Data-driven stochastic inversion under functional uncertainties

Mohamed Reda El Amri, Céline Helbert, Olivier Lepreux, Miguel Munoz Zuniga, Clémentine Prieur, Delphine Sinoquet

DATA-DRIVEN STOCHASTIC INVERSION UNDER FUNCTIONAL UNCERTAINTIES*

MOHAMED REDA EL AMRI [§¶], CELINE HELBERT[†], OLIVIER LEPREUX[‡], MIGUEL MUNOZ ZUNIGA [¶], CLEMENTINE PRIEUR[§], AND DELPHINE SINOQUET[¶]

Abstract. In this paper, we propose a new methodology to deal with an uncertain functional input in inversion problems through computer experiments. This study is motivated by an automotive application. In this context, the simulator code takes a double set of simulation inputs : deterministic control variables and functional random variables. This framework is characterized by two features. The first feature is the high computational cost of simulations, which makes the inversion in the presence of uncertainties unaffordable. The second feature is that the probability density of the functional input V is only known through a sample of realizations. The proposed method involves two imbricated tasks. A first task based on a bayesian approach aims at wisely choosing the new evaluations of the code in order to estimate the excursion set with a limited number of costly simulations. The second task targets on efficiently estimating the expectation over the functional random variable. As the uncertain variable is observable through a finite training sample, we present three ways to infer the distribution from data. Our method is illustrated and calibrated on an analytical example. It is then applied on the automotive industrial test case where the objective is to identify the set of control parameters leading to meet the pollutant emissions standards of a vehicle.

Key words. functional random variable, design of experiments, set estimation, Gaussian process models

AMS subject classifications. 60G15, 62D05, 62P30

1. Introduction.

In recent years, computer models are omnipresent in engineering and sciences, because the corresponding physical experimentation is costly or impossible to execute. Numerical models are adopted to study the behaviour of such physical phenomena. As underlined in [4, 7], practitioners are not only interested in the response of their model for a given set of inputs (forward problem) but also in recovering the set of input values leading to a prescribed value or range for the output of interest. The problem of estimating such set is called hereafter inversion problem. We will consider a system that evolves in an uncertain environment, the uncertainties appear for example due to manufacturing tolerances or environmental conditions. The numerical simulator modelling the system, denoted f, takes two types of input variables : a set of control variables $x \in \mathbb{X}$, and a set of uncertain variables $v \in \mathcal{V}$. Robust inversion consists in seeking the set of control variables $x \in \mathbb{X}$ such that $\sup_{v\in\mathcal{V}} f(x,v)$ is smaller than a threshold c. Then, the difficulty of solving the robust inversion problem strongly depends on the uncertainty set $\mathcal{V}$. In our setting, $\mathcal{V}$ is a functional space, and we consider the inversion problem under uncertainty as a stochastic inversion problem. It means that we assume that the uncertainty has a probabilistic description. Let V denote the random variable, valued in $\mathcal{V}$, modelling the uncertainty. In our framework, we are interested in constructing the set : $\Gamma^* := \{x \in \mathbb{X} \ , \ \mathbb{E}_{\mathrm{V}}[f(x,\mathrm{V})] \leq c\}$, with $c \in \mathbb{R}$. Other sets could be considered such

*Submitted to the editors DATE.

Funding: ....

[†]Université de Lyon, UMR 5208, Ecole Centrale de Lyon, Institut Camille Jordan (celine.helbert@ec-lyon.fr).

[‡]IFPEN, Lyon, France (olivier.lepreux@ifp.fr).

[§]Université Grenoble Alpes, France (clementine.prieur@univ-grenoble-alpes.fr,mohamed-reda.el-amri@univ-grenoble-alpes.fr).

[¶]IFPEN, Rueil-Malmaison, France (delphine.sinoquet@ifp.fr, miguel.munoz-zuniga@ifp.fr, mohamed-reda.el-amri@ifp.fr).

as $\Gamma_\alpha = \{x \in \mathbb{X}\ ,\ \mathbb{P}_V(f(x,\mathrm{V}) \leq c) \geq 1-\alpha\}$, $\alpha \in [0,1]$. Most of the methodology presented in our paper could be adapted to that case.

In our framework, the probability distribution of V is only known from a set of realizations. We thus aim at replacing the expectation in the definition of Γ^* by a Monte Carlo estimate. In that sense, our methodology is a data-driven procedure.

Inverse problems have already been carried out in many applications, notably reliability engineering (see, e.g., [4], [7]), climatology (see, e.g., [5], [14]) and many other fields. In the literature, one way to solve the problem is to adopt a sequential sampling strategy based on Gaussian process emulators. The idea is that gaussian process emulators, which capture prior knowledge about the regularity of the unknown function $g : x \mapsto \mathbb{E}_{\mathrm{V}}[f(x,\mathrm{V})]$, make it possible to assess the uncertainty about Γ^* given a set of evaluations of g. More specifically, these sequential strategies for the estimation of an excursion set are closely related to the field of Bayesian global optimization (see, e.g., [7]). In the case of inversion problems, Stepwise Uncertainty Reduction (SUR) strategies based on set measures were introduced in [31]. More recently, a parallel implementation of these strategies have been proposed in [8] and applied to the problem of construction of an excursion set. Briefly, the strategy SUR gives sequentially the next location where to evaluate the function g in order to minimize an uncertainty function. The key contribution of the present paper is to propose a data-driven adaptation of that procedure in the presence of functional uncertainties.

The paper is organised as follows. In Section 2 we introduce the Bayesian framework and fundamental notions of the infill strategy, Stepwise Uncertainty Reduction (SUR). In Section 3, we present a new method and recall two existing ones to quantify the uncertainty associated with functional random variable: one based on a probabilistic modelling approach and the other one related to the so-called scenario approach. We also define a new more efficient method for estimating the expectation using tools from both existing methods. In Section 4, we introduce our data-driven methodology and describe our algorithms. Finally, in Section 5, we illustrate the overall procedure on an analytical example and then apply it to an industrial test case.

2. Background on SUR strategies. Let $f : \mathbb{X} \times \mathcal{V} \longrightarrow \mathbb{R}$ denote the unknown real-valued continuous function, where $\mathbb{X}$ is a bounded subset of $\mathbb{R}^p$, $p \geq 1$, and $\mathcal{V}$ a functional space on which a functional random variable V is defined. Moreover, we suppose that a finite set of N realizations of the functional random variable V is available. In the following, we consider the expectation as the robustness measure and we are interested in characterizing the set of control variables which leads a system to satisfy a safe behaviour:

$$\begin{aligned} \Gamma^* &:= \{x \in \mathbb{X}\ ,\ \mathbb{E}_{\mathrm{V}}[f(x,\mathrm{V})] \in C\} \\ &:= \{x \in \mathbb{X}\ ,\ g(x) \in C\} \quad \text{with} \quad C = (\infty, c],\ c \in \mathbb{R}. \end{aligned} \tag{1}$$

While the function f depends on two separate types of inputs (control and uncertain variables), our objective function g depends only on the control variables, i.e., for each setting of control variables, the objective function is the mean of f over the unknown distribution of the uncertain variable.

