allows the global sensitivity indices to be deduced directly from the polynomial chaos coefficients [Crestaux et al., 2009, Sudret, 2008]. To work well, all these numerical methods require special structure to be present. The use of a Gaussian process requires that the function that the computational model establishes between the input variables and the quantity of interest be sufficiently smooth and can be approximated as a trajectory of a Gaussian process with a trend and a covariance function that can be identified from the available samples. The use of Gaussian quadrature and polynomial chaos expansions require this function to lend itself well to being approximated with a polynomial of sufficiently low order.

In this paper, we propose an alternative numerical method that is specifically intended for settings in which the computational model is computationally expensive and can be evaluated for only a small number of samples. As a means of improving accuracy in such small-data settings, we explore the use of probabilistic learning. The objective of the probabilistic learning is to learn from the available samples a probabilistic model that can be used to generate additional samples, on the basis of which

the global sensitivity indices are then computed, ideally with higher accuracy. Clearly, as for the aforementioned numerical methods, to work well, such a probabilistic learning method requires special structure to be present, albeit of a different kind than for the aforementioned numerical methods. Here, the available samples must in some way encapsulate information that can be used to generate additional samples that are statistically consistent with the available samples, that is, so that the additional samples can be considered as being sampled from the same probability distribution.

In this paper, we use the *probabilistic learning on manifolds* (PLoM) method that has been recently introduced by Soize and Ghanem in [Soize and Ghanem, 2016] with complementary developments in Ghanem and Soize [2018], Soize and Ghanem [2020a,b], Soize et al. [2019] and applications and validations in Farhat et al. [2019], Ghanem et al. [2018, 2019], Soize and Farhat [2019]. The PLoM method allows available samples to be used more effectively by generating statistically consistent additional samples, especially when the samples concentrate near a lower dimensional manifold. Please note that we do not seek in this paper to provide new validations and performance evaluations of the PLoM method. Rather, we use the PLoM method as a particular instance of a probabilistic learning method in order to propose and demonstrate the interest of probabilistic learning in the computation of global sensitivity indices in small-data settings. In this regard, it should also be noted that while we focus on the computation of (variance-based) global sensitivity indices in this paper, the proposed use of probabilistic learning extends to other types of (global) sensitivity analysis, including sensitivity-analysis methods using divergence functions from statistics and information geometry Borgonovo [2007], Da Veiga [2014], Da Veiga and Gamboa [2013], Laurent [1996].

This paper is organized as follows. First, in Secs. 2 and 3, we recall the definition of global sensitivity indices and the main numerical methods available for their computation. Then, in Sec. 4, we describe the proposed PLoM method for small-data settings. Subsequently, Sec. 5 is devoted to the illustration. Finally, after the conclusion in Sec. 6, in A, we provide further details about the PLoM method and its implementation.

2 Global sensitivity indices

We will now recall the definition of global sensitivity indices. Let us consider a computational model that establishes a (typically nonlinear) function of a set of input variables into a predicted quantity of interest, written as

$$q = f(w_1, \dots, w_{n_w}), \quad f : \mathbb{R}^{n_w} \rightarrow \mathbb{R}, \quad (1)$$

in which w_1, \dots, w_{n_w} are the values of the n_w input variables and q is the value of the quantity of interest. We assume that the values of the input variables and the value of the quantity of interest are scalars. But please note that most of the material to follow can be readily extended to settings in which they are vectors, matrices, or even of other types.

The values of the input variables w_1, \dots, w_{n_w} are collected in a vector $\mathbf{w} = (w_1, \dots, w_{n_w})$. The indices are collected in an index set denoted by $\mathcal{I} = \{1, \dots, n_w\}$. In global sensitivity analysis, these indices are also called factors, and this terminology will be used in the following. For two subsets of factors u and v , we denote by $u \subseteq v$ that u is a subset of v and by $u \subset v$ that u is a proper subset of v (which could also have been written more explicitly as $u \subsetneq v$). We denote by $|u|$ the cardinality of u and by $-u$ the complement $\mathcal{I} \setminus u$ of u in \mathcal{I} . For the subset of factors $u = \{k_1, \dots, k_{|u|}\}$, we denote by \mathbf{w}_u the subvector of values of the input variables $\mathbf{w}_u = (w_{k_1}, \dots, w_{k_{|u|}})$.

Global sensitivity analysis is a probabilistic method based on describing uncertainties in the input variables with a probabilistic representation. In global sensitivity analysis, it is most often assumed that the uncertain input variables are independent, and we do so here, thus resulting in a joint

probability distribution for the uncertain input variables

$$P_{\mathbf{W}} = P_{W_1} \otimes \dots \otimes P_{W_{n_w}} \quad (2)$$

that is the product of the marginal probability distributions $P_{W_1}, \dots, P_{W_{n_w}}$ for the individual uncertain input variables. Here and throughout the remainder of this paper, uppercase letters are used for random variables and lowercase letters for their realizations. Clearly, for the subset of factors $u = \{k_1, \dots, k_{|u|}\}$, the joint probability distribution

$$P_{\mathbf{W}_u} = P_{W_{k_1}} \otimes \dots \otimes P_{W_{k_{|u|}}} \quad (3)$$

is also the product of the marginal probability distributions $P_{W_{k_1}}, \dots, P_{W_{k_{|u|}}}$ for the uncertain input variables indexed by $k_1, \dots, k_{|u|}$. Please note that most of the material to follow can be extended to settings in which the uncertain input variables are dependent, but certain properties to follow, such as the orthogonality of the component functions and the normalized variance components forming a partition of unity, then do not hold; please see also [Chastaing et al., 2012, 2015, Hart and Gremaud, 2018, Tarantola and Mara, 2017].

Global sensitivity analysis then analyzes the induced probabilistic representation of the quantity of interest:

$$Q = f(\mathbf{W}). \quad (4)$$

For the material to follow to be well-posed, the uncertain quantity of interest must be of the second order:

$$\int_{\mathbb{R}} |q|^2 dP_Q(q) = \int_{\mathbb{R}^{n_w}} |f(\mathbf{w})|^2 dP_{\mathbf{W}}(\mathbf{w}) < +\infty, \quad (5)$$

in which P_Q is the probability distribution of the uncertain quantity of interest, that is, P_Q is the image of $P_{\mathbf{W}}$ under f .

Global sensitivity indices are most often defined based on a decomposition of the function f introduced in (1) into an expansion of component functions:

$$f(\mathbf{w}) = \sum_{u \subseteq \mathcal{I}} f_u(\mathbf{w}_u), \quad (6)$$

in which the component functions f_u , also called effects, as we will do in the following, are recursively defined by

$$f_u(\mathbf{w}_u) = f_u^*(\mathbf{w}_u) - \sum_{v \subset u} f_v(\mathbf{w}_v), \quad (7)$$

with

$$f_u^* = \arg \min_{g_u^* \in L^2_{P_{\mathbf{W}_u}}(\mathbb{R}^{|u|}, \mathbb{R})} \int_{\mathbb{R}^{n_w}} |f(\mathbf{w}) - g_u^*(\mathbf{w}_u)|^2 dP_{\mathbf{W}}(\mathbf{w}). \quad (8)$$

Thus, for the subset of factors $u = \{k_1, \dots, k_{|u|}\}$, the effect f_u is obtained by determining the regression function f_u^* that solves the regression problem that seeks the mean-square-best approximation of the function f of $\mathbf{w} = (w_1, \dots, w_{n_w})$ with a function of only $\mathbf{w}_u = (w_{k_1}, \dots, w_{k_{|u|}})$ and then subtracting from it the effects f_v for the subsets of factors v for which v is a proper subset of u . As usual in regression problems, the regression function f_u^* is the orthogonal projection of f onto the subspace of functions of only \mathbf{w}_u in the geometry of the space of $P_{\mathbf{W}}$ -square-integrable functions. And, as can also be deduced from a simple calculation within the calculus of variations, it is also obtained by averaging f over the uncertain input variables indexed by the factors not in u :

$$f_u^*(\mathbf{w}_u) = \int_{\mathbb{R}^{n_w - |u|}} f(\mathbf{w}) dP_{\mathbf{W}_{-u}}(\mathbf{w}_{-u}). \quad (9)$$

The effect f_\emptyset for the subset of factors taken as the empty set \emptyset is obtained by averaging f over all the uncertain input variables, that is, it is the mean \underline{q} of the uncertain quantity of interest:

$$f_\emptyset = \underline{q} = \int_{\mathbb{R}^{n_w}} f(\mathbf{w}) dP_{\mathbf{W}}(\mathbf{w}). \quad (10)$$

For subsets of factors with a cardinality of 1, that is, for singletons, the effects are also called main effects. The main effect $f_{\{k\}}$ for the singleton $\{k\}$, which we will denote more concisely by f_k , is given by

$$f_k(w_k) = f_k^*(w_k) - f_\emptyset = \int_{\mathbb{R}^{n_w-1}} f(w_1, \dots, w_{n_w}) dP_{W_1}(w_1) \dots dP_{W_{k-1}}(w_{k-1}) dP_{W_{k+1}}(w_{k+1}) \dots dP_{W_{n_w}}(w_{n_w}) - \underline{q}. \quad (11)$$

For subsets of factors with a cardinality larger than 1, the effects are also called interaction effects. Because they are obtained via orthogonal projections in the geometry of the space of $P_{\mathbf{W}}$ -square-integrable functions, the effects are orthogonal functions in this geometry of the space of $P_{\mathbf{W}}$ -square-integrable functions:

$$\int_{\mathbb{R}^{n_w}} f_u(\mathbf{w}_u) f_v(\mathbf{w}_v) dP_{\mathbf{W}}(\mathbf{w}) = 0 \quad \text{if } u \neq v. \quad (12)$$

It follows from (10) and (12) that

$$\int_{\mathbb{R}^{|u|}} f_u(\mathbf{w}_u) dP_{\mathbf{W}_u}(\mathbf{w}_u) = 0 \quad \text{if } u \neq \emptyset. \quad (13)$$

Bringing f_\emptyset , that is, \underline{q} , to the left-hand side in (6) and averaging the square of both hand sides over all the uncertain input variables results in the variance decomposition

$$V = \sum_{\substack{u \subseteq \mathcal{I} \\ u \neq \emptyset}} V_u, \quad (14)$$

in which it is the square of the mean-square norm thus obtained in the left-hand side that provides the variance V of uncertain quantity of interest in the left-hand side:

$$V = \int_{\mathbb{R}^{n_w}} |f(\mathbf{w}) - \underline{q}|^2 dP_{\mathbf{W}}(\mathbf{w}), \quad (15)$$

and the squares of the mean-square norms thus obtained in the right-hand side provide the so-called variance components V_u in the right-hand side:

$$V_u = \int_{\mathbb{R}^{|u|}} |f_u(\mathbf{w}_u)|^2 dP_{\mathbf{W}_u}(\mathbf{w}_u) \quad \text{if } u \neq \emptyset. \quad (16)$$

Because the effects are orthogonal functions in the geometry of the space of $P_{\mathbf{W}}$ -square-integrable functions, there are no double-product terms in this variance decomposition. Clearly, by dividing both hand sides in (14) by the variance, it is seen that the normalized variance components form a partition of unity:

$$1 = \sum_{\substack{u \subseteq \mathcal{I} \\ u \neq \emptyset}} \frac{V_u}{V}. \quad (17)$$

Based on the variance components in the right-hand side in (14), global sensitivity analysis defines for each subset of factors $u = \{k_1, \dots, k_{|u|}\}$ with $|u| \geq 1$ an unnormalized (also called closed) global sensitivity index as

$$D_u = \sum_{\substack{v \subseteq u \\ v \neq \emptyset}} V_v; \quad (18)$$

it is interpreted as the portion of the variance of the uncertain quantity of interest that is explained as stemming from the uncertain input variables indexed by the factors $k_1, \dots, k_{|u|}$. Based on these unnormalized global sensitivity indices, global sensitivity analysis defines (normalized) global sensitivity indices as

$$S_u = \frac{D_u}{V}. \quad (19)$$

Because the effects are orthogonal functions in the geometry of the space of $P_{\mathbf{W}}$ -square-integrable functions, the unnormalized global sensitivity index D_u is also obtained as

$$\begin{aligned} D_u &= \sum_{\substack{v \subseteq u \\ v \neq \emptyset}} \int_{\mathbb{R}^{|v|}} |f_v(\mathbf{w}_v)|^2 dP_{\mathbf{W}_v}(\mathbf{w}_v) \\ &= \int_{\mathbb{R}^{|u|}} \left| f_u^*(\mathbf{w}_u) - \sum_{\substack{v \subseteq u \\ v \neq \emptyset}} f_v(\mathbf{w}_v) - \underline{q} \right|^2 dP_{\mathbf{W}_u}(\mathbf{w}_u) + \sum_{\substack{v \subseteq u \\ v \neq \emptyset}} \int_{\mathbb{R}^{|v|}} |f_v(\mathbf{w}_v)|^2 dP_{\mathbf{W}_v}(\mathbf{w}_v) \\ &= \int_{\mathbb{R}^{|u|}} \left| f_u^*(\mathbf{w}_u) - \underline{q} \right|^2 dP_{\mathbf{W}_u}(\mathbf{w}_u), \end{aligned} \quad (20)$$

and, hence, with (9), it is also obtained as

$$D_u = \int_{\mathbb{R}^{|u|}} \left| \int_{\mathbb{R}^{n_w - |u|}} f(\mathbf{w}) dP_{\mathbf{W}_{-u}}(\mathbf{w}_{-u}) - \underline{q} \right|^2 dP_{\mathbf{W}_u}(\mathbf{w}_u) = \int_{\mathbb{R}^{|u|}} \left| \int_{\mathbb{R}^{n_w - |u|}} f(\mathbf{w}) dP_{\mathbf{W}_{-u}}(\mathbf{w}_{-u}) \right|^2 dP_{\mathbf{W}_u}(\mathbf{w}_u) - \underline{q}^2. \quad (21)$$

As usual in regression problems, the expression of the regression function f_u^* in (9) may also be viewed as providing $f_u^*(\mathbf{w}_u)$ as the conditional expectation $E\{Q|\mathbf{W}_u = \mathbf{w}_u\}$ of $Q = f(\mathbf{W})$ given $\mathbf{W}_u = \mathbf{w}_u$, thus allowing the decomposition of f into an expansion of component functions in (6) to be written equivalently as follows:

$$q = f(w_1, \dots, w_{n_w}) = f_\emptyset + \sum_{k=1}^{n_w} f_k(w_k) + \sum_{j < k} f_{\{j,k\}}(w_j, w_k) + \sum_{i < j < k} f_{\{i,j,k\}}(w_i, w_j, w_k) + \dots + f_{\mathcal{I}}(w_1, \dots, w_{n_w}) \quad (22)$$

in which

$$f_\emptyset = \underline{q} = E\{Q\}, \quad (23)$$

the main effects are given by

$$f_k(w_k) = E\{Q|W_k = w_k\} - E\{Q\}, \quad (24)$$

the interaction effects for pairs of factors are given by

$$f_{\{j,k\}}(w_j, w_k) = E\{Q|(W_j, W_k) = (w_j, w_k)\} - f_j(w_j) - f_k(w_k) - E\{Q\} \quad (25)$$

and the interaction effects for triples of factors are given by

$$\begin{aligned} f_{\{i,j,k\}}(w_i, w_j, w_k) &= E\{Q|(W_i, W_j, W_k) = (w_i, w_j, w_k)\} \\ &\quad - f_{\{i,j\}}(w_i, w_j) - f_{\{i,k\}}(w_i, w_k) - f_{\{j,k\}}(w_j, w_k) - f_i(w_i) - f_j(w_j) - f_k(w_k) - E\{Q\}. \end{aligned} \quad (26)$$

With this equivalence, the expression in (21) provides the unnormalized global sensitivity index D_u as the variance of the conditional expectation $E\{Q|\mathbf{W}_u = \mathbf{w}_u\}$:

$$D_u = \text{Var}\{f_u^*(\mathbf{W}_u)\} \quad \text{with} \quad f_u^*(\mathbf{w}_u) = E\{Q|\mathbf{W}_u = \mathbf{w}_u\}. \quad (27)$$

3 Computation of global sensitivity indices

The different ways of defining global sensitivity indices, which we recalled in the previous section, are the basis for different ways of numerically computing global sensitivity indices, the main variants of which we will now recall.

3.1 Monte Carlo method

For the estimation of global sensitivity indices, the principle of Monte Carlo methods is to approximate the mathematical expectations involved in the expression of the unnormalized global sensitivity indices as variances of conditional expectations as in (21) and equivalently (27) by using Monte Carlo integration. In these equations, the expression for D_u has a nested structure: an outer integral averages over the uncertain input variables indexed by the factors in u the square of an expression that involves an inner integral that averages over the uncertain input variables indexed by the factors not in u the function f . Because of this nested structure, a straightforward application of a Monte Carlo method would require the computational model to be evaluated for the product of the numbers of samples used for Monte Carlo estimations of the outer integral and the inner integral. To avoid the high number of evaluations of the computational model that such an approach would require, Monte Carlo methods for the estimation of global sensitivity indices remove this nested structure by increasing the multiplicity of the inner integral [Owen, 2013, Saltelli, 2002, Sobol, 2001]:

$$D_u = \int_{\mathbb{R}^{|u|}} \int_{\mathbb{R}^{n_w - |u|}} f(\mathbf{w}_u : \mathbf{w}_{-u}) dP_{\mathbf{W}_{-u}}(\mathbf{w}_{-u}) \int_{\mathbb{R}^{n_w - |u|}} f(\mathbf{w}_u : \tilde{\mathbf{w}}_{-u}) dP_{\mathbf{W}_{-u}}(\tilde{\mathbf{w}}_{-u}) dP_{\mathbf{W}_u}(\mathbf{w}_u) - q^2.$$

Here, the notation involving the colon signifies that a vector $\mathbf{z} = (\mathbf{v}_u : \mathbf{w}_{-u})$ is such that $z_k = v_k$ if k is in u and $z_k = w_k$ otherwise. Thus, Monte Carlo methods for the estimation of global sensitivity indices avoid the iterated integral and take the form of so-called fixing methods that use two sets of samples of the uncertain input variables along with the corresponding samples of the uncertain quantity of interested evaluated with the computational model:

$$\{(q^{(j)}, \mathbf{w}_u^{(j)} : \mathbf{w}_{-u}^{(j)}), 1 \leq j \leq N\}, \quad \text{in which } q^{(j)} = f(\mathbf{w}_u^{(j)} : \mathbf{w}_{-u}^{(j)}), \quad (28)$$

$$\{(\tilde{q}^{(j)}, \mathbf{w}_u^{(j)} : \tilde{\mathbf{w}}_{-u}^{(j)}), 1 \leq j \leq N\}, \quad \text{in which } \tilde{q}^{(j)} = f(\mathbf{w}_u^{(j)} : \tilde{\mathbf{w}}_{-u}^{(j)}), \quad (29)$$

whereby the samples of the uncertain input variables form pairs that share the same independently sampled values for the uncertain input variables indexed by the factors in u but involve different, independently sampled values for the uncertain input variables indexed by the factors not in u . The Monte Carlo estimate D_u^{MC} of D_u is then obtained as

$$D_u^{\text{MC}} = \frac{1}{N} \sum_{j=1}^N f(\mathbf{w}_u^{(j)} : \mathbf{w}_{-u}^{(j)}) f(\mathbf{w}_u^{(j)} : \tilde{\mathbf{w}}_{-u}^{(j)}) - \left(\frac{1}{N} \sum_{j=1}^N f(\mathbf{w}_u^{(j)} : \mathbf{w}_{-u}^{(j)}) \right)^2 = \frac{1}{N} \sum_{j=1}^N q^{(j)} \tilde{q}^{(j)} - \left(\frac{1}{N} \sum_{j=1}^N q^{(j)} \right)^2. \quad (30)$$

Estimating in this way the main effect global sensitivity indices for n_w uncertain input variables requires a total of $N \times (n_w + 1)$ evaluations of the computational model. There exist central limit theorems for such Monte Carlo estimators, which indicate that accuracy improves with the square root of the number of samples N , see, for instance, [Janon et al., 2014a].

3.2 Nonparametric kernel method

The principle of nonparametric kernel methods for the estimation of global sensitivity indices is to approximate the regression function involved in the expression of the unnormalized global sensitivity indices as variances of conditional expectations as in (21) and equivalently (27) by using

nonparametric kernel regression. Such nonparametric kernel methods involve rewriting the regression function in these equations equivalently as follows:

$$f_u^*(\mathbf{w}_u) = E\{Q|\mathbf{W}_u = \mathbf{w}_u\} = \int_{\mathbb{R}} q p_{Q|\mathbf{W}_u}(q|\mathbf{w}_u) dq = \frac{\int_{\mathbb{R}} q p_{(Q,\mathbf{W}_u)}(q,\mathbf{w}_u) dq}{\int_{\mathbb{R}} p_{(Q,\mathbf{W}_u)}(q,\mathbf{w}_u) dq} \quad (31)$$

and then approximating $p_{(Q,\mathbf{W}_u)}$ by using kernel density estimation, see, for instance, [Luo et al., 2014]. In (31), $p_{(Q,\mathbf{W}_u)}$ is the joint probability density function (PDF) of the uncertain quantity of interest and the uncertain input variables indexed by the factors in u , and $p_{Q|\mathbf{W}_u}$ is the conditional PDF of the uncertain quantity of interest given the uncertain input variables indexed by the factors in u . Based on a set of independently sampled values of the uncertain input variables and the corresponding samples of the quantity of interest evaluated by using the computational model,

$$\{(q^{(j)}, \mathbf{w}^{(j)}), 1 \leq j \leq N\}, \quad \text{in which } q^{(j)} = f(\mathbf{w}^{(j)}), \quad (32)$$

the kernel density estimate $p_{(Q,\mathbf{W}_u)}^{\text{KDE}}$ of $p_{(Q,\mathbf{W}_u)}$ takes the form

$$p_{(Q,\mathbf{W}_u)}^{\text{KDE}}(q, \mathbf{w}_u) = \frac{1}{N} \sum_{j=1}^N K_Q(q - q^{(j)}) K_{\mathbf{W}_u}(\mathbf{w}_u - \mathbf{w}_u^{(j)}), \quad (33)$$

in which K_Q is a univariate kernel with appropriately chosen bandwidth parameter for the uncertain quantity of interest and $K_{\mathbf{W}_u}$ is a $|u|$ -variate kernel with appropriately chosen bandwidth parameters for the uncertain input variables indexed by the factors in u . In this paper, we use Gaussian kernels with bandwidth parameters chosen in accordance with Silverman's rule of thumb, see, for instance, [Scott, 2015], a reference to which we also refer for details about alternative kernels and choices of values of bandwidth parameters. Provided that K_Q satisfies the moment conditions

$$\int_{\mathbb{R}} K_Q(q - q^{(j)}) dq = 1, \quad \int_{\mathbb{R}} q K_Q(q - q^{(j)}) dq = q^{(j)}, \quad (34)$$

introducing the kernel density estimate in the numerator and denominator in (31) leads to

$$\int_{\mathbb{R}} q p_{(Q,\mathbf{W}_u)}^{\text{KDE}}(q, \mathbf{w}_u) dq = \frac{1}{N} \sum_{j=1}^N \underbrace{\int_{\mathbb{R}} q K_Q(q - q^{(j)}) dq}_{=q^{(j)}} K_{\mathbf{W}_u}(\mathbf{w}_u - \mathbf{w}_u^{(j)}) = \frac{1}{N} \sum_{j=1}^N q^{(j)} K_{\mathbf{W}_u}(\mathbf{w}_u - \mathbf{w}_u^{(j)}), \quad (35)$$

$$\int_{\mathbb{R}} p_{(Q,\mathbf{W}_u)}^{\text{KDE}}(q, \mathbf{w}_u) dq = \frac{1}{N} \sum_{j=1}^N \underbrace{\int_{\mathbb{R}} K_Q(q - q^{(j)}) dq}_{=1} K_{\mathbf{W}_u}(\mathbf{w}_u - \mathbf{w}_u^{(j)}) = \frac{1}{N} \sum_{j=1}^N K_{\mathbf{W}_u}(\mathbf{w}_u - \mathbf{w}_u^{(j)}). \quad (36)$$

Therefore, the following nonparametric kernel regression estimate is obtained:

$$f_u^{\star, \text{KDE}}(\mathbf{w}_u) = \frac{\frac{1}{N} \sum_{j=1}^N q^{(j)} K_{\mathbf{W}_u}(\mathbf{w}_u - \mathbf{w}_u^{(j)})}{\frac{1}{N} \sum_{j=1}^N K_{\mathbf{W}_u}(\mathbf{w}_u - \mathbf{w}_u^{(j)})}. \quad (37)$$

By using a quadrature rule for integration with respect to the kernel density estimate of the PDF of the uncertain input variables indexed by the factors in u , specifically, by using a quadrature rule with N_G nodes $\mathbf{w}_{G,u}^{(i)}$ and weights $v_G^{(i)}$ constructed so that the integral $\int_{\mathbb{R}^{|u|}} g(\mathbf{w}_u) \frac{1}{N} \sum_{j=1}^N K_{\mathbf{W}_u}(\mathbf{w}_u - \mathbf{w}_u^{(j)}) d\mathbf{w}_u$ of a function g from $\mathbb{R}^{|u|}$ into \mathbb{R} , with the kernel density estimate of the PDF of the

uncertain input variables indexed by the factors in u acting as weight function, is approximated with $\sum_{j=1}^{N_G} g(\mathbf{w}_{G,u}^{(j)}) v_G^{(j)}$, the nonparametric kernel estimate D_u^{KDE} of D_u is obtained as

$$D_u^{\text{KDE}} = \sum_{i=1}^{N_G} \left(\frac{\frac{1}{N} \sum_{j=1}^N q^{(j)} K_{\mathbf{W}_u}(\mathbf{w}_{G,u}^{(i)} - \mathbf{w}_u^{(j)})}{\frac{1}{N} \sum_{j=1}^N K_{\mathbf{W}_u}(\mathbf{w}_{G,u}^{(i)} - \mathbf{w}_u^{(j)})} \right)^2 v_G^{(i)} - \left(\frac{1}{N} \sum_{j=1}^N q^{(j)} \right)^2. \quad (38)$$

In this paper, to estimate main-effect global sensitivity indices, we use a trapezoidal quadrature rule. But please note that alternative quadrature rules, such as Gaussian quadrature or Monte Carlo integration, can also be used, as in [Luo et al., 2014]. In comparison with the aforementioned Monte Carlo methods, there is no need to generate an additional set of samples and associated evaluations of the computational model per global sensitivity index. Estimating all the global sensitivity indices by using this nonparametric kernel method requires N evaluations of the computational model.