The estimation of Γ^* by a systematic exploration of $\mathbb{X}$ requires far too many evaluations of g. Therefore, statistical methods based on a reduced number of evaluation points are widely used to overcome this latter difficulty (see [30], [2], [7]) by focusing the evaluations on the 'promising' subregion of the control space.

These methods usually begin by an exploration phase, during which the output of the code is computed on an experimental design of size n. This initial design is then se-

quentially expanded by adding new goal oriented points. This procedure is introduced below.

2.1. Random closed set and bayesian framework. In a Bayesian framework, we assume that g is a realization of an almost surely continuous Gaussian process $Y \sim GP(m, k)$ with a mean structure m, defined as, $m(x) = \mathbb{E}[Y_x]$, $x \in \mathbb{X}$, and a covariance kernel k, defined as, $k(x, x') := Cov(Y_x, Y_{x'})$, $x, x' \in \mathbb{X}$. Due to the stochastic nature of $(Y_x)_{x\in\mathbb{X}}$, the associated excursion set,

$$\Gamma := \{x \in \mathbb{X} \ , \ Y_x \in C\} \tag{2}$$

is a random closed set. From the assumption that g is a realization of Y, the true unknown set Γ^* can be seen as a realization of the random closed set Γ. Therefore, we need to build a Stepwise Uncertainty Reduction strategy (SUR) that aims at reducing uncertainty on Γ by adding new evaluation points step by step. In this context such strategy relies on the notion of uncertainty for random sets. This latter will be characterized by the Vorob'ev expectation and deviation for random sets introduced in Subsection 2.2. The principle of SUR strategies are alo recalled in Subsection 2.3.

2.2. Vorob'ev approach. Γ defines a random closed set. Several ways to define the expectation of a random set have been developed in the literature (see e.g. [22]). In this paper, we focus on Vorob'ev expectation, closely related to quantiles and level-sets which have been the subject of many developments. Let us define *the coverage probability function* of a random set Γ as

$$p(x) = \mathbb{P}(x \in \Gamma), \quad x \in \mathbb{X}. \tag{3}$$

The coverage function p is not always available in an analytical closed form. For instance [12], [11] propose to replace p by an empirical counterpart and established consistency of plug-in estimators under weak assumptions. Here thanks to the Gaussianity assumption we will work with a closed form criterion. For $\alpha \in [0, 1]$, the α-level set of $p(x)$, also known as the *Vorob'ev Quantile*, is

$$Q_\alpha = \{x \in \mathbb{X} \ : \ p(x) \geq \alpha\}. \tag{4}$$

Let μ be a Borel σ-finite measure defined on $\mathbb{X}$, the Vorob'ev expectation is defined as Q_{α^*} with α^* such that the volume of Q_{α^*} matches the mean volume of Γ in the following sense :

$$\forall \beta > \alpha^* \ , \ \mu(Q_\beta) < \mathbb{E}[\mu(\Gamma)] \leq \mu(Q_{\alpha^*}). \tag{5}$$

Finally we introduce a notion of variability, based on the concept of expected distance between two random sets. First, let us consider two random closed sets A, B, the expected distance between A, B with respect to the measure μ is defined as

$$d_\mu(A, B) = \mathbb{E}[\mu(A \triangle B)], \tag{6}$$

where $A \triangle B$ is the symmetric difference of A and B. The quantity $\mathbb{E}[\mu(\Gamma \triangle Q_{\alpha^*})]$ is called *Vorob'ev deviation*.

In the following, we use the Vorob'ev expectation and deviation to quantify the variability of Γ conditionally to available observations. Let us denote the initial design points as $\mathcal{X}_n = (x_1, x_2, ..., x_n) \in \mathbb{X}^n$ and the responses at these points as

$\mathbf{g}_{\mathcal{X}_\mathbf{n}} = (g(x_1), g(x_2), ..., g(x_n)) \in \mathbb{R}^n$. We note $Y_{\mathcal{X}_n} = (Y_{x_1}, Y_{x_2}, ..., Y_{x_n})$ the random vector associated to the random process Y considered at $\mathcal{X}_n$. The main object of interest is then the conditional probability distribution of the random closed set Γ given the n observations. We know that the posterior distribution of the process Y given the n available observations remains Gaussian and is characterized by the posterior mean $m_n(x) = \mathbb{E}[Y_x \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}]$, $x \in \mathbb{X}$, and the posterior covariance kernel, $k_n(x, x') := Cov(Y_x, Y_{x'} \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}})$. The *coverage probability function* (3) and *Vorob'ev Quantile* (4) given the n observations can easily be computed as follows :

$$
\begin{aligned}
p_n(x) &= \mathbb{P}(x \in \Gamma \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}) = \mathbb{P}(Y_x \leq c \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}) = \Phi\Big(\frac{c - m_n(x)}{\sqrt{k_n(x,x)}}\Big), \\
Q_{n,\alpha} &= \{x \in \mathbb{X} \ : \ p_n(x) \geq \alpha\} = \{x \in \mathbb{X} \ : \ m_n(x) + \Phi^{-1}(\alpha)\sqrt{k_n(x,x)} \leq c\},
\end{aligned} \tag{7}
$$

where $\Phi(.)$ denotes the cumulated distribution function (c.d.f.) of the standard Gaussian distribution. The Vorob'ev expectation Q_{n,α_n^*} can be determined by tuning α to a level α_n^* such that $\mu(Q_{n,\alpha_n^*}) = \mathbb{E}[\mu(\Gamma) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}]$, knowing that $\mathbb{E}[\mu(\Gamma) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}] = \int_{\mathbb{X}} p_n(x)\mu(\mathrm{dx})$, this can be done through simple dichotomy or more advanced technique as, e.g., brent's method.
Once the Vorob'ev expectation is determined, the computation of Vorob'ev deviation $\mathbb{E}[\mu(\Gamma \triangle Q_{n,\alpha_n^*}) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}]$ can be expressed as a function of the posterior coverage probability function p_n and does not require simulations of Γ. Indeed,

$$
\begin{aligned}
\mathbb{E}_n[\mu(\Gamma \triangle Q_{n,\alpha_n^*})] &= \mathbb{E}_n\Big[\mu(\Gamma \cap Q^c_{n,\alpha_n^*}) + \mu(Q_{n,\alpha_n^*} \cap \Gamma^c)\Big] \\
&= \mathbb{E}_n\Big[\int_{\mathbb{X}} 1_{\{x\in\Gamma, x\notin Q_{n,\alpha_n^*}\}} + 1_{\{x\notin\Gamma, x\in Q_{n,\alpha_n^*}\}}\mu(\mathrm{dx})\Big] \\
&= \int_{Q^c_{n,\alpha_n^*}} \mathbb{E}_n[1_{\{x\in\Gamma\}}]\mu(\mathrm{dx}) + \int_{Q_{n,\alpha_n^*}} \mathbb{E}_n[1_{\{x\notin\Gamma\}}]\mu(\mathrm{dx}) \\
&= \int_{Q^c_{n,\alpha_n^*}} p_n(x)\mu(\mathrm{dx}) + \int_{Q_{n,\alpha_n^*}} (1 - p_n(x))\mu(\mathrm{dx}),
\end{aligned} \tag{8}
$$

where $\mathbb{E}_n[.] = \mathbb{E}[. \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}]$. It must be emphasized that the Vorob'ev expectation and deviation depend on the n available observations. Therefore when more evaluation points are added, these two quantities change. The aim of the following strategy is to wisely choose the next evaluations to adaptively reduce the uncertainty on Vorob'ev expectation. Such a strategy using definitions previously introduced has been proposed in [7].