3.3 Spectral method

The principle of spectral methods for the estimation of global sensitivity indices is to approximate the decomposition of f into an expansion of component functions as in (6) with a polynomial chaos expansion. Based on an experimental design of well-chosen values of the uncertain input variables for which the computational model is evaluated,

$$\{(q_{\text{DOE}}^{(j)}, \mathbf{w}_{\text{DOE}}^{(j)}), 1 \leq j \leq N_{\text{DOE}}\}, \quad \text{in which } q_{\text{DOE}}^{(j)} = f(\mathbf{w}_{\text{DOE}}^{(j)}), \quad (39)$$

such spectral methods involve the approximation of f with a polynomial chaos expansion of the form

$$q^p = f^p(\mathbf{w}) = \sum_{|\boldsymbol{\alpha}|=0}^p f_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{w}), \quad (40)$$

where the $\Psi_{\boldsymbol{\alpha}}$ are multivariate polynomials set up as products of univariate polynomials as $\Psi_{\boldsymbol{\alpha}}(\mathbf{w}) = \psi_{\alpha_1}^{(1)}(w_1) \times \dots \times \psi_{\alpha_{n_w}}^{(n_w)}(w_{n_w})$, in which for $k = 1, \dots, n_w$ the univariate polynomials $\psi_0^{(k)} = 1, \psi_1^{(k)}, \psi_2^{(k)}, \dots$ are of increasing degree and P_{W_k} -orthonormal, that is, they are such that $\int_{\mathbb{R}} \psi_{\alpha}^{(k)}(w_k) \psi_{\beta}^{(k)}(w_k) dP_{W_k}(w_k) = \delta_{\alpha\beta}$, with $\delta_{\alpha\beta}$ the Kronecker delta equal to 1 if $\alpha = \beta$ and 0 otherwise. In (40), $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{n_w})$ is a multi-index and $|\boldsymbol{\alpha}| = \alpha_1 + \dots + \alpha_{n_w}$ its modulus. There exist many methods for choosing the values of the uncertain input variables in the experimental design, including fully-tensorized and sparse grids of nodes of quadrature rules, and there exist many methods for fitting the so-called polynomial chaos coefficients $f_{\boldsymbol{\alpha}}$ in (40) to the data set in (39), including regression and collocation methods; please refer, for instance, to [Le Maître and Knio, 2010]. Once the polynomial chaos expansion is available, owing to Ψ_0 being 1 and the $\Psi_{\boldsymbol{\alpha}}$ being products of P_{W_k} -orthonormal basis functions, approximations to the unnormalized global sensitivity indices can be deduced directly from the polynomial chaos coefficients as follows [Crestaux et al., 2009, Sudret, 2008]:

$$D_u^{\text{PCE}} = \sum_{\boldsymbol{\alpha} \in \mathcal{J}_u} f_{\boldsymbol{\alpha}}^2, \quad \text{in which } \mathcal{J}_u = \{\boldsymbol{\alpha} : \boldsymbol{\alpha} \neq \mathbf{0} \text{ and } \alpha_k = 0 \text{ if } k \text{ is not in } u\}. \quad (41)$$

4 Probabilistic learning on manifolds

Especially when the computational model is computationally expensive, it may not be possible (within an available computational budget) to evaluate the computational model sufficiently many times to obtain sufficiently many samples for estimates obtained with Monte Carlo methods as in (30) and those obtained with nonparametric kernel methods as in (38) to be sufficiently accurate.

For such small-data settings, we propose to explore the probabilistic learning on manifolds (PLoM) presented in [Soize and Ghanem, 2016], as we will now describe in more detail.

We assume that a number N_{ir} of independently sampled values $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(N_{\text{ir}})}$ of the uncertain input variables has been generated and the corresponding samples $q^{(1)}, \dots, q^{(N_{\text{ir}})}$ of the uncertain quantity of interest have been obtained by evaluating the computational model, thus resulting in the so-called initial data set of length N_{ir} given by

$$\{\mathbf{x}^{(j)} = (q^{(j)}, \mathbf{w}^{(j)}), 1 \leq j \leq N_{\text{ir}}\} \quad \text{in which } q^{(j)} = f(\mathbf{w}^{(j)}). \quad (42)$$

Here, the subscript “ir” stands for “initial realizations.” We assume that N_{ir} cannot be taken sufficiently large for estimates obtained with Monte Carlo methods as in (30) and those obtained with nonparametric kernel methods as in (38) to be sufficiently accurate. As a means of improving accuracy, we explore the use of probabilistic learning, specifically, the PLoM method. The objective of the probabilistic learning is to deduce from the available samples in the initial data set, but, importantly, without any further evaluations of the computational model, statistically consistent additional samples collected in a so-called learned data set of length N_{ar} , with N_{ar} larger or much larger than N_{ir} :

$$\{\mathbf{x}_{\text{ar}}^{(j)} = (q_{\text{ar}}^{(j)}, \mathbf{w}_{\text{ar}}^{(j)}), 1 \leq j \leq N_{\text{ar}}\}. \quad (43)$$

Here, the subscript “ar” stands for “additional realizations.” We then use these additional samples to estimate the global sensitivity indices. Because these additional samples do not form pairs that share the same values for some uncertain input variables but involve different values for others, these estimates cannot be obtained by using Monte Carlo methods as in (30). Consequently, we use a nonparametric kernel method as in (38):

$$D_u^{\text{PLoM}} = \sum_{i=1}^{N_G} \left(\frac{\frac{1}{N_{\text{ar}}} \sum_{j=1}^{N_{\text{ar}}} q_{\text{ar}}^{(j)} K_{\mathbf{W}_u}(\mathbf{w}_{G,u}^{(i)} - \mathbf{w}_{\text{ar},u}^{(j)})}{\frac{1}{N_{\text{ar}}} \sum_{j=1}^{N_{\text{ar}}} K_{\mathbf{W}_u}(\mathbf{w}_{G,u}^{(i)} - \mathbf{w}_{\text{ar},u}^{(j)})} \right)^2 v_G^{(i)} - \left(\frac{1}{N_{\text{ar}}} \sum_{j=1}^{N_{\text{ar}}} q_{\text{ar}}^{(j)} \right)^2, \quad (44)$$

in which the $\mathbf{w}_{G,u}^{(i)}$ and $v_G^{(i)}$ are the nodes and weights of a quadrature rule such that the integral $\int_{\mathbb{R}^{|u|}} g(\mathbf{w}_u) \frac{1}{N_{\text{ar}}} \sum_{j=1}^{N_{\text{ar}}} K_{\mathbf{W}_u}(\mathbf{w}_u - \mathbf{w}_{\text{ar},u}^{(j)}) d\mathbf{w}_u$ of a function g from $\mathbb{R}^{|u|}$ into \mathbb{R} , with the kernel density estimate of the PDF of the uncertain input variables indexed by the factors in u acting as weight function, is approximated with $\sum_{i=1}^{N_G} g(\mathbf{w}_{G,u}^{(i)}) v_G^{(i)}$.

Whereas estimates obtained with Monte Carlo methods as in (30) and those obtained with nonparametric kernel methods as in (38) have a single number of samples N with respect to which convergence must be assessed, estimates obtained with the PLoM method have two numbers of samples N_{ir} and N_{ar} with respect to which convergence must be assessed. Because we assumed the computational model to be computationally expensive, the total computational cost of the global sensitivity analysis can be expected to be determined mainly by the number of samples N_{ir} in the initial data set. Because the computational model is not used to generate the additional samples in the learned data set, it can be expected that N_{ar} can be readily taken sufficiently large for estimates to be converged with respect to N_{ar} , without accruing a significant contribution to the total computational cost.

For the PLoM method to be of interest, the initial data set must encapsulate information that the PLoM method can extract to generate the additional samples, and the PLoM method must use this information more effectively than Monte Carlo methods and nonparametric kernel methods can. In this paper, we approach this issue from a convergence point of view: we consider that the PLoM method is of interest when the number of samples N_{ir} required for estimates obtained with the PLoM method to be sufficiently converged is lower than the number of samples N required for estimates obtained with Monte Carlo methods and nonparametric kernel methods to be sufficiently converged.

In the remainder of this section, we provide an overview of the PLoM method. We limit ourselves in the remainder of this section to describing and putting into equations the main steps involved in the PLoM method. We provide in A additional details about the PLoM method and its implementation. And we also refer to [Soize and Ghanem, 2016] for additional details, including illustrative examples with samples concentrating near circles in a plane and near a helix in 3D, to Ghanem and Soize [2018], Soize and Ghanem [2020a,b], Soize et al. [2019] for complementary developments, and to Farhat et al. [2019], Ghanem et al. [2018, 2019], Soize and Farhat [2019] for applications and validations.

In the PLoM method, first, a principal component analysis is used to normalize and, if appropriate, construct a linear reduced-dimensional representation of the available samples in the initial data set:

$$\boldsymbol{\eta}_d^{(j)} = [\lambda]^{-1/2}[\varphi]^T(\mathbf{x}^{(j)} - \underline{\mathbf{x}}), \quad j = 1, \dots, N_{\text{ir}}; \quad (45)$$

here, the vector $\underline{\mathbf{x}}$ in \mathbb{R}^n with $n = n_w + 1$ is the sample mean of the available samples, the matrix $[\varphi]$ is the $(n \times \nu)$ -dimensional matrix that collects in its columns the eigenvectors associated with the ν largest magnitude eigenvalues of the sample covariance matrix of the available samples, and the matrix $[\lambda]$ is the $(\nu \times \nu)$ -dimensional diagonal matrix that collects along its diagonal these largest magnitude eigenvalues. The main purpose of the principal component analysis is to normalize the scales and correlations of the physical degrees of freedom (the input variables and the quantity of interest). This normalization of scales and correlations is exploited throughout the remainder of the PLoM method. For instance, it will enable in (46) the use of an isotropic kernel. The linear dimension reduction is not mandatory, that is, ν can be kept equal to the number n of physical degrees of freedom. If $\nu = n$, the representation in (45) involves only a normalization; otherwise, it involves both a normalization and a linear dimension reduction.

The PLoM method interprets the samples $\boldsymbol{\eta}_d^{(1)}, \dots, \boldsymbol{\eta}_d^{(N_{\text{ir}})}$ obtained in (45) as N_{ir} independent samples of a random variable \mathbf{H} with values in \mathbb{R}^ν , whose probability distribution is unknown but assumed to be concentrated in a subset near a manifold in \mathbb{R}^ν . The PLoM method collects these samples in a matrix $[\eta_d] = [\boldsymbol{\eta}_d^{(1)} \dots \boldsymbol{\eta}_d^{(N_{\text{ir}})}]$ and views this matrix as one realization of a random matrix $[\mathbf{H}]$, also with unknown probability distribution. The PLoM method seeks to construct a probabilistic model of this unknown data-generating probability distribution (for a group of N_{ir} samples collected together in a matrix). The additional samples are ultimately obtained by sampling from this probabilistic model. The construction of a generator for groups of N_{ir} samples collected together in a matrix is inherent to the PLoM method. It enables a dimension reduction in the dimension of the problem that indexes the samples. In the PLoM method, this dimension reduction is obtained by using the diffusion-map nonlinear dimension-reduction method [Coifman et al., 2005], and it is fundamental to preserving the concentration near the manifold, as explained next.

Specifically, the PLoM method uses the diffusion-map method to identify the manifold that the samples concentrate near to. This method involves defining an $(N_{\text{ir}} \times N_{\text{ir}})$ -dimensional so-called affinity matrix $[k]$ whose entries

$$[k]_{ij} = \exp\left(-\frac{1}{4\varepsilon_{\text{diff}}} \|\boldsymbol{\eta}_d^{(i)} - \boldsymbol{\eta}_d^{(j)}\|^2\right), \quad i, j = 1, \dots, N_{\text{ir}}, \quad (46)$$

gauge the similarity between the available samples. Upon considering the right eigenvalue problem for the so-called transition matrix $[p] = [b]^{-1} [k]$,

$$[p] \mathbf{g}^{(j)} = \Lambda_j \mathbf{g}^{(j)}, \quad j = 1, \dots, N_{\text{ir}}, \quad (47)$$

in which $[b]$ is the diagonal matrix such that $[b]_{ii} = \sum_{j=1}^{N_{\text{ir}}} [k]_{ij}$ for $i = 1, \dots, N_{\text{ir}}$, for a given integer $\kappa \geq 0$, the diffusion map Φ_κ , which embeds the available samples into a reduced-dimensional space \mathbb{R}^m , is then obtained as

$$\Phi_\kappa : \boldsymbol{\eta}_d^{(j)} \mapsto \Phi_\kappa(\boldsymbol{\eta}_d^{(j)}) = (\Lambda_1^\kappa g_j^{(1)}, \dots, \Lambda_m^\kappa g_j^{(m)}); \quad (48)$$

here, $0 \leq m \leq N_{\text{ir}}$, and, strictly speaking, the embedding results in a reduced-dimensional representation only if m is smaller than N_{ir} . As we have already mentioned, achieving a dimension reduction and thus m being smaller than N_{ir} is fundamental to the PLoM method. The diffusion-map method is a nonlinear dimension-reduction method that does not provide a closed-form out-of-sample extension for the diffusion map Φ_κ : for a supplementary sample in \mathbb{R}^ν , there is no closed-form expression for its image in \mathbb{R}^m , and, for a supplementary sample in \mathbb{R}^m , there is no closed-form expression for its pre-image in \mathbb{R}^ν . In the PLoM method, the eigenvectors $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$ are used in an unusual and original way to construct an alternative, linear, reduced-dimensional representation:

$$[\eta_d]^T = [g][z_d]^T, \text{ in which } [\eta_d] = [\boldsymbol{\eta}_d^{(1)} \dots \boldsymbol{\eta}_d^{(N_{\text{ir}})}], [g] = [\mathbf{g}^{(1)} \dots \mathbf{g}^{(m)}], \text{ and } [z_d] = [\eta_d][g]([g]^T[g])^{-1}. \quad (49)$$

As already mentioned, it should be noted that in this representation, the dimension reduction pertains to the matrix dimension that indexes the samples (N_{ir} is reduced to m) and not the usual physical dimension (ν is kept).

Then, to obtain the probabilistic model of the unknown data-generating probability distribution (for a group of N_{ir} samples collected together in a matrix), the PLoM method begins with inferring from the available samples $\boldsymbol{\eta}_d^{(1)}, \dots, \boldsymbol{\eta}_d^{(N_{\text{ir}})}$ an estimate $p_{\mathbf{H}}^{\text{KDE}}$ of the probability density function of the aforementioned random variable \mathbf{H} by using multi-dimensional kernel density estimation. However, the estimate $p_{\mathbf{H}}^{\text{KDE}}$ thus obtained does not account explicitly for the concentration near the manifold and is thus not used to sample from directly. Instead, the PLoM method proceeds with constructing a joint probability density function $p_{[\mathbf{H}]}^{\text{KDE}}$ by multiplying N_{ir} copies of $p_{\mathbf{H}}^{\text{KDE}}$ such that $[\eta] = [\boldsymbol{\eta}^{(1)} \dots \boldsymbol{\eta}^{(N_{\text{ir}})}] \mapsto p_{[\mathbf{H}]}^{\text{KDE}}([\eta]) = p_{\mathbf{H}}^{\text{KDE}}(\boldsymbol{\eta}^{(1)}) \times \dots \times p_{\mathbf{H}}^{\text{KDE}}(\boldsymbol{\eta}^{(N_{\text{ir}})})$. The probabilistic model is then obtained by improving $p_{[\mathbf{H}]}^{\text{KDE}}$ by a transport of $p_{[\mathbf{H}]}^{\text{KDE}}$ to the subspace spanned by the eigenvectors $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$. That is, the probability distribution $p_{[\mathbf{H}]}^{\text{KDE}}$ on the space of all $(\nu \times N_{\text{ir}})$ -dimensional matrices is improved by a transport to the subspace of only those $(\nu \times N_{\text{ir}})$ -dimensional matrices $[\eta]$ that admit a reduced-dimensional representation $[\eta]^T = [g][z]^T$ with $[z]$ a $(\nu \times m)$ -dimensional matrix. Thus, the concentration near the manifold is preserved by requiring that like the available samples $\boldsymbol{\eta}_d^{(1)}, \dots, \boldsymbol{\eta}_d^{(N_{\text{ir}})}$, the additional samples in the learned data set (to be generated by sampling from the probabilistic model) belong to the subspace spanned by the eigenvectors $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$.

To generate the additional samples, the PLoM method uses the following reduced-dimensional representation:

$$[\mathbf{H}^m]^T = [g][\mathbf{Z}]^T; \quad (50)$$

here, $[\mathbf{H}^m]$ is a $(\nu \times N_{\text{ir}})$ -dimensional random matrix whose probability distribution is the probabilistic model of the unknown data-generating probability distribution (for a group of N_{ir} samples collected together in a matrix) that must be sampled from, and $[\mathbf{Z}]$ is a $(\nu \times m)$ -dimensional random matrix. With this reduced-dimensional representation, the problem of generating samples of $[\mathbf{H}^m]$ is transformed into a problem of generating samples of $[\mathbf{Z}]$. To generate samples of $[\mathbf{Z}]$, the PLoM method uses a Markov chain Monte Carlo (MCMC) method. This MCMC method is obtained by constructing an Itô-stochastic-differential-equation-based (ISDE-based) MCMC method for sampling from the kernel density estimate deduced from the normalized available samples in (45) and then projecting N_{ir} independent copies of this ISDE-based MCMC method onto the subspace spanned by the eigenvectors $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$ (this ISDE-based MCMC method can be viewed as belonging to the class of Hamiltonian Monte Carlo methods). It should be noted that $[z_d]$ is used to define an initial value for the MCMC generator of $[\mathbf{Z}]$ (see A).

Finally, once the MCMC method has provided samples of $[\mathbf{Z}]$, denoted here by $[z^{(1)}], \dots, [z^{(n_{\text{MC}})}]$, with $n_{\text{MC}} = N_{\text{ar}}/N_{\text{ir}}$, the additional samples to form the learned data set are obtained as follows:

$$\mathbf{x}_{\text{ar}}^{(j)} = \underline{\mathbf{x}} + [\varphi][\lambda]^{1/2}[z^{(\ell)}][g_i^{(1)} \dots g_i^{(m)}]^T, \quad j = (\ell-1)N_{\text{ir}}+i, \quad i = 1, \dots, N_{\text{ir}}, \quad \ell = 1, \dots, n_{\text{MC}}. \quad (51)$$

Thus, the PLoM method introduces several parameters to be set. Particularly notable are the parameter $\varepsilon_{\text{diff}}$ of the kernel involved in the definition of the affinity matrix $[k]$ in (46) and the dimension m of the reduced basis $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$ in (49). In [Soize et al., 2019], the parameters $\varepsilon_{\text{diff}}$ and m have been interpreted as parameters that control scales retained in the probabilistic model used to generate the additional samples and in terms of the concentration of this probabilistic model near the identified manifold. In [Soize and Ghanem, 2020b], mathematical results are provided that further elucidate how the dimension reduction achieved by choosing m smaller than N_{ir} is fundamental to preserving the concentration near the manifold. In this paper, we set the values of these parameters based on the scale-separation argument proposed in [Soize et al., 2019] and recalled in A. In addition to $\varepsilon_{\text{diff}}$ and m , the PLoM method introduces bandwidth parameters s_ν and \hat{s}_ν for the kernel density estimate that must be deduced from the normalized available samples in (45), a damping parameter f_0 involved in the ISDE, a pseudotime step Δt involved in the discretization of this ISDE, and parameters k_0 and M_0 involved in the extraction of samples from the sample path of the Markov chain produced by the ISDE-based MCMC method; please see A for more details.

It should be noted that in the PLoM method as it is set up in Soize and Ghanem [2016] with complementary developments in Ghanem and Soize [2018], Soize and Ghanem [2020a], Soize et al. [2019] and recalled in A, it is assumed that there is in the computational model a source of uncertainty \mathbf{U} beyond the source of uncertainty \mathbf{W} . Owing to the presence of \mathbf{U} , the random variable $\mathbf{X} = (Q, \mathbf{W})$ with, now, $Q = f(\mathbf{W}, \mathbf{U})$ can reasonably be expected to admit a PDF, thus lending a justification to deducing a kernel density estimate from the normalized available samples in (45) in obtaining the MCMC method. In our context, such a source of uncertainty \mathbf{U} is not present explicitly. One way of still lending a justification to the use of a kernel density estimate is to consider that the computational model is complex and involves a number of numerical approximations (discretizations, iterative solvers with tolerances, ...) whose impact on the quantity of interest is complex and ultimately manifests itself as a numerical noise that can be construed as a source of uncertainty \mathbf{U} .

Finally, it should be noted that when the PLoM method is used to compute sensitivity indices in global sensitivity analysis, two factors primarily determine the accuracy that is attained. On the one hand, in the PLoM method, a probabilistic model is learned from the available samples and then used to generate additional samples. Any discrepancy that may exist between this probabilistic model and the true, but unknown, data-generating probability distribution can manifest itself as a source of inaccuracy in the computed global sensitivity indices. In this paper, we control this source of inaccuracy by assessing the convergence of the computed sensitivity indices with respect to the number of available samples in the initial data set. On the other hand, we estimate the global sensitivity indices from the additional samples. Here, the use of only a finite number of additional samples can manifest itself as a source of inaccuracy in the computed global sensitivity indices. In this paper, we control this source of inaccuracy by assessing the convergence of the computed global sensitivity indices with respect to the number of additional samples in the learned data set. It can be expected that since the PLoM method is computationally expedient, the number of additional samples in the learned data set can be readily taken as large as desired, thus allowing the potential source of inaccuracy stemming from the use of only a finite number of additional samples to be made as small as desired.

5 Illustration

We will now demonstrate the proposed probabilistic learning method and compare it with the aforementioned other methods in an illustration concerned with forecasting the contribution of the Antarctic ice sheet to sea-level rise.

5.1 Computational ice-sheet model

We used the fast Elementary Thermomechanical Ice Sheet (f.ETISh) model, a computational ice-sheet model that reduces the three-dimensional nature of ice-sheet flow to a two-dimensional problem while keeping the essential characteristics of ice-sheet thermodynamics and ice-stream flow Pattyn [2017]. The f.ETISh model solves a coupled thermo-mechanical system of equations for the ice-flow velocity, the ice thickness, the ice temperature, and the bedrock elevation defined with respect to sea level. For a comprehensive overview of the governing equations, we refer to Pattyn [2017]. As described in Pattyn [2017], the ice-flow velocity and the ice thickness are obtained by solving a hybrid combination of governing equations according to the shallow-ice approximation and governing equations according to the shallow-shelf approximation. The governing equations according to the shallow-ice approximation are solved only where the ice is grounded, that is, where the ice rests on the bedrock, and the governing equations according to the shallow-shelf approximation are solved both where the ice is grounded and where it is floating, that is, in the so-called ice shelves. The governing equations according to the shallow-shelf approximation take a different form in the grounded and floating portions, as a consequence of friction with the bedrock only being present in the grounded portion and differences in the mechanical behavior that dominates the ice-flow velocity and stress fields between the grounded and floating portions. The temperature is obtained by solving a diffusion-advection equation. The bedrock elevation is obtained by solving a bedrock relaxation model. Input data for the f.ETISh model consist of input data deduced from present-day observations, such as present-day ice-sheet geometry and topography, a basal-sliding coefficient inferred with a data assimilation method, present-day atmospheric temperature and precipitation, and present-day ocean temperature. The discretization of space is a finite-difference discretization on staggered grids and the discretization of time is implicit.

5.2 Test problem

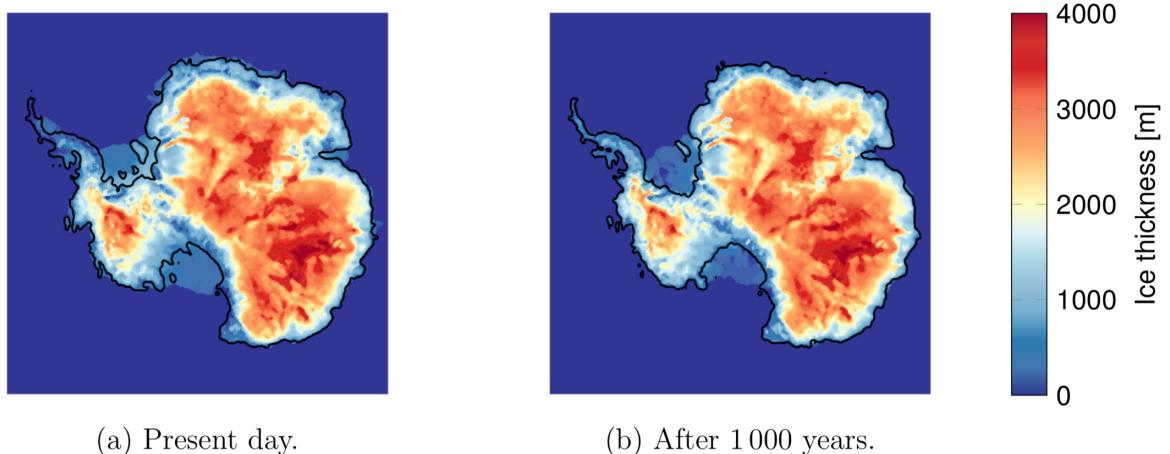


Figure 1: Ice thickness [m] (a) at present day and (b) after 1 000 years obtained with nominal values of input variables.

We set up a test problem using the f.ETISh model in a manner that is similar to the problem set up in [Bulthuis et al., 2019]. We simulated the response of the Antarctic ice sheet over the next 1 000 years, starting from its present-day configuration, under the atmospheric forcing scenario RCP 4.5 defined for the next hundred years by the IPCC and extended in time as reported in [Golledge et al., 2015] (Fig. 1). In the test problem as we set it up using the f.ETISh model, an increase

in atmospheric temperature has an impact on the surface mass balance through an increase in precipitation and surface melting, and, in addition, an increase in atmospheric temperature induces an increase in the ocean temperature, which in turn has an impact on the mass balance underneath the ice shelves through an increase in the strength of the overturning ocean circulation in ice-shelf cavities. We used a square spatial grid with a length of 5 600 km in each direction and a spatial resolution of 20 km, and we used a time step of 0.05 year. The computing time of a single model evaluation on two threads of a SkyLake 2.3 GHz CPU was approximately 8 hours.