2.3. SUR strategies. The principle of stepwise uncertainty reduction (SUR) (see, e.g., [4]; [8]) is to define an uncertainty measure, depending on the objective to be fulfilled, and to sequentially choose the points that decrease most this uncertainty. In other words, the aim of the SUR strategy is to construct a sequence of evaluation locations in order to reduce the *expected* uncertainty on a quantity of interest.
Here, we work in the particular setting where g is a sample path of a random process Y. The uncertainty function for an estimate of Γ is defined as a function $\mathcal{H}^{\text{uncert}}$ that associates to any finite sequence of observations $(\mathcal{X}_n, \mathbf{g}_{\mathcal{X}_\mathbf{n}})$ a real value representing the uncertainty on the estimation of Γ. When n observations are available, we denote by $\mathcal{H}_n^{\text{uncert}}$ the uncertainty at step n. We assume that we have r evaluations left. The objective of the SUR strategy is to find r optimal locations $x_{n+1}, ..., x_{n+r}$ such that the uncertainty $\mathcal{H}_{n+r}^{\text{uncert}}$ is as small as possible.

In what follows, we consider the Vorob'ev deviation as the uncertainty function, at step n,

$$\mathcal{H}_n^{\text{uncert}} = \mathbb{E}[\mu(\Gamma \triangle Q_{n,\alpha_n^*}) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}] \tag{9}$$

One way of constructing the optimal sequence $x_{n+1}, ..., x_{n+r}$ is to choose at each step the point that gives the smallest uncertainty $\mathcal{H}_{n+1}^{\text{uncert}}$,

$$\mathcal{H}_{n+1}^{\text{uncert}}(x) = \mathbb{E}[\mu(\Gamma \triangle Q_{n+1,\alpha_n^*}) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}\ , Y_x] \tag{10}$$

We note that the future uncertainty $\mathcal{H}_{n+1}^{\text{uncert}}$ is function of Y_x given $Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}$. Therefore, at each step we choose the point that gives the smallest uncertainty in expectation, that is :

$$\begin{aligned} x_{n+1} &= \text{argmin}_{x \in \mathbb{X}}\ \mathbb{E}_{n,x}[\mathcal{H}_{n+1}^{\text{uncert}}(x)] \\ &= \text{argmin}_{x \in \mathbb{X}}\ \mathcal{J}_n(x), \end{aligned} \tag{11}$$

where $\mathbb{E}_{n,x}$ denotes the expectation with respect to $Y_x | Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}$.
After having evaluated the function g at the optimal location x_{n+1}, we update the parameters of the posterior mean and covariance, and we restart until the evaluation budget r is exhausted. Such strategy is called *one-step lookahead*, which means that we select the next evaluation point as if it were the last one.
In practice, it is possible to run the model evaluations in parallel. Thus, we aim at finding an optimal **batch** of q points to reduce the uncertainty. Such strategy is called *batch sequential one-step lookahead*. In this setting, the sampling criterion $\mathcal{J}_n$ is the expected uncertainty at next step assuming the batch of q points $\mathbf{x} = (x^{(1)}, x^{(2)}, ..., x^{(q)}) \in \mathbb{X}^q$ is evaluated.

$$\mathcal{J}_n(\mathbf{x}) = \mathbb{E}_{n,\mathbf{x}}[\mathcal{H}_{n+q}^{\text{uncert}}(\mathbf{x})], \tag{12}$$

where $\mathbb{E}_{n,\mathbf{x}}$ denotes the expectation with respect to $Y_\mathbf{x} | Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_\mathbf{n}}$. As previously, we aim at choosing the batch of q points that gives the smallest *expected* uncertainty,

$$(x_{n+1}, x_{n+2}, ..., x_{n+q}) = \text{argmin}_{\mathbf{x} \in \mathbb{X}^q}\ \mathcal{J}_n(\mathbf{x}). \tag{13}$$

PROPOSITION 2.1 ($\mathcal{J}_n$ criterion). *Under the previous assumptions, the criterion $\mathcal{J}_n$ can be expanded in a closed-form expression as (the proof can be found in [7], Chapter 4.2),*

$$\begin{aligned} \mathcal{J}_n(\mathbf{x}) = \int_{\mathbb{X}} \Bigg(& 2\Phi_2\Bigg(\Bigg(\begin{bmatrix} a_{n+q}(u) \\ \Phi^{-1}(\alpha_n^*) - a_{n+q}(u) \end{bmatrix}; \begin{bmatrix} 1 + \gamma_{n+q}(u) & -\gamma_{n+q}(u) \\ -\gamma_{n+q}(u) & \gamma_{n+q}(u) \end{bmatrix} \Bigg) \Bigg) \\ & + p_n(u) + \Phi\Big(\frac{a_{n+q}(u) - \Phi^{-1}(\alpha_n^*)}{\sqrt{\gamma_{n+q}(u)}}\Big) \Bigg) d\mu(u) \end{aligned} \tag{14}$$

where

$$\begin{aligned} \mathbf{x} &= (x^{(1)}, x^{(2)}, ..., x^{(q)}) \in \mathbb{X}^q \\ a_{n+q}(u) &= \frac{c - m_n(u)}{s_{n+q}(u)} \\ \gamma_{n+q}(u) &= \mathrm{b}_{n+q}^t(u) K_q \mathrm{b}_{n+q}(u) \\ \mathrm{b}_{n+q}(u) &= \frac{K_q^{-1} k(\mathbf{x}, u)}{s_{n+q}(u)}\ ,\ u \in \mathbb{X} \end{aligned}$$

with $s_{n+q}(u) = \sqrt{k_{n+q}(u,u)}$, $k(\mathbf{x},u) = (k(x^{(1)},u), k(x^{(2)},u), ..., k(x^{(q)},u))^t$. K_q *is the covariance matrix with elements* $[k(x^{(i)},x^{(j)})]_{i,j=1,...,q}$ *and* $\Phi_2(.,\Sigma)$ *is the c.d.f. of the centered bivariate Gaussian distribution with covariance matrix* Σ.

For more theoretical perspectives on the SUR strategies, see [3] and references therein.

The sampling criterion $\mathcal{J}_n$ (14) is used in order to select the next evaluations of the function g. Once these locations are obtained by minimizing the sampling criterion, we have to evaluate the function g on these points, i.e., to estimate $g : x \mapsto \mathbb{E}_{\mathrm{V}}[f(x,\mathrm{V})]$, and to do so, we have to evaluate the function f for different realizations of the functional random variable V. This issue of efficiently evaluating an expectation over a functional random variable is discussed in the following section.

3. Uncertainty quantification of random functional variables. The uncertainty of the functional variable must be quantified. Uncertainty modelling of functional variables has already been studied in various contexts (see, e.g. [25]). In Subsection 3.3, we present a new method to characterize the uncertainty associated to functional variables, but first we recall two existing methods. The first one described in Subsection 3.1 relies on the theory of functional analysis, while the second one introduced in Subsection 3.2 relies on a discrete approximation of the functional distribution.