5.3 Sources of uncertainty

Uncertain input variables	Symbol	Nominal	Min	Max	Units
Calving multiplier factor	w_1	1	0	2.0	
Ocean melt factor	w_2	0.3	0.1	0.5	
Shelf anisotropy factor	w_3	0.5	0.2	1	
East Antarctic relaxation time	w_4	3 000	1 000	5 000	yr
West Antarctic relaxation time	w_5	3 000	1 000	5 000	yr
Quantity of interest	Symbol	Nominal	Min	Max	Units
Contribution to global mean sea level	q				m

Table 1: List of uncertain input variables and quantity of interest.

We considered five input variables to be uncertain (Table 1), in a manner that is similar to the problem set up in [Bulthuis et al., 2019], where the reader may also find additional information about the definition and physical interpretation of these input variables. The first uncertain input variable, the calving multiplier factor, controls the magnitude of the calving rate at the edges of ice shelves. The second uncertain input variable, the ocean melt factor, controls the magnitude of the change in the ocean temperature on the continental shelf as a function of the change in the background atmospheric temperature under the atmospheric forcing scenario. The third uncertain input variable, the shelf anisotropy factor, controls the magnitude of the change in ice viscosity as shear-dominated stress behavior of grounded ice transitions to extension-dominated stress behavior of floating ice. The fourth and fifth uncertain input variables, the East and West Antarctic relaxation times, control the characteristic time scales of the bedrock uplift due to deglaciation in East and West Antarctica. We represented these uncertain input variables as independent uniform random variables with ranges of values determined by the minimum and maximum values in Table 1.

5.4 Predicted quantity of interest

As the predicted quantity of interest, we considered the contribution to global mean sea level stemming from the evolution of the Antarctic ice sheet after 1 000 years. Under the aforementioned probabilistic description of the uncertain input variables, we found using the Monte Carlo method with 3 000 samples for this quantity of interest a mean value of 0.46 m, a standard deviation of 0.49 m, and thus a variance of 0.24 m² and a coefficient of variation of 108%.

5.5 Global sensitivity analysis: Monte Carlo method

We used the Monte Carlo method as in (30) as described in Sec. 3.1 to estimate the main-effect global sensitivity indices for the first, third, and fifth uncertain input variables. We used for N values up to 3 000, so that the Monte Carlo estimation of these three main-effect global sensitivity

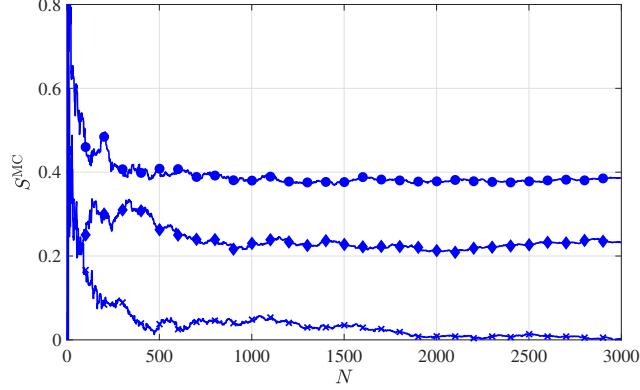


Figure 2: Estimates of the main-effect global sensitivity indices for the first (circle), third (diamond), and fifth (x-mark) uncertain input variables obtained by using the Monte Carlo method as a function of N .

indices required $3000 \times (3 + 1) = 12\,000$ model evaluations. Figure 2 shows the convergence of the estimates as a function of N . We can observe that the uncertainties introduced in the calving factor and the shelf anisotropy factor are more significant in inducing uncertainty in the contribution to global mean sea level than the uncertainty introduced in the West Antarctic relaxation time. We can observe that reasonably converged estimates of the main-effect global sensitivity indices for the first and third uncertain input variables are obtained after a value for N of about 1000 and that reasonably converged estimates of the main-effect global sensitivity index for the fifth uncertain input variable are obtained after a value for N of about 2000. The results obtained with $N = 3\,000$ are $S_1^{\text{MC}} = 0.3869$, $S_3^{\text{MC}} = 0.2322$, and $S_5^{\text{MC}} = 0.0026$. Please note that we did not compute the main-effect global sensitivity indices for the second and fourth uncertain input variables because we preferred to avoid the high computational cost of the additionally required $3\,000 \times 2 = 6\,000$ model evaluations.

5.6 Global sensitivity analysis: Spectral method

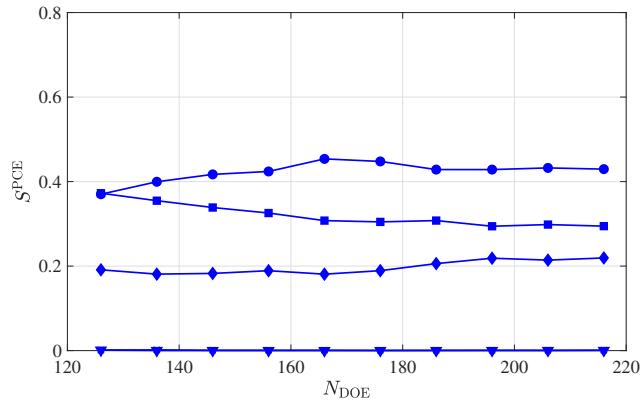


Figure 3: Estimates of the main-effect global sensitivity indices for the first (circle), second (square), third (diamond), fourth (triangle), and fifth (x-mark) uncertain input variables obtained by using the spectral method as a function of N_{DOE} .

We used the spectral method as in (41) as described in Sec. 3.2 to estimate the main-effect global sensitivity indices for the uncertain input variables. Especially to facilitate a comparison of the performance of the spectral method with the performance of the probabilistic learning method to be presented next, we used for N_{DOE} values up to only 224. We used independent samples to form the

experimental design of values of the uncertain input variables for which the computational model is evaluated as in (39). We used multivariate polynomials set up as products of univariate normalized Legendre polynomials up to a total degree of 4; a polynomial chaos expansion of dimension $n_w = 5$ and a total degree of $p = 4$ as in (40) involves $(4+5)!/4!/5! = 126$ terms. We used a regression method to determine the polynomial chaos coefficients. Figure 3 shows the convergence of the estimates as a function of N_{DOE} . We can observe that especially for the main-effect global sensitivity index for the first uncertain input variable, the estimate obtained with the spectral method after $N_{\text{DOE}} = 224$ evaluations of the computational model in Fig. 3 is not yet converged to the reasonably converged value obtained with the Monte Carlo method after $N = 3000$ evaluations of the computational model in Fig. 2. The results obtained with $N_{\text{DOE}} = 224$ and $p = 4$ are $S_1^{\text{PCE}} = 0.4294$, $S_2^{\text{PCE}} = 0.2945$, $S_3^{\text{PCE}} = 0.2192$, $S_4^{\text{PCE}} = 0.0003$, and $S_5^{\text{PCE}} = 0.0009$.

5.7 Global sensitivity analysis: PLoM method

We used the PLoM method as in Sec. 4 to estimate the main-effect global sensitivity indices for the uncertain input variables. The maximum value considered for the length of the initial data set was 224.

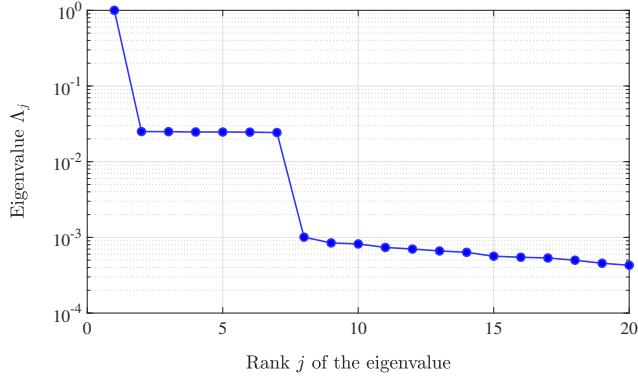


Figure 4: Eigenvalues of $[p]$ for $N_{\text{ir}} = 224$ and $\varepsilon_{\text{diff}} = 20$.

For $N_{\text{ir}} = 224$, we set the values of the parameters of the PLoM method as follows:

- We set $\nu = n = 6$, so that the representation of the available samples as in (45) (and (52) in A) involves only a normalization and not a dimension reduction.
- We set the parameter $\varepsilon_{\text{diff}}$ of the kernel involved in the definition of the affinity matrix $[k]$ in (46) (and (54) in A) and the dimension m of the reduced basis $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$ in (49) (and (59) in A) following the arguments in (57) with (56) in A, leading to $\varepsilon_{\text{diff}}^{\text{opt}} = 20$ and $m^{\text{opt}} = 8$. The constant eigenvector $\mathbf{g}^{(1)}$ was removed for the reasons detailed in A. For $\varepsilon_{\text{diff}} = 20$, Fig. 4 shows the eigenvalues of $[p]$.
- We set the bandwidth parameters s_ν and \hat{s}_ν for the kernel density estimate that must be deduced from the normalized available samples in (45) following (65) in A, leading to $s_\nu = 0.543$ and $\hat{s}_\nu = 0.478$. We set the damping parameter f_0 in the ISDE (equation (61) in A) equal to $f_0 = 1.5$.
- We set the pseudotime step Δt involved in the discretization of the ISDE (equations (67)–(69) in A) equal to $\Delta t = 0.150192$. We set the parameters k_0 , M_0 , and n_{MC} involved in the extraction of samples from the trajectory of the Markov chain produced by the ISDE-based MCMC method as in (51) (and Sec. A.6 in A) equal to $k_0 = 200$, $M_0 = 100$, and $n_{\text{MC}} = 400$, thus leading to $N_{\text{ar}} = 89\,600 = 400 \times 224$.

- In estimating the global sensitivity indices as in (44), we used an $N_G = 200$ -node trapezoidal integration rule.

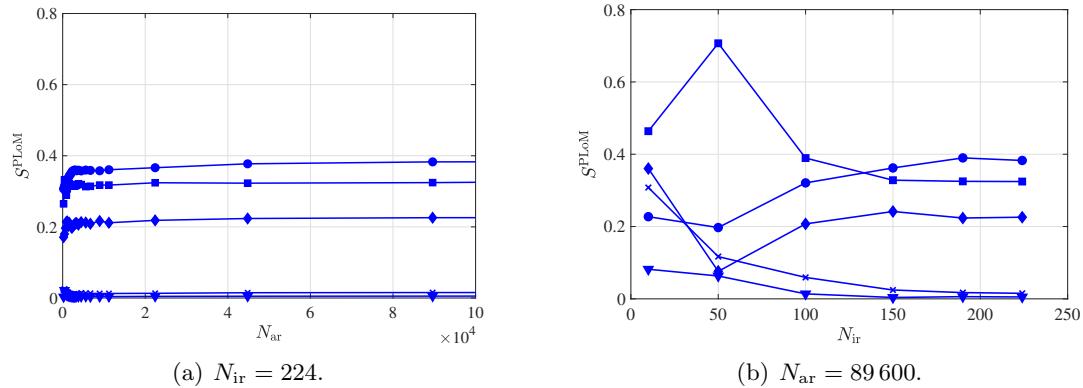


Figure 5: Estimates of the main-effect global sensitivity indices for the first (circle), second (square), third (diamond), fourth (triangle), and fifth (x-mark) uncertain input variables obtained by using the PLoM method (a) as a function of N_{ar} for $N_{\text{ir}} = 224$ and (b) as a function of N_{ir} for $N_{\text{ar}} = 89\,600$. For the values of N_{ir} different from 224, the values of n_{MC} are different from 400.

Figure 5(a) shows the convergence of the estimates as a function of N_{ar} for $N_{\text{ir}} = 224$. The PLoM method was computationally expedient, and we could readily take N_{ar} sufficiently large for estimates to be converged with respect to N_{ar} without accruing a significant contribution to the total computational cost. Figure 5(b) shows the convergence of the estimates as a function of N_{ir} for $N_{\text{ar}} = 89\,600$. We can observe that the estimates obtained with the PLoM method are well converged with respect to the length N_{ir} of the initial data set even for small initial data sets with lengths of only $N_{\text{ir}} = 190$ or only $N_{\text{ir}} = 224$. We can also observe that the estimates obtained with the PLoM method even for small initial data sets with lengths of only $N_{\text{ir}} = 190$ or only $N_{\text{ir}} = 224$ in Fig. 5(b) correspond well with the converged values obtained with the Monte Carlo method in Fig. 2. Notably, we can observe that for the main-effect global sensitivity index for the first uncertain input variable, the estimate obtained with the PLoM method even for the small initial data set with a length of only $N_{\text{ir}} = 224$ in 5(b) corresponds better with the reasonably converged value obtained with the Monte Carlo method with $N = 3\,000$ in Fig. 2 than the estimate obtained with the spectral method with $N_{\text{DOE}} = 224$ in Fig. 3. We carried out a final computation with the PLoM method with $N_{\text{ar}} = 1\,000\,000$ for $N_{\text{ir}} = 224$. The results obtained with $N_{\text{ir}} = 224$ and $N_{\text{ar}} = 1\,000\,000$ are $S_1^{\text{PLoM}} = 0.3963$, $S_2^{\text{PLoM}} = 0.3353$, $S_3^{\text{PLoM}} = 0.2361$, $S_4^{\text{PLoM}} = 0.0066$, and $S_5^{\text{PLoM}} = 0.0182$.

5.8 Global sensitivity analysis: Comparison of the numerical methods

Uncertain input variables	D_k^{MC} [m ²]	S_k^{MC}	D_k^{PCE} [m ²]	S_k^{PCE}	D_k^{PLOM} [m ²]	S_k^{PLOM}
Calving multiplier factor	0.0956 ± 0.0100	0.3869	0.1040	0.4294	0.0960	0.3963
Ocean melt factor			0.0714	0.2945	0.0812	0.3353
Shelf anisotropy factor	0.0578 ± 0.0093	0.2322	0.0531	0.2192	0.0572	0.2361
East Antarctic relaxation time			7.2687E-5	0.0003	0.0016	0.0066
West Antarctic relaxation time	6.6151E-4 ± 0.0094	0.0026	2.1806E-4	0.0009	0.0044	0.0182

Table 2: Unnormalized and normalized global sensitivity indices obtained with the Monte Carlo method with $N = 3\,000$, the spectral method with $N_{\text{DOE}} = 224$ and $p = 4$, and the PLoM method with $N_{\text{ir}} = 224$ and $N_{\text{ar}} = 1\,000\,000$.

Table 2 lists the unnormalized and normalized global sensitivity indices obtained with the Monte Carlo, spectral, and PLoM methods. In addition, Table 2 provides 95%-confidence intervals for the Monte Carlo estimates of the unnormalized global sensitivity indices obtained by using the central limit theorem for the Monte Carlo estimator. These 95%-confidence intervals quantify the error that may be present in the Monte Carlo estimates of the unnormalized global sensitivity indices owing to their estimation from only the finite number of $N = 3\,000$ samples. We can observe that the estimates of the unnormalized global sensitivity indices obtained with the spectral and PLoM methods fall within the 95%-confidence intervals for the corresponding estimates obtained with the Monte Carlo method. We can also observe that for the first and third uncertain input variables, the estimates obtained with the PLoM method correspond better with the estimates obtained with the Monte Carlo method than the estimates obtained with the spectral method. This is not the case for the fifth uncertain input variable, but the global sensitivity index for the fifth uncertain input variable is very small, and the 95%-confidence interval for its Monte Carlo estimate indicates that this estimate may be subject to a large relative error due to its estimation from only $N = 3\,000$ samples.

Since a single evaluation of the computational model on two threads of a SkyLake 2.3 GHz CPU of the employed cluster takes about 8 hours, the aggregate computing time for the $3\,000 \times (3 + 1) = 12\,000$ evaluations of the computational model to obtain the Monte Carlo estimates was 96 000 hours. By contrast, the aggregate computing time for the 224 evaluations of the computational model to obtain the spectral and PLoM estimates was only 1 792 hours.

6 Conclusion

We have proposed to use the PLoM method for the computation of global sensitivity indices in small-data settings in which the computational model can be evaluated only a small number of times. The probabilistic learning on manifolds serves to learn from the available samples a probabilistic model that is used to generate statistically consistent additional samples, on the basis of which estimates of the global sensitivity indices are then computed. We provided an illustration in which a usual Monte Carlo method required a data set of several thousands of samples to attain a reasonable convergence of the estimates, but the proposed PLoM method was able to attain converged estimates based on only a small data set of only a few hundred samples. In this illustration, these results indicated that the small data set encapsulated information that the PLoM method was able to exploit to accelerate the convergence of the estimates of the global sensitivity indices with respect to the number of samples.

Acknowledgements

Maarten Arnst and Christian Soize acknowledge the Fund for Scientific Research (F.R.S.-FNRS) for its financial support for research stays, and Kevin Bulthuis acknowledges the F.R.S.-FNRS for its financial support for his F.R.S.-FNRS research fellowship. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the F.R.S.-FNRS under Grant No. 2.5020.11 and by the Walloon Region.

A Probabilistic learning on manifolds (PLoM)

In this Appendix, we summarize in greater detail the probabilistic learning on manifolds (PLoM) method. The PLoM method was introduced in Soize and Ghanem [2016] and complementary developments can be found in Ghanem and Soize [2018], Soize and Ghanem [2020a,b], Soize et al. [2019].

Applications and validations can be found in Farhat et al. [2019], Ghanem et al. [2018, 2019], Soize and Farhat [2019]. In addition, we give formulas for setting the values of parameters $\varepsilon_{\text{diff}}$, m , s_ν , and \hat{s}_ν .

In this Appendix, all random variables are defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$. An example of the use of the PLoM method is as follows. We consider a computationally expensive large-scale stochastic computational model. The vector-valued random response is a non-Gaussian \mathbb{R}^{n_q} -valued random variable $\mathbf{Q} = (Q_1, \dots, Q_{n_q})$ written as $\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbf{U})$ in which $\mathbf{W} = (W_1, \dots, W_{n_w})$ is a non-Gaussian \mathbb{R}^{n_w} -valued random parameter controlling the system, \mathbf{U} is a non-Gaussian \mathbb{R}^{n_u} -valued random variable not used for controlling the system, such as a random variable related to uncertainties, and \mathbf{f} is a deterministic measurable nonlinear function with values in \mathbb{R}^{n_q} representing the computational model. The available information is a given *initial data set* $D_{N_{\text{ir}}}$ (training set) of length N_{ir} that is constructed as $\{\mathbf{x}^{(j)} = (\mathbf{q}^{(j)}, \mathbf{w}^{(j)}), j = 1, \dots, N_{\text{ir}}\}$, in which the $\mathbf{q}^{(j)} = \mathbf{f}(\mathbf{w}^{(j)}, \mathbf{u}^{(j)})$ are N_{ir} independent realizations of \mathbf{Q} calculated with the computational model, with the $\mathbf{w}^{(j)}$ and $\mathbf{u}^{(j)}$ being N_{ir} independent realizations of \mathbf{W} and \mathbf{U} . Hence, $\{\mathbf{x}^{(j)}, j = 1, \dots, N_{\text{ir}}\}$ collects N_{ir} independent realizations in $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ with $n = n_q + n_w$ of the non-Gaussian random vector $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ with values in $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$. Knowing only this initial data set, the objective is, for instance, to construct, for a given \mathbf{w} , an estimate $h^{N_{\text{ir}}}(\mathbf{w})$ of $h(\mathbf{w}) = E\{H(\mathbf{Q})|\mathbf{W} = \mathbf{w}\}$, the conditional expectation of $H(\mathbf{Q})$ given $\mathbf{W} = \mathbf{w}$, with H a given deterministic function. For instance, beyond the scope of the computation of global sensitivity indices as considered in this paper, h could be the objective function of an optimization problem for which \mathbf{w} could be a design parameter. If each evaluation of a $\mathbf{q}^{(j)}$ is computationally expensive, then taking N_{ir} sufficiently large for $h^{N_{\text{ir}}}(\mathbf{w})$ to be sufficiently converged to $h(\mathbf{w})$ may not be computationally tractable. The objective of the PLoM method is then to generate $N_{\text{ar}} \gg N_{\text{ir}}$ additional realizations $\{\mathbf{x}_{\text{ar}}^{(j)}, j = 1, \dots, N_{\text{ar}}\}$ of \mathbf{X} . As soon as the set $\{\mathbf{x}_{\text{ar}}^{(j)}, j = 1, \dots, N_{\text{ar}}\}$ is obtained, the additional realizations of \mathbf{Q} and \mathbf{W} may be extracted as $\mathbf{x}_{\text{ar}}^{(j)} = (\mathbf{q}_{\text{ar}}^{(j)}, \mathbf{w}_{\text{ar}}^{(j)})$ for $j = 1, \dots, N_{\text{ar}}$ to obtain the *learned data set* $D_{N_{\text{ar}}}$.

A.1. Normalization of the initial dataset by using principal component analysis. Let $[\mathbf{X}] = [\mathbf{X}^{(1)} \dots \mathbf{X}^{(N_{\text{ir}})}]$ be the random matrix with values in $\mathbb{M}_{n, N_{\text{ir}}}$ whose columns are N_{ir} independent copies of \mathbf{X} . Therefore, the N_{ir} independent realizations $\mathbf{x}^{(j)}$ of \mathbf{X} with values in \mathbb{R}^n may be collected in the columns of a matrix $[x_d] = [\mathbf{x}^{(1)} \dots \mathbf{x}^{(N_{\text{ir}})}]$ in $\mathbb{M}_{n, N_{\text{ir}}}$, which may be viewed as one realization of $[\mathbf{X}]$. A normalization $[\mathbf{H}]$ of $[\mathbf{X}]$ is obtained via a mean-square approximation $[\mathbf{X}^\nu]$ to $[\mathbf{X}]$ by using a classical principal component analysis:

$$[\mathbf{X}^\nu] = [\underline{x}] + [\varphi][\lambda]^{1/2}[\mathbf{H}]. \quad (52)$$

In (52), the random matrix $[\mathbf{H}] = [\mathbf{H}^{(1)} \dots \mathbf{H}^{(N_{\text{ir}})}]$ takes values in $\mathbb{M}_{\nu, N_{\text{ir}}}$ with $\nu \leq n$ and its columns are N_{ir} independent copies of a random vector denoted by \mathbf{H} . The diagonal of the $(\nu \times \nu)$ -dimensional diagonal matrix $[\lambda]$ is composed of ν positive eigenvalues of the empirical estimate of the covariance matrix of \mathbf{X} (computed by using $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_{\text{ir}})}$) and $[\varphi]$ is the $(n \times \nu)$ -dimensional matrix that collects in its columns the associated eigenvectors that satisfy the normalization condition $[\varphi]^T[\varphi] = [I_\nu]$. The matrix $[\underline{x}]$ belongs to $\mathbb{M}_{n, N_{\text{ir}}}$ and has identical columns, each one being equal to the empirical estimate $\underline{\mathbf{x}}$ in \mathbb{R}^n of the mean value of \mathbf{X} (computed by using $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_{\text{ir}})}$). The realization $[\eta_d] = [\eta_d^{(1)} \dots \eta_d^{(N_{\text{ir}})}]$ in $\mathbb{M}_{\nu, N_{\text{ir}}}$ of $[\mathbf{H}]$ (associated with the realization $[x_d]$ of $[\mathbf{X}]$) is computed as

$$[\eta_d] = [\lambda]^{-1/2}[\varphi]^T([x_d] - [\underline{x}]). \quad (53)$$

If n is small, ν may be chosen equal to the rank n_{rank} of the aforementioned empirical estimate of the covariance matrix of \mathbf{X} . If n is high, a statistical reduction may be used as usual and therefore $\nu < n_{\text{rank}}$. Then, $[\mathbf{X}^\nu]$ is a mean-square approximation to $[\mathbf{X}]$ for a given tolerance of the mean-square error.

A.2. Diffusion-maps basis. An algebraic basis of $\mathbb{R}^{N_{\text{ir}}}$ is constructed by using diffusion maps as proposed in Coifman et al. [2005]. We introduce the matrix $[k]$, assumed to be in $\mathbb{M}_{N_{\text{ir}}}^+$, with entries given for $i = 1, \dots, N_{\text{ir}}$ and $j = 1, \dots, N_{\text{ir}}$ by

$$[k]_{ij} = \exp \left(-\frac{1}{4\varepsilon_{\text{diff}}} \left\| \boldsymbol{\eta}_{\text{d}}^{(i)} - \boldsymbol{\eta}_{\text{d}}^{(j)} \right\|^2 \right), \quad (54)$$

in which the kernel depends on a real smoothing parameter $\varepsilon_{\text{diff}} > 0$. Let $[b]$ be the positive-definite diagonal real matrix in $\mathbb{M}_{N_{\text{ir}}}^+$ such that $[b]_{ii} = \sum_{j=1}^{N_{\text{ir}}} [k]_{ij}$ for $i = 1, \dots, N_{\text{ir}}$. Let $[p] = [b]^{-1}[k]$, so that $[p]$ may be viewed as a transition matrix in $\mathbb{M}_{N_{\text{ir}}}$ of a Markov chain. For given m with $1 < m \leq N_{\text{ir}}$, let $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$ be the right eigenvectors in $\mathbb{R}^{N_{\text{ir}}}$ of $[p]$ such that $[p]\mathbf{g}^{(j)} = \Lambda_j \mathbf{g}^{(j)}$, with the real eigenvalues Λ_j ordered such that $\Lambda_1 = 1 > \Lambda_2 > \dots > \Lambda_m$. The vectors $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$ are referred to as the *diffusion-maps basis*, and they are collected in a matrix $[g] = [\mathbf{g}^{(1)} \dots \mathbf{g}^{(m)}]$ in $\mathbb{M}_{N_{\text{ir}},m}$. The normalization condition of these eigenvectors is set such that $[g]^T [b] [g] = [I_m]$. The eigenvector $\mathbf{g}^{(1)}$ associated with the largest eigenvalue $\Lambda_1 = 1$ is a constant vector. For $m = N_{\text{ir}}$, the diffusion-maps basis is an algebraic basis of $\mathbb{R}^{N_{\text{ir}}}$. It can be seen that $[g]$ is the solution to the following generalized eigenvalue problem for positive-definite matrices with the normalization condition $[g]^T [b] [g] = [I_m]$:

$$[k] [g] = [b] [g] [\Lambda]. \quad (55)$$

This construction introduces two parameters: the dimension $m \leq N_{\text{ir}}$ and the smoothing parameter $\varepsilon_{\text{diff}} > 0$. The value of m is *a priori* dependent on the value of $\varepsilon_{\text{diff}}$. A method has been proposed in Soize et al. [2019] for setting their values as follows. Let $\varepsilon_{\text{diff}} \mapsto \hat{m}(\varepsilon_{\text{diff}})$ be the function from $\mathbb{R}_0^+ =]0, +\infty[$ into \mathbb{N} such that

$$\hat{m}(\varepsilon_{\text{diff}}) = \arg \min_{j|j \geq 3} \left\{ \frac{\Lambda_j(\varepsilon_{\text{diff}})}{\Lambda_2(\varepsilon_{\text{diff}})} < 0.1 \right\}. \quad (56)$$