3.1. Dimension reduction and density estimation (Fpca). Let $\mathrm{V_t}$ be a real-valued random process, which is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and indexed on the bounded interval $I \subset \mathbb{R}$. Without loss of generality, we will consider a centred stochastic process with finite variance. In this work, the stochastic process $\mathrm{V_t}$ is observable through a finite sample of N realizations. In this context, uncertainty quantification methodology followed by most of authors (e.g. [27], [13]) can be separated into two steps, hereafter called the Fpca method.
First, the variable is decomposed on a truncated basis of functions. There are different prominent examples of fixed functional basis, we can cite B-Spline basis or wavelet basis (see [20] for more details about wavelets and wavelet transforms). The Karhunen-Loève (KL) expansion is another way of representing a stochastic process. It is based on the spectral expansion of the covariance function of the process. This covariance function, which can be viewed as a kernel, is defined as :

$$C(t,s) = \mathrm{Cov}(\mathrm{V}_t, \mathrm{V}_s) = \mathbb{E}[\mathrm{V}_t \mathrm{V}_s], \tag{15}$$

and the corresponding integral operator is :

$$\mathcal{T}_C \zeta(t) = \int_I C(t,s)\zeta(s)\mathrm{ds}. \tag{16}$$

We suppose that the covariance function is continuous in t, s. By Mercer's theorem, the corresponding integral operator $\mathcal{T}_C$ has an orthonormal basis of eigenfunctions $\{\psi_i(t)\}$. We define :

$$\eta_i = \int_I \mathrm{V}_t \psi_i(t)\mathrm{dt}, \tag{17}$$

then η_i are centered uncorrelated random variables with $\mathrm{var}(\eta_i) = \lambda_i$, where λ_i is the eigenvalue corresponding to ψ_i. We note that the use of the KL expansion is limited as the covariance function of the process is not known a priori. Nevertheless, we can

replace it by an empirical estimation based on the sample of N available realizations.. There are a number of computational approaches that can be used to compute the eigenfunctions. The most straightforward is to discretize, and to solve the eigenfunctions problem in its discretized form. Alternatively, the eigenfunctions can be decomposed in terms of a suitable basis of functions (e.g., B-splines), and the eigenequation (16) solved with respect to the coefficients in this basis (see, e.g., [27]). The KL expansion then takes the following form $\mathrm{V}_t = \sum_{i=1}^{\infty} \eta_i \psi_i(t)$. This latter can be approximately represented by a truncated KL expansion :

$$\mathrm{V}_t \simeq \sum_{i=1}^{m} \eta_i \psi_i(t), \tag{18}$$

where m denotes the truncation argument. We emphasize that among many possible decompositions, as Fourier series, wavelets or B-splines, of a random process, the KL expansion is optimal in the sense that the mean-square error resulting from a finite representation of the process is minimized.

The second step of the method consists in estimating the joint probability density function of the m independent random variables η. For this latter purpose we use in the following a Gaussian Mixture Model (GMM) which considered the seeked distribution as the weighted sum of multidimensional Gaussian distributions (see, e.g., [24] for a detailed overview on the subject). The GMM parameters are estimated by the expectation-maximization algorithm (EM).

Finally, in order to generate a new sample of realizations of V_t, we start by sampling independently the scalar random variables η whose probability distribution is estimated beforehand, then we obtain the desired curve using the linear combination of equation (18).

3.2. Scenario modelling. Another method, suggested by Lilburne et al. [19], is to represent the uncertainty of the functional variable by directly sampling from the set of N realizations of V, denoted by $\Xi = \{v_1, ..., v_N\}$. The functional probability distribution is approximated by a uniform discrete distribution over Ξ. This method was applied in [19] to calculate the first and total-order sensitivity indices for spatial models for simulating nitrate transport from paddock to groundwater. Ruffo et al. [29] also used this approach to perform sensitivity analysis of a model for oil reservoir production forecasting. This approach is likewise used in many other fields (see, e.g., [15]; [18]; [1]; [21]).

3.3. Weighted scenario modelling. The two approaches, described previously, suffer from various pathologies. The first strategy [Subsection 3.1] depends on the approach used for the estimation of the KL coefficients distribution : non-parametric (distribution-free) model which makes no assumption about the probability distribution, or parametric model like GMM. To preserve the underlying curve structure, one should choose a large value for the truncation argument. However, parametric or non-parametric approaches for density estimation suffer from the curse of dimensionality. The second strategy [Subsection 3.2] has the merit of keeping the shape of curves, but the realizations are randomly selected from the initial set Ξ. However, sampling a few representative curves among Ξ is a difficult task. This is the reason why we propose here an optimal space filling strategy instead of the crude method of Lilbrune et al. [19]. Our aim is to explain at best the variability of V with a reduced sample of realizations. This approach can be related to randomized

quasi-Monte Carlo method which often improves the representativeness of the sample by reducing the sampling error. First, we place ourselves in a finite dimensional euclidean space $\mathbb{R}^m$ by Karhunen-Loève decomposition (see Subsection 3.1), this can be illustrated by the following diagram and Figure 1

$$\Xi = \{v_{ij}\}_{i=1:N}^{j=1:f} \longrightarrow \mathcal{G} = \{\eta_i\}_{i=1:N} \tag{19}$$

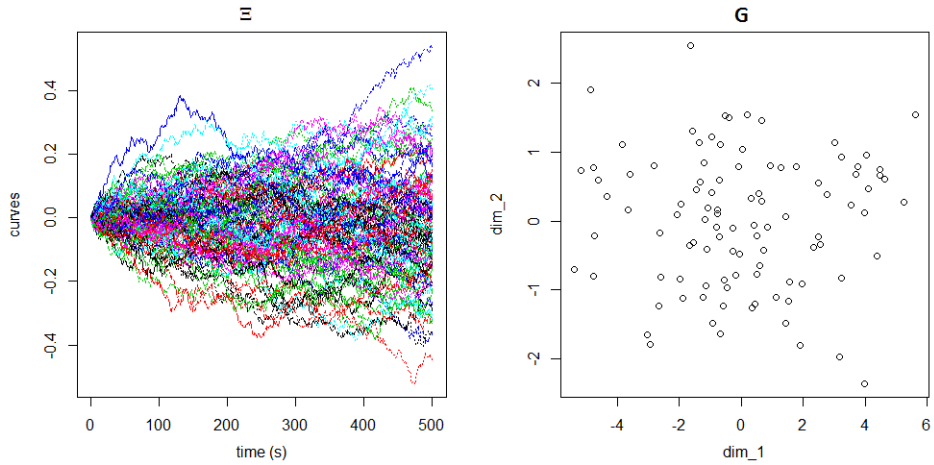

FIG. 1. *Ξ the sample of* 100 *realizations of* V *(left) and their representatives in the space of coefficients η (right), where* $m = 2$.

where $v_{ij} = v_i(t_j)$ is the discretized version of the function v_i at the points $t_1, ..., t_f$ of the interval I and $\eta_i \in \mathbb{R}^m$ the associated KL decomposition coefficients. Once the reduction dimension has been performed, each realization in Ξ will be represented by a point in $\mathcal{G}$. In the second step, to optimally explore the range of variation of V, we construct a space filling design, denoted D, using $\mathcal{G}$ as the candidate set, i.e., $D \subset \mathcal{G}$, such that D aims at obtaining the best coverage of the space of the coefficients $\mathcal{G}$. Before going further, let us briefly recall the notion of space-filling design, taking a purely model-free stance.

3.3.1. Space-Filling Design and quality criterion. Let us define by $D = \{d_1, ..., d_l\}$ a collection of l points. We denote

$$\texttt{dist}_{ij} = ||d_i - d_j||, \tag{20}$$

the euclidean distance between two design points d_i and d_j of D. One must then attempt to make the smallest distance between neighboring points in D as large as possible. We call a design that maximizes $\phi_{Mm}(D) = \min_{i \neq j} \texttt{dist}_{ij}$, a **maximin-distance design** (see Johnson et al [17]). There are several other intrinsic criteria in literature such as discrepancy that measures whether the distribution of the points of D is close to a uniform distribution. See Pronzato et al. [26] for a detailed overview on the subject.

In the following, we consider the maximin-distance criterion to construct our design, and since we want to select points from the set $\mathcal{G}$ of coefficients $\{\eta_i\}_{i=1:N}$, the design D can be obtained by finding the design of l points among N, that maximizes the criterion

$$\max \min_{i \neq j} ||\eta_i - \eta_j||. \tag{21}$$

3.3.2. Sequential Space-filling Design. Finding the design D is a computationally difficult problem. We could adapt the optimal design algorithms used in the literature such as simulated annealing (see Morris et al. [23]) and stochastic evolutionary algorithm (see Jin et al. [16]) for our purpose. Different from theirs, we propose a one-point-at-time greedy algorithm for the generation of our design. The sequential construction is described below

$$\begin{aligned} &\text{Initialisation} \ : \text{random sampling of } d_1 \in \mathcal{G} \\ &\forall l \geq 2 \quad , \quad D_l = \{d_i\}_{i=1:l} \\ &\qquad\qquad d_l = \underset{s \in \mathcal{G}}{\arg\max}\ \phi_{Mm}\Big(D_{l-1} \cup \{s\}\Big) \\ &\qquad\qquad \mathcal{G} = \{\eta_j\}_{j=1:N} \end{aligned} \tag{22}$$

The algorithm starts with a random point d_1, the next point is chosen among the points in $\mathcal{G}$ in order to maximize the maximin-distance criterion, which means that the next point is selected so that it is as far as possible from those previously selected. The advantages are twofold, the points are chosen optimally and also sequentially. As the selected points belong to the set $\mathcal{G}$, we recover the corresponding curves

$$\forall l \geq 2 \quad , \quad D_l = \{d_i\}_{i=1:l} \longrightarrow \tilde{D}_l = \{v_{(i)}\}_{i=1:l} \tag{23}$$

Figure 2 shows a construction of 10 points design applied to the example of Figure 1. In case the dimension reduction may lead to represent two different curves by the same point, thus in this case, we choose randomly one of them.

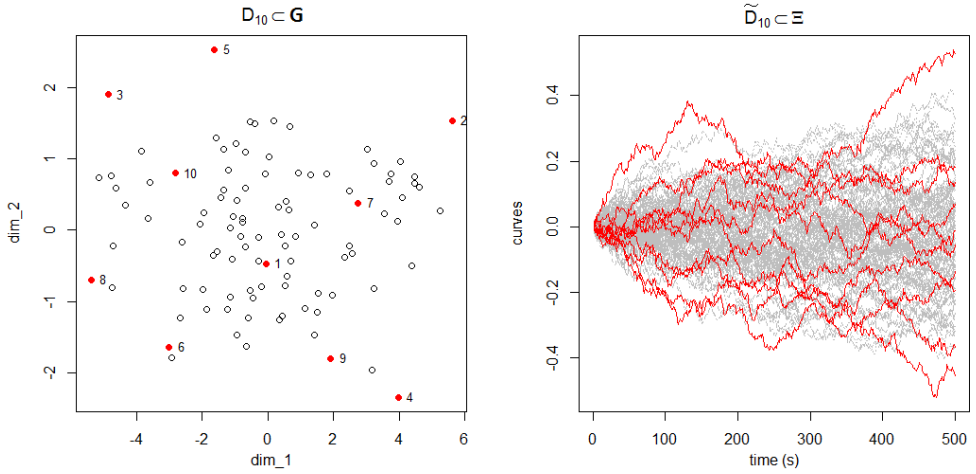

FIG. 2. *Sequential design of* 10 *points (red points) in the space of the coefficients* $\mathcal{G}$ *(left) and their corresponding red curves in* Ξ *(right).*

3.3.3. Weight calculus. The proposed method produces a weighted mean that has less variability than the arithmetic mean of a simple random sample. To do so, one must calculate the weight associated to each selected curve. We propose to use the concept of Voronoi cell: we associate to each design point a region consisting of all

points closer to that design point than to any other, and we derive the weight calculus

$$\forall l \geq 2\ ,\ \forall i \in \{1,..,l\}\ ,\ w_i^{(l)} = |S_{li}| \ \text{ and } \ |.| \text{ denotes the cardinality of set} \quad S_{li} = \{\eta \in \mathcal{G} \mid \forall h \in \{1,...,l\}\ ,\ d(\eta, d_i) \leq d(\eta, d_h)\} \tag{24}$$

The principal advantage of such a method is its performance to select even the extreme curves and to associate to them a small weight by definition of their uncommon nature. Such an approach can be considered as a variance reduction method. Note the complexity for the computation of the weights due to sequential approach.

4. Implementation. The whole computational aspect is carried out in the **R** environment : we use `DiceKriging` package [28] for Gaussian modelling and the sampling criterion $\mathcal{J}_n$ (14), used in order to select the next evaluation x_{n+1} of the function g, is already implemented in the `KrigInv` package [10]. We exploit the kriging update formulas [9] for faster updates of posterior mean and covariance. The sequentiality of our method to estimate the robustness measure on x_{n+1} leads us to define a *stopping criterion* on the expectation estimation $\hat{m}$. Thus, at each step in the estimation, we evaluate the absolute difference between two consecutive estimations of the expectation,

$$e_l(x_{n+1}) = |\hat{m}_{l-1}(x_{n+1}) - \hat{m}_l(x_{n+1})| \tag{25}$$

where $\hat{m}_i(x_{n+1}) = \frac{1}{\sum_{j=1}^{i} w_j^{(i)}} \sum_{j=1}^{i} w_j^{(i)} f(x_{n+1}, v_{(j)})$, we denote by $|.|$ the absolute value function. In the following, the *stopping criterion* must satisfy the condition,

$$e_{l-l_0}(x_{n+1}) \leq e_{l-l_0+1}(x_{n+1}) \leq ... \leq e_l(x_{n+1}) \leq \epsilon \tag{26}$$

it ensures that the quantities e_l are smaller than a tolerance ϵ on the l_0 previous steps in the estimation. These two parameters are set by practitioners. It allows to use fewer curves without sacrificing estimation accuracy.
The strategy SUR could be stopped if the allocated number of simulations is reached. However, we define in this work an additional stopping criterion based on the Vorob'ev deviation and close to the one defined for the expectation estimate. Thus, the strategy is carried out until the following stopping criterion is verified

$$e_{l-l_0}^{\text{SUR}} \leq e_{l-l_0+1}^{\text{SUR}} \leq ... \leq e_l^{\text{SUR}} \leq \epsilon, \tag{27}$$

where $e_i^{\text{SUR}} = |\ \mathbb{E}_{i-1}[\mu(\Gamma \triangle Q_{i-1,\alpha_{i-1}^*})] - \mathbb{E}_i[\mu(\Gamma \triangle Q_{i,\alpha_i^*})]\ |$ is the absolute error between two successive Vorob'ev deviations. The condition (27) tests if all the quantities are smaller than a tolerance ϵ on l_0 consecutive steps.
The global methodology to perform robust inversion in presence of functional uncertainty proposed in this paper is summarized in Algorithm 1.