If \hat{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 $\hat{m}(\varepsilon_{\text{diff}}^{\text{opt}})$ such that

$$\{\hat{m}(\varepsilon_{\text{diff}}^{\text{opt}}) < \hat{m}(\varepsilon_{\text{diff}}), \forall \varepsilon_{\text{diff}} \in]0, \varepsilon_{\text{diff}}^{\text{opt}}[\} \cap \{\hat{m}(\varepsilon_{\text{diff}}^{\text{opt}}) = \hat{m}(\varepsilon_{\text{diff}}), \forall \varepsilon_{\text{diff}} \in]\varepsilon_{\text{diff}}^{\text{opt}}, 1.5\varepsilon_{\text{diff}}^{\text{opt}}[\}.$$
 \quad (57)

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

A.3. Reduced-dimensional representation $[\mathbf{H}^m]$ of $[\mathbf{H}]$. The vectors $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$ span a subspace of $\mathbb{R}^{N_{\text{ir}}}$ that characterizes, for the optimal values m^{opt} and $\varepsilon_{\text{diff}}^{\text{opt}}$ of m and $\varepsilon_{\text{diff}}$, the local geometry of the data set $\{\boldsymbol{\eta}_{\text{d}}^{(j)}, j = 1, \dots, N_{\text{ir}}\}$. A reduced-dimensional representation $[\mathbf{H}^m]$ of $[\mathbf{H}]$ is obtained by representing each column of the $\mathbb{M}_{N_{\text{ir}},\nu}$ -valued random matrix $[\mathbf{H}]^T$ in this subspace of $\mathbb{R}^{N_{\text{ir}}}$ spanned by $\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(m)}$. Consequently, introducing the random matrix $[\mathbf{Z}]$ with values in $\mathbb{M}_{\nu,m}$, we write

$$[\mathbf{H}^m] = [\mathbf{Z}][g]^T. \quad (58)$$

Asuming that for all $i = 1, \dots, \nu$, the vector $(\eta_{\text{di}}^{(1)}, \dots, \eta_{\text{di}}^{(N_{\text{ir}})})$ is not a constant vector, the basis vector $\mathbf{g}^{(1)}$ associated with $\Lambda_1 = 1$, which is a constant vector, could be removed from the representation defined by (58). In order to calculate a realization $[z_{\text{d}}]$ in $\mathbb{M}_{\nu,m}$ of $[\mathbf{Z}]$ associated with $[\eta_{\text{d}}]$, which will be used as an initial value of the MCMC generator of $[\mathbf{Z}]$, and since the matrix $[g]^T [g]$ in \mathbb{M}_m is invertible, we introduce the following least-squares approximation $[z_{\text{d}}]$ of $[\eta_{\text{d}}]$:

$$[z_{\text{d}}] = [\eta_{\text{d}}][a], \quad (59)$$

in which

$$[a] = [g]([g]^T [g])^{-1}. \quad (60)$$

A.4. Generation of the additional realizations $\{\boldsymbol{\eta}_{\text{ar}}^{(j)}, j = 1, \dots, N_{\text{ar}}\}$ of $[\mathbf{H}^m]$. The generation of additional realizations $[z_{\text{ar}}^1], \dots, [z_{\text{ar}}^{n_{\text{MC}}}]$ of $[\mathbf{Z}]$ is carried out by using an unusual MCMC method introduced in Soize and Ghanem [2016]. This MCMC method is based on a reduced-order Itô stochastic differential equation (ISDE) constructed by projecting onto the diffusion-maps basis an ISDE related to a dissipative Hamiltonian dynamical system for which the invariant measure is defined by the PDF of $[\mathbf{H}]$ (obtained using an unusual Gaussian kernel-density estimation method [Soize, 2015] from $[\eta_{\text{d}}]$) Soize [2008]. The existence and uniqueness of this invariant measure can be obtained as in Soize [1988, 1994]. This MCMC method preserves the concentration of the probability measure (if such a concentration exists) and avoids a scatter phenomenon. This reduced-order ISDE, for which the stochastic process $\{([\mathbf{Z}(t)], [\mathbf{Y}(t)]), t \in \mathbb{R}^+\}$ with values in $\mathbb{M}_{\nu,m} \times \mathbb{M}_{\nu,m}$ is the solution, is written, for $t > 0$, as

$$\begin{cases} d[\mathbf{Z}(t)] = [\mathbf{Y}(t)]dt, \\ d[\mathbf{Y}(t)] = [\mathcal{L}([\mathbf{Z}(t)])]dt - \frac{1}{2}f_0[\mathbf{Y}(t)]dt + \sqrt{f_0}d[\mathbf{V}(t)], \end{cases} \quad (61)$$

and it is completed with the initial condition $[\mathbf{Z}(0)] = [z_{\text{d}}]$ and $[\mathbf{Y}(0)] = [\mathcal{N}][a]$ almost surely.

(i) The random matrix $[\mathcal{L}([\mathbf{Z}(t)])]$ with values in $\mathbb{M}_{\nu,m}$ is such that $[\mathcal{L}([\mathbf{Z}(t)])] = [L([\mathbf{Z}(t)][g]^T)][a]$. For $[u] = [\mathbf{u}^{(1)} \dots \mathbf{u}^{(N_{\text{ir}})}]$ in $\mathbb{M}_{\nu,N_{\text{ir}}}$ with the $\mathbf{u}^{(j)} = (u_1^{(j)}, \dots, u_{\nu}^{(j)})$ in \mathbb{R}^{ν} , the entries of the matrix $[L([u])]$ in $\mathbb{M}_{\nu,N_{\text{ir}}}$ for $i = 1, \dots, \nu$ and $j = 1, \dots, N_{\text{ir}}$ are given by

$$[L([u])]_{ij} = \frac{1}{p(\mathbf{u}^{(j)})} \{\nabla_{\mathbf{u}} p(\mathbf{u}^{(j)})\}_i, \quad (62)$$

$$p(\mathbf{u}) = \frac{1}{N_{\text{ir}}} \sum_{j=1}^{N_{\text{ir}}} \exp \left(-\frac{1}{2\hat{s}_{\nu}^2} \left\| \frac{\hat{s}_{\nu}}{s_{\nu}} \boldsymbol{\eta}_{\text{d}}^{(j)} - \mathbf{u} \right\|^2 \right), \quad (63)$$

$$\nabla_{\mathbf{u}} p(\mathbf{u}) = \frac{1}{\hat{s}_{\nu}^2} \frac{1}{N_{\text{ir}}} \sum_{j=1}^{N_{\text{ir}}} \left(\frac{\hat{s}_{\nu}}{s_{\nu}} \boldsymbol{\eta}_{\text{d}}^{(j)} - \mathbf{u} \right) \exp \left(-\frac{1}{2\hat{s}_{\nu}^2} \left\| \frac{\hat{s}_{\nu}}{s_{\nu}} \boldsymbol{\eta}_{\text{d}}^{(j)} - \mathbf{u} \right\|^2 \right), \quad (64)$$

in which s_{ν} is the bandwidth chosen in accordance with Silverman's rule of thumb (more specifically, its extension to multivariate product kernels) and \hat{s}_{ν} its modification introduced in Soize [2015] (this modification is motivated by the normalization of the initial data set by using principal components described above; this modification results in the covariance matrix of the kernel density estimate being the identity matrix):

$$s_{\nu} = \left(\frac{4}{N_{\text{ir}}(2 + \nu)} \right)^{1/(\nu+4)}, \quad \hat{s}_{\nu} = \frac{s_{\nu}}{\sqrt{s_{\nu}^2 + \frac{N_{\text{ir}}-1}{N_{\text{ir}}}}}. \quad (65)$$

(ii) $[\mathbf{V}(t)] = [\mathcal{W}^{(1)}(t) \dots \mathcal{W}^{(N_{\text{ir}})}(t)][a]$ where the $\{[\mathcal{W}^{(j)}(t)], t \in \mathbb{R}^+\}$ are N_{ir} independent \mathbb{R}^{ν} -valued normalized Wiener stochastic processes.

(iii) $[\mathcal{N}]$ is an $\mathbb{M}_{\nu,N_{\text{ir}}}$ -valued normalized Gaussian random matrix that is independent of $[\mathbf{V}]$.

(iv) The free parameter f_0 , whose value must be set such that $0 < f_0 \leq 4$, allows the dissipation term of the nonlinear second-order dynamical system (dissipative Hamiltonian system) to be controlled in order to suppress the transient induced by the initial condition (because the initial condition does not correspond to a realization of the invariant measure). A common value is $f_0 = 1.5$, and the upper bound of $f_0 = 4$ would correspond to the critical damping value if the dynamical system

was linear (which is not the case).

(v) We then have $[\mathbf{Z}] = \lim_{t \rightarrow +\infty} [\mathcal{Z}(t)]$ in probability distribution, which allows us to generate the additional realizations $[z_{\text{ar}}^{(1)}], \dots, [z_{\text{ar}}^{(n_{\text{MC}})}]$ and then the additional realizations $[\eta_{\text{ar}}^{(1)}], \dots, [\eta_{\text{ar}}^{(n_{\text{MC}})}]$ by using (58) such that $[\eta_{\text{ar}}^{(j)}] = [z_{\text{ar}}^{(j)}][g]^T$ (see Sec. A.6 of this Appendix).

A.5. Algorithm for solving the reduced-order ISDE. The reduced-order ISDE is solved with a step Δt of the continuous index parameter t (pseudotime). The pseudotime integration scheme is based on the use of the pseudotime instants t_k such that $t_k = k\Delta t$ for $k \geq 0$, at which $[\mathcal{Z}_k]$ denotes the pseudotime-discrete approximation to $[\mathcal{Z}(t_k)]$, $[\mathcal{Y}_k]$ denotes the pseudotime-discrete approximation to $[\mathcal{Y}(t_k)]$, and $[\mathcal{W}_k] = [\mathcal{W}(t_k)]$, with $[\mathcal{Z}_0] = [z_d]$, $[\mathcal{Y}_0] = [\mathcal{N}][a]$, and $[\mathcal{W}_0] = [0_{\nu,m}]$. For $k \geq 0$, let $[\Delta \mathcal{W}_{k+1}] = [\Delta \mathcal{W}_{k+1}][a]$ be the sequence of random matrices with values in $\mathbb{M}_{\nu,m}$, in which the increments $[\Delta \mathcal{W}_{k+1}]$ are independent random matrices with values in $\mathbb{M}_{\nu,N_{\text{ir}}}$ whose entries $[\Delta \mathcal{W}_{k+1}]_{ij}$ for $i = 1, \dots, \nu$ and $j = 1, \dots, N_{\text{ir}}$ are independent and centered Gaussian random variables such that

$$E\{[\Delta \mathcal{W}_{k+1}]_{ij} [\Delta \mathcal{W}_{k+1}]_{i'j'}\} = \Delta t \delta_{ii'} \delta_{jj'}. \quad (66)$$

For $k \geq 0$, the Störmer–Verlet scheme Burrage et al. [2007], Hairer et al. [2002], Soize [2017] is applied to the reduced-order ISDE:

$$[\mathcal{Z}_{k+\frac{1}{2}}] = [\mathcal{Z}_k] + \frac{\Delta t}{2} [\mathcal{Y}_k], \quad (67)$$

$$[\mathcal{Y}_{k+1}] = \frac{1-b}{1+b} [\mathcal{Y}_k] + \frac{\Delta t}{1+b} [\mathcal{L}_{k+\frac{1}{2}}] + \frac{\sqrt{f_0}}{1+b} [\Delta \mathcal{W}_{k+1}], \quad (68)$$

$$[\mathcal{Z}_{k+1}] = [\mathcal{Z}_{k+\frac{1}{2}}] + \frac{\Delta t}{2} [\mathcal{Y}_{k+1}], \quad (69)$$

with the initial condition defined in Sec. A.4, $b = f_0 \Delta t / 4$, and $[\mathcal{L}_{k+\frac{1}{2}}]$ the $\mathbb{M}_{\nu,m}$ -valued random matrix

$$[\mathcal{L}_{k+\frac{1}{2}}] = [\mathcal{L}([\mathcal{Z}_{k+\frac{1}{2}}])] = [L([\mathcal{Z}_{k+\frac{1}{2}}][g]^T)][a]. \quad (70)$$

A.6. The additional realizations $\{\mathbf{x}_{\text{ar}}^{(j)}, j = 1, \dots, N_{\text{ar}}\}$ of \mathbf{X}^ν . Let $\{([\mathcal{Z}_k(\theta)], [\mathcal{Y}_k(\theta)]), k = 0, \dots, M\}$ be one sample path of the $(\mathbb{M}_{\nu,m} \times \mathbb{M}_{\nu,m})$ -valued stochastic process $\{([\mathcal{Z}_k], [\mathcal{Y}_k]), k = 0, \dots, M\}$ computed by using the algorithm presented in Sec. A.5. Let k_0 be an integer such that for $k \geq k_0$, the pseudotime-discrete solution is converged to the stationary solution. Let M_0 be a given positive integer such that the pairs $[\mathcal{Z}_k]$ and $[\mathcal{Z}_{k+M_0}]$ are random matrices that are approximatively independent. For $\ell = 1, \dots, n_{\text{MC}}$ with n_{MC} such that $N_{\text{ar}} = n_{\text{MC}} \times N_{\text{ir}}$, for $i = 1, \dots, N_{\text{ir}}$, and for $i' = 1, \dots, \nu$, we set $\{\eta_{\text{ar}}^{(j)}\}_{i'} = [[\mathcal{Z}_{k_0+\ell M_0}(\theta)][g]^T]_{i'i}$ with $j = (\ell-1)N_{\text{ir}} + i$. The realizations $\{\mathbf{x}_{\text{ar}}^{(j)}, j = 1, \dots, N_{\text{ar}}\}$ of \mathbf{X}^ν are then calculated for $j = 1, \dots, N_{\text{ar}}$ as $\mathbf{x}_{\text{ar}}^{(j)} = \underline{\mathbf{x}} + [\varphi][\lambda]^{1/2} \eta_{\text{ar}}^{(j)}$.