Algorithm 1

Create an initial design of experiments (Doe) of n points in the control space $\mathbb{X}$
$l \leftarrow 2$
while Stopping criterion (27) not met (SUR) **do**
  $x_{n+1} \leftarrow$ Sampling criterion $\mathcal{J}_n$
  while Stopping criterion (26) not met (Expectation Estimation) **do**
    Augment design D_l and calculate weight w_l using (22) and (24)
    $l \leftarrow l + 1$
  end while
  Update Doe
  $n \leftarrow n + 1$
end while
end

5. Numerical tests. In this section we apply our new methodology based on the SUR strategy combined with the weighted scenario modelling of the functional uncertainties. It also possible to combine the SUR strategy with a dimension reduction of the functional uncertainty (Subsection 3.1) or with a basic scenario modelling (Subsection 3.2). On an analytical test case, we compare our approach with these two alternatives. We then present in Subsection 5.2 an application to the industrial automotive test case which motivate our study.

5.1. Analytical example. In this example, we define the function f as follows:

(28)
$$f : (x, \mathrm{V}) \mapsto |0.1\cos(x_1 \max_{t\in T} \mathrm{V}_t)\sin(x_2).(x_1 + x_2 \min_{t\in T} \mathrm{V}_t)^2|. \int_T (30 + \mathrm{V}_t)^{\frac{x_1.x_2}{20}} dt. \max_{t\in T} \mathrm{V}_t$$

where the control variable x lies in $\mathbb{X} = [1.5, 5] \times [3.5, 5]$, and V is a standard Brownian motion with state space in $\mathbb{R}$ and $T = [0, 1]$. We suppose that a sample of N realizations of V is available, denoted by Ξ, and these realizations are discretized uniformly on 100 points of T. The objective is to construct the set $\Gamma^* := \{x \in \mathbb{X}\ ,\ g(x) = \mathbb{E}_{\mathrm{V}}[f(x, \mathrm{V})] \leq c\}$, where $c = 1.2$.
Here we consider a Gaussian process prior $(Y_x)_{x\in\mathbb{X}} \sim GP(m, k)$, with constant mean function and Matérn covariance kernel with $\nu = 5/2$. The initial DoE consists of a 9 points LHS design optimized by maximin criterion. The covariance kernel hyperparameters are estimated by Maximum Likelihood Estimation (MLE). Figure 3 shows the initial design of experiments and the target set Γ^* obtained from a 30×30 grid experiment, where at each grid point the expectation is approximated by a Monte Carlo Method over 5000 realizations of V. We aim at estimating the set Γ^* using the SUR strategy to choose the next evaluation point as defined in Section 2, and the methods presented in Section 3 to provide an estimation of the expectation.

We proceed to add one point at each iteration of the SUR strategy until the condition (27) for $(l_0, \epsilon) = (4, 5.10^{-3})$ is reached. The covariance parameters are re-estimated at each step by MLE. Since this criterion is based on the Vorob'ev deviation, the objective is to reduce the uncertainty on the estimate until stability. For the sequential estimation of the expectation, we test the sensitivity to the parameters (l_0, ϵ) of criterion (26) (see Table 1).
The estimation of the expectation at the proposed point by SUR is carried out with the 3 methods (Fpca, Scenario, W.Scenario) detailed in Section 3. As presented in Section 4, the estimation is done sequentially and it depends on the *stopping criteria*

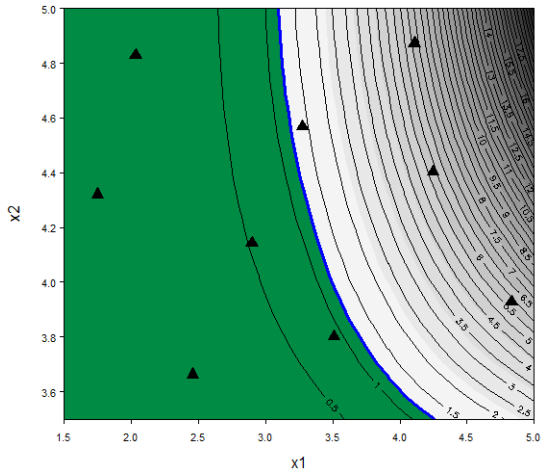

FIG. 3. *Analytical example. Contour plot of the function g, the set of interest (green) with boundary (blue line), the initial design of experiments (black triangles).*

parameters l_0, ϵ and on the truncation argument m. This latter is set at $m = 7$ in order to explain 97% of the variance. Our approach detailed in Subsection 3.3 is sequential by definition like the two others. The Monte Carlo approach (called Scenario method) is sequential because at each step a curve is drawn with replacement from the available sample Ξ. The same goes for the probabilistic approach (Fpca), at each step we add a new curve built as explained in Subsection 3.1. The first test consists in fixing the available sample of realizations of V (N=$\texttt{card}(\Xi)$=150). For this fixed sample, we compare the results obtained by the 3 uncertainty quantification methods for different l_0 and ϵ. The Table 1 lists the parameters tested in this section.

l_0	4	2	3	4
ϵ	10^{-2}	5.10^{-3}	5.10^{-3}	5.10^{-3}

TABLE 1
Analytical example. Estimation of expectation stopping criteria parameters

To compare the performance of the various methods we use the ratio between the volume of the symmetric difference between the true set Γ^* and the estimated set at last iteration, $\mu(\Gamma^* \triangle Q_{n_{\text{last}},\alpha^*_{n_{\text{last}}}})$ and the volume of the true set, $\mu(\Gamma^*)$.
Because of the sampling steps, the whole expectation estimation methods (Scenario, Fpca and Weighted Scenario) have a stochastic behaviour. Indeed, the Scenario and Fpca methods are stochastic by nature, the Weighted Scenario method depends on the starting curve which is randomly chosen from the available set of realizations Ξ. To account for these variabilities in the tests, the performance of each method is averaged over 30 independent runs. The results are summarized in Table 2.

	$\mu(\Gamma^* \triangle Q_{n_{\text{last}}, \alpha^*_{n_{\text{last}}}})/\mu(\Gamma^*)$			Number of calls to f		
(l_0, ϵ)	Scenario	Fpca	W.Scenario	Scenario	Fpca	W.Scenario
(4,1.e-2)	8.01 %	8.17 %	**5.09 %**	2372 (26)	2263 (28)	**513 (15)**
(2,5.e-3)	9.40 %	9.48 %	**7.85 %**	2189 (28)	2047 (26)	**365 (19)**
(3,5.e-3)	8.06 %	8.05 %	**4.77 %**	3121 (28)	2962 (28)	**525 (15)**
(4,5.e-3)	6.56 %	7.53 %	**4.26 %**	3913 (26)	3242 (24)	**573 (12)**

TABLE 2

Analytical example. Results obtained at the last iteration for different values of l_0, ϵ and methods. (.) : the average total number of iterations (SUR)

Table 2 indicates that the three methods are sensitive to the parameters l_0 and ϵ. Bigger is the parameter l_0, i.e., one seeks a stability of the estimation, smaller is the error but higher is the number of calls to the function. The weighted scenario method performs well both in terms of error and number of calls to the function f. The cumulative number of calls to f has been improved by a factor greater than 4 in comparison with the two other methods, even if in term of error the three methods remain close and of reasonable quality.

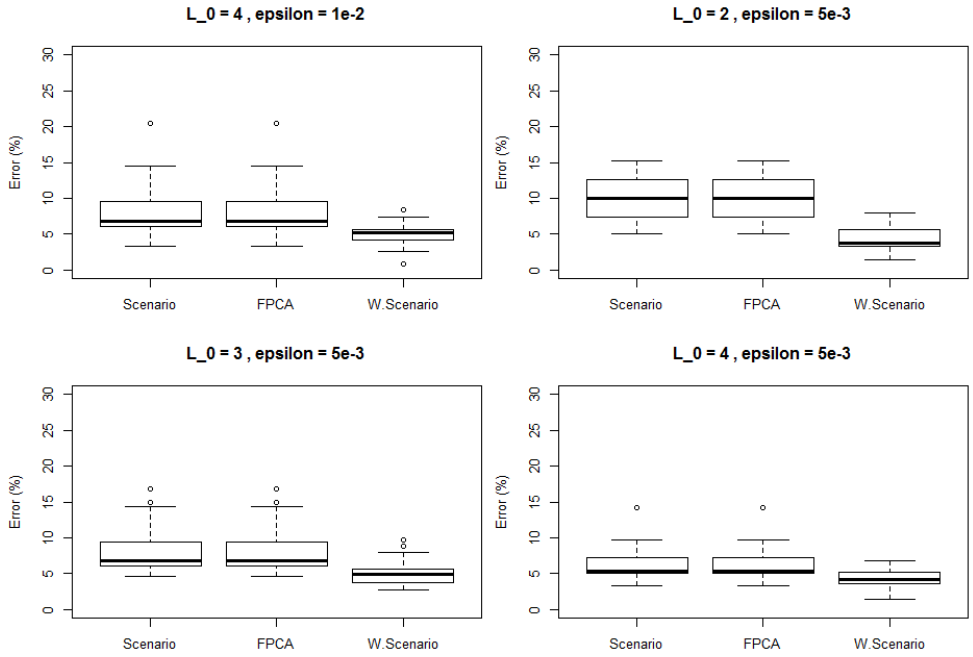

FIG. 4. *Analytical example. Boxplots of the error estimation for different values l_0, ϵ and methods (30 runs for each boxplot)*

Figure 4 presents a comparison of the errors at last iteration, i.e., when the stopping criteria associated to the SUR strategy is verified, for each method and each criterion associated to expectation estimation. This shows that the recovering errors are quite constant for the Weighted Scenario method. Thus, the starting point seems to have little influence on the quality of the recovering, and the latter can be chosen arbitrarily.

In the following, the stopping criteria for SUR ($l_0 = 4, \epsilon = 5.10^{-3}$) and for estimation of the expectation ($l_0 = 4, \epsilon = 10^{-2}$) are chosen because they offer a good compromise between the accuracy and the number of model evaluations.

Table 3 compares the sensitivity of the methods to the size of the available sample Ξ, denoted by N. In each cell of the table, we perform 20×20 independent runs. Indeed, for each value of N, we generate 20 training samples Ξ of size N and for each sample we perform 20 runs for each method. The table summarizes the results averaged over the 400 runs.

	$\mu(\Gamma^* \triangle Q_{n_{\text{last}}, \alpha^*_{n_{\text{last}}}})/\mu(\Gamma^*)$			Cumulative number of calls to f		
	Scenario	Fpca	W.Scenario	Scenario	Fpca	W.Scenario
N = 50	9.26%	9.33 %	**5.73 %**	3059 (34)	2988 (34)	**387** (14)
N = 100	9.05 %	9.19 %	**5.52 %**	3054 (34)	2786 (33)	**477** (16)
N = 200	8.61 %	8.84 %	**4.26 %**	3131 (35)	2850 (34)	**499** (17)

TABLE 3

Analytical example. Results obtained at the last iteration for different sample size and methods. (.) : the average total number of iterations (SUR)

We note that for a larger sample size, the recovering error is smaller. This can be explained by the fact that with a large sample size, the available information on variable V enables an effective estimation of the expectation.
We know that the Weighted Scenario and the probabilistic modelling (Fpca) depend on the truncation argument. To better understand the effect of the number of dimensions m, we fix the stopping criteria for the SUR strategy and expectation estimation, and we consider different values of m= $\{2, 3, 4, 5, 6\}$. Each cell of Table 4 represents the result averaged over 14×20 independent runs. For each m, we generate 14 samples Ξ of size N=200, and for each of them we perform 20 runs of each method.

	$\mu(\Gamma^* \triangle Q_{n_{\text{last}}, \alpha^*_{n_{\text{last}}}})/\mu(\Gamma^*)$		Cumulative number of calls to f	
	Fpca	W.Scenario	Fpca	W.Scenario
m = 2	12.21 %	**5.81 %**	2861 (34)	**536** (18)
m = 3	10.74 %	**4.7 %**	2851 (33)	**544** (14)
m = 4	10.17 %	**4.46 %**	2893 (34)	**515** (18)
m = 5	9.93 %	**4.4 %**	2980 (35)	**500** (18)
m = 6	8.89 %	**4.21 %**	2885 (34)	**494** (18)

TABLE 4

Analytical example. Results obtained at the last iteration for different values of the truncation argument. (.) : the average total number of iterations (SUR)

Table 4 shows that for all values of m, the Weighted Scenario outperforms the probabilistic Fpca modelling. As shown in Table 5, for high truncation argument, the explained variance increases, that explains the decrease of the estimation error for the probabilistic modelling (Fpca). In the other hand, the Weighted Scenario seems to be quite constant for m ≥ 3. This can be explained by the fact that the KL expansion is only used to define a space filling design, and the information lost by the truncation is recovered by tacking the corresponding curve in the set Ξ. On the contrary, the probabilistic modelling which is based on Fpca gives better results when m is higher.