A.7. Convergence of the probabilistic learning with respect to dimension N_{ir} of the initial data set. The additional realizations $\{\mathbf{x}_{\text{ar}}^{(j)}, j = 1, \dots, N_{\text{ar}}\}$ of \mathbf{X}^ν depend on N_{ir} , ν , and m , in which ν and m depend, in general, on N_{ir} . To emphasize these dependencies, \mathbf{X}^ν is denoted by $\mathbf{X}^{N_{\text{ir}}, \nu(N_{\text{ir}}), m(N_{\text{ir}})}$. By construction, $\mathbf{X}^{N_{\text{ir}}, \nu(N_{\text{ir}}), m(N_{\text{ir}})}$ is a second-order random variable with values in \mathbb{R}^n . The convergence of the PLoM method can be analyzed via the mean-square convergence of the sequence of random variables $\mathbf{X}^{N_{\text{ir}}, \nu(N_{\text{ir}}), m(N_{\text{ir}})}$ indexed by N_{ir} . This convergence could also be studied in distribution (mean-square convergence implies convergence in distribution). Nevertheless, the following comment must be made. The PLoM method is proposed for improving the construction of the generator of \mathbf{X} for the small-data case, that is, for N_{ir} relatively small because one evaluation of the response of the computational model is assumed to be computationally expensive. This means

that, in practice, the value of N_{ir} is constrained by the total computational cost, so that it may not be possible to take the value of N_{ir} as high as one may wish. Consequently, a convergence analysis could run as follows. Subsets $D_{N_1} \subset D_{N_2} \subset \dots \subset D_{N_\mu}$ of the initial dataset can be introduced with $1 < N_1 < N_2 < \dots < N_{\max}$. Then, the convergence can be analyzed with respect to the finite number of values $N_1 < N_2 < \dots < N_\mu$ of N_{ir} . If a reasonable convergence cannot be obtained using such a convergence analysis of the $\mathbf{X}^{N_j, \nu(N_j), m(N_j)}$ for $j = 1, \dots, \mu$, this convergence analysis would suggest that new evaluations have to be carried out by using the computational model in order to increase the value of N_{\max} .

References

- I. Alhossen, F. Baudoin, F. Bugarin, S. Segonds, and G. Teyssèdre. Use of Sobol indexes for efficient parameter estimation in a charge transport model. *IEEE T. Dielect. El. In.*, 26(2):584–592, 2019. DOI: 10.1109/TDEI.2018.007702.
- M. Arnst and J.-P. Ponthot. An overview of nonintrusive characterization, propagation, and sensitivity analysis of uncertainties in computational mechanics. *Int. J. Uncertain. Quan.*, 4(5):387–421, 2014. DOI: 10.1615/Int.J.UncertaintyQuantification.2014006990.
- E. Borgonovo. A new uncertainty importance measure. *Reliab. Eng. Syst. Safe.*, 92(6):771–784, 2007. DOI: 10.1016/j.ress.2006.04.015.
- K. Bulthuis, M. Arnst, S. Sun, and F. Pattyn. Uncertainty quantification of the multi-centennial response of the Antarctic ice sheet to climate change. *The Cryosphere*, 13:1349–1380, 2019. DOI: 10.5194/tc-13-1349-2019.
- K. Burrage, I. Lenane, and G. Lythe. Numerical methods for second-order stochastic differential equations. *SIAM J. Sci. Comput.*, 29(1):245–264, 2007. DOI: 10.1137/050646032.
- G. Chastaing, F. Gamboa, and C. Prieur. Generalized Hoeffding–Sobol decomposition for dependent variables — application to sensitivity analysis. *Electron. J. Stat.*, 6:2420–2448, 2012. DOI: 10.1214/12-EJS749.
- G. Chastaing, F. Gamboa, and C. Prieur. Generalized Sobol sensitivity indices for dependent variables: numerical methods. *J. Stat. Comput. Sim.*, 85(7):1306–1333, 2015. DOI: 10.1080/00949655.2014.960415.
- R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *P. Natl. Acad. Sci. USA*, 102(21):7426–7431, 2005. DOI: 10.1073/pnas.0500334102.
- T. Crestaux, O. Le Maître, and J.-M. Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliab. Eng. Syst. Safe.*, 94(7):1161–1172, 2009. DOI: 10.1016/j.ress.2008.10.008.
- S. Da Veiga. Global sensitivity analysis with dependence measures. *J. Stat. Comput. Sim.*, 85(7):1283–1305, 2014. DOI: 10.1080/00949655.2014.945932.
- S. Da Veiga and F. Gamboa. Efficient estimation of sensitivity indices. *J. Nonparametr. Stat.*, 25(3):573–595, 2013. DOI: 10.1080/10485252.2013.784762.
- C. Farhat, R. Tezaur, T. Chapman, P. Avery, and C. Soize. A feasible probabilistic learning method for model-form uncertainty in vibration analysis. *AIAA J.*, 57(1), 2019. DOI: 10.2514/1.J057797.
- R. Ghanem and C. Soize. Probabilistic nonconvex constrained optimization with fixed number of function evaluations. *Int. J. Numer. Meth. Eng.*, 113(4):719–741, 2018. DOI: 10.1002/nme.5632.

- R. Ghanem, D. Higdon, and H. Owhadi. *Handbook of Uncertainty Quantification*. Springer, Cham, Switzerland, 2017.
- R. Ghanem, C. Soize, and C. Thimmiisetty. Optimal well-placement using probabilistic learning. *Data-Enabled Discovery and Applications*, 2, 2018. DOI: 10.1007/s41688-017-0014-x.
- R. Ghanem, C. Soize, C. Safta, X. Huan, G. Lacaze, J. Oefelein, and H. Najm. Design optimization of a scramjet under uncertainty using probabilistic learning on manifolds. *J. Comput. Phys.*, 399, 2019. DOI: 10.1016/j.jcp.2019.108930.
- N. Golledge, D. Kowalewski, T. Naish, R. Levy, C. Fogwill, and E. Gasson. The multi-millennial Antarctic commitment to future sea-level rise. *Nature*, 526:421–425, 2015. DOI: 10.1038/nature15706.
- E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer, Heidelberg, Germany, 2002.
- J. Hart and P. Gremaud. An approximation theoretic perspective of Sobol' indices with dependent variables. *Int. J. Uncertain. Quan.*, 8(6):483–493, 2018. DOI: 10.1615/Int.J.UncertaintyQuantification.2018026498.
- A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM-Probab. Stat.*, 18:342–364, 2014a. DOI: 10.1051/ps/2013040.
- A. Janon, M. Nodet, and C. Prieur. Uncertainties assessment in global sensitivity indices estimation from metamodels. *Int. J. Uncertain. Quan.*, 4(1):21–36, 2014b. DOI: 10.1615/Int.J.UncertaintyQuantification.2012004291.
- M. Lamboni. Global sensitivity analysis: An efficient numerical method for approximating the total sensitivity index. *Int. J. Uncertain. Quan.*, 6(1):1–17, 2016. DOI: 10.1615/Int.J.UncertaintyQuantification.2016012354.
- B. Laurent. Efficient estimation of integral functionals of a density. *The Annals of Statistics*, 24(2): 659–681, 1996. Retrieved May 15, 2020, from www.jstor.org/stable/2242667.
- O. Le Maître and O. Knio. *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*. Springer, Heidelberg, Germany, 2010.
- X. Luo, Z. Lu, and X. Xu. Non-parametric kernel estimation for the ANOVA decomposition and sensitivity analysis. *Reliab. Eng. Syst. Safe.*, 130:140–148, 2014. DOI: 10.1016/j.ress.2014.06.002.
- J. Nossent, P. Elsen, and W. Bauwens. Sobol' sensitivity analysis of a complex environmental model. *Environ. Modell. Softw.*, 26(12):1515–1525, 2011. DOI: 10.1016/j.envsoft.2011.08.010.
- J. Oakley and A. O'Hagan. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. Roy. Stat. Soc. B*, 66(3):751–769, 2004. DOI: 10.1111/j.1467-9868.2004.05304.x.
- A. Owen. Variance components and generalized Sobol' indices. *SIAM-ASA J. Uncertain.*, 1(1): 19–41, 2013. DOI: 10.1137/120876782.
- F. Pattyn. Sea-level response to melting of Antarctic ice shelves on multi-centennial timescales with the fast Elementary Thermomechanical Ice Sheet model (f.ETISh v1.0). *The Cryosphere*, 11:1851–1878, 2017. DOI: 10.5194/tc-11-1851-2017.
- E. Qian, B. Peherstorfer, D. O'Malley, V. Vesselinov, and K. Willcox. Multifidelity Monte Carlo estimation of variance and sensitivity indices. *SIAM-ASA J. Uncertain.*, 6(2):683–706, 2018. DOI: 10.1137/17M1151006.

- R. Rosolem, H. Gupta, W. Shuttleworth, X. Zeng, and L. Gonçalves. A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis. *J. Geophys. Res.*, 117(D7):1–18, 2012. DOI: 10.1029/2011JD016355.
- A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.*, 145(2):280–297, 2002. DOI: 10.1016/S0010-4655(02)00280-1.
- D. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley and Sons, New York, 2015.
- I. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simulat.*, 55(1–3):271–280, 2001. DOI: 10.1016/S0378-4754(00)00270-6.
- C. Soize. Steady-state solution of Fokker–Planck equation in higher dimension. *Probabilist. Eng. Mech.*, 3(4):196–206, 1988. DOI: 10.1016/0266-8920(88)90012-4.
- C. Soize. *The Fokker–Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions*. World Scientific Publishing Co Pte Ltd, Singapore, 1994.
- C. Soize. Construction of probability distributions in high dimension using the maximum entropy principle. applications to stochastic processes, random fields and random matrices. *Int. J. Numer. Meth. Eng.*, 76(10):1583–1611, 2008. DOI: 10.1002/nme.2385.
- C. Soize. Polynomial chaos expansion of a multimodal random vector. *SIAM-ASA J. Uncertain.*, 3(1):34–60, 2015. DOI: 10.1137/140968495.
- C. Soize. *Uncertainty Quantification. An Accelerated Course with Advanced Applications in Computational Engineering*. Springer, New York, 2017.
- C. Soize and C. Farhat. Probabilistic learning for modeling and quantifying model-form uncertainties in nonlinear computational mechanics. *Int. J. Numer. Meth. Eng.*, 117:819–843, 2019. DOI: 10.1002/nme.5980.
- C. Soize and R. Ghanem. Data-driven probability concentration and sampling on manifold. *J. Comput. Phys.*, 321:242–258, 2016. DOI: 10.1016/j.jcp.2016.05.044.
- C. Soize and R. Ghanem. Physics-constrained non-Gaussian probabilistic learning on manifolds. *Int. J. Numer. Meth. Eng.*, 121(1):110–145, 2020a. DOI: 10.1002/nme.6202.
- C. Soize and R. Ghanem. Probabilistic learning on manifolds. *arXiv preprint arXiv:2002.12653*, math.ST:1–41, 2020b.
- C. Soize, R. Ghanem, C. Safta, X. Huan, Z. Vane, J. Oefelein, G. Lacaze, H. Najm, Q. Tang, and X. Chen. Entropy-based closure for probabilistic learning on manifolds. *J. Comput. Phys.*, 388: 528–533, 2019. DOI: 10.1016/j.jcp.2018.12.029.
- B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Safe.*, 93(7):964–979, 2008. DOI: 10.1016/j.ress.2007.04.002.
- S. Tarantola and T. Mara. Variance-based sensitivity indices of computer models with dependent inputs: The Fourier amplitude sensitivity test. *Int. J. Uncertain. Quan.*, 7(6):511–523, 2017. DOI: 10.1615/Int.J.UncertaintyQuantification.2017020291.