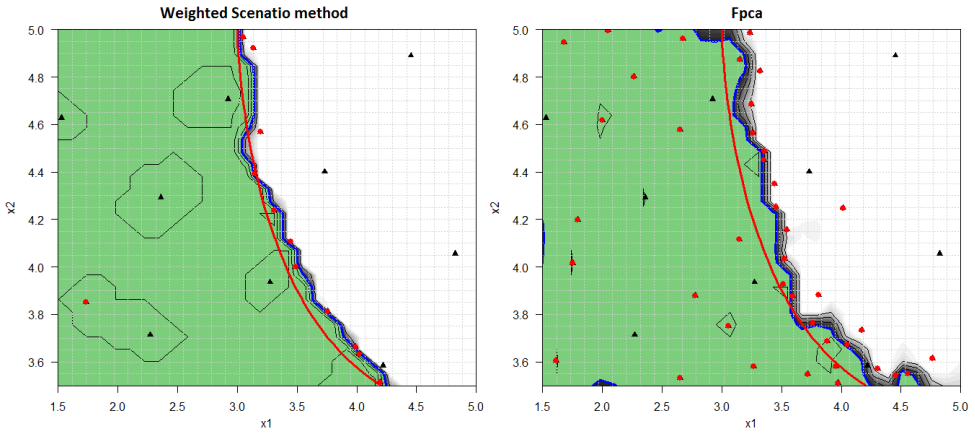

FIG. 5. *Analytical example. Results at the last iteration in the case $m = 4$. Coverage function, boundary of the true set (red), estimate set (green). The initial design of experiments (black triangles), the added points (red circles)*

However the errors in Table 4 seem to be bounded below. To go below that bound, we probably need to increase the size of Ξ. Figure 5 shows the coverage functions of the random set Γ obtained at the last iteration for the two methods and there corresponding estimate sets.

m	2	3	4	5	6
Explained variance	90.2 %	93.4 %	95.1 %	96 %	96.7%

TABLE 5
Analytical example. The explained variance in function of m

5.2. IFPEN test case. In this section we test the proposed method on a automotive test case from the IFPEN. The problem concerns an after-treatment device of diesel vehicles, called Selective Catalytic Reduction (SCR). This latter consists on a basic process of chemical reduction of nitrogen oxides (NO_x) to diatomic nitrogen (N_2) and water (H_2O) by the reaction of NO_x and ammonia NH_3. The reaction itself occurs in the SCR catalyst. Ammonia is provided by a liquid-reductant agent injected upstream of the SCR catalyst. The amount of ammonia introduced into the reactor is a critical quantity: overdosing causes undesirable ammonia slip downstream of the catalyst, whereas under-dosing causes insufficient NO_x reduction. In practice, ammonia slip is restricted to a prescribed threshold.
We use an emission-oriented simulator developed by IFPEN, which models the vehicle, its engine and the exhaust after-treatment system. It takes the vehicle driving cycle profile as input and provides the time-series of corresponding exhaust emissions as output. A realistic SCR control law is used in this simulator. See [6] for an example of such a control law. In this study, we choose two control variables as input and a functional one considered as random. The control variables are parameters of the SCR control law. They set the targeted level of NH_3 storage in the catalyst and then are indirectly related to the NH_3 injected. They lie in $\mathbb{X} = [0, 0.6]^2$. The functional random variable describes the evolution of vehicle speed on I, with $I = [0, 5400s]$. These functional uncertainties come from an available sample of 100 real driving cycles. A

subset of that sample is represented in Figure 6.

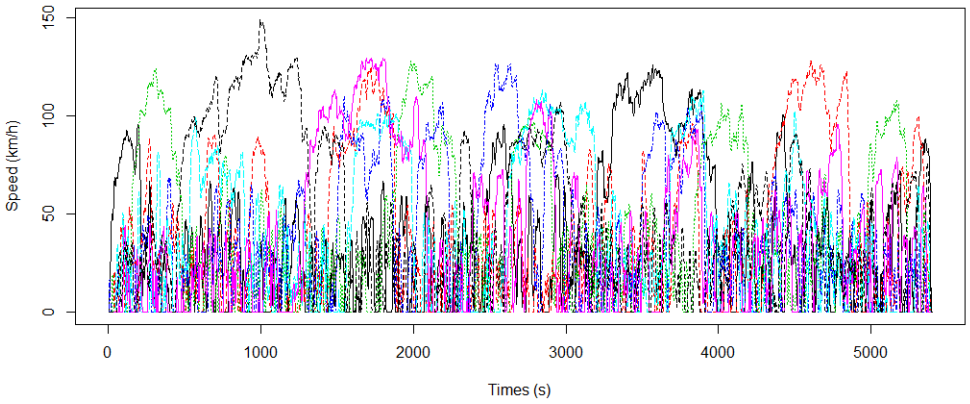

FIG. 6. *Automotive test case. Sample of 7 real driving cycles.*

In short, the ammonia emissions peak during a driving cycle is modelled as a function,

$$\begin{array}{rcccl} f & : & \mathbb{X}\times\mathcal{V} & \to & \mathbb{R} \\ & & (x,\mathrm{V}) & \mapsto & f(x,\mathrm{V}) = \max_{t\in I} \mathrm{NH}_3^{\mathtt{slip}}(t) \end{array}$$

We are interested in recovering the set $\Gamma^* = \{x \in \mathbb{X}\ ,\ g(x) = \mathbb{E}_{\mathrm{V}}[f(x,\mathrm{V})] \leq c\}$, with $c = 30ppm$. Conducting these studies with full grid simulations requires many hours of computational time, and the use of meta-models allows to tackle this computational issues.

Here we consider a Gaussian process prior $(Y_x)_{x\in\mathbb{X}} \sim GP(m,k)$, with constant mean function and Matérn covariance kernel with $\nu = 5/2$. The initial DoE consists of a 8 points LHS design optimized with respect to the maximin criterion. The covariance kernel hyper-parameters are estimated by maximizing the likelihood.

As for the analytical example, we proceed to add one point at each iteration for the SUR strategy until the stopping criterion with $(l_0,\epsilon) = (4, 5.10^{-3})$ is verified. Concerning the expectation estimation, we set the stopping criterion parameters at $(l_0,\epsilon) = (4, 10^{-2})$ and the truncation argument is set at m=20 in order to explain 80% of the variance. The algorithm was stopped at the 37-point design because the Vorob'ev deviation appears to have stabilized, in other words, the absolute error between the Vorob'ev deviations of the points 33-37 are smaller than 0.005, as shown in Figure 7. We note that for each additional point, the new observed response affects the estimation of the excursion set and its uncertainty. Thus, although the Vorob'ev deviation generally decreases, it is not a monotonic decreasing. The stopping criterion is constructed to check the stability of convergence by taking into account the last four iterations.

In searching for the true set, the SUR algorithm heavily visits the boundary region of Γ^* and allows itself to explore also potentially interesting regions (cf. Figure 8). In each added point, Figure 8 shows the number of necessary driving cycles to estimate the expectation. We remark that instead of taking the whole sample (100 driving cycles), it was sufficient to sequentially and wisely choose a reduced and representative number of driving cycles below 20. In the present case, the excursion domain Γ^* is well recovered by the algorithm. Actually, after 37 iterations (815 evaluations) the whole domain $\mathbb{X}$ has an excursion probability close to either 0 or 1.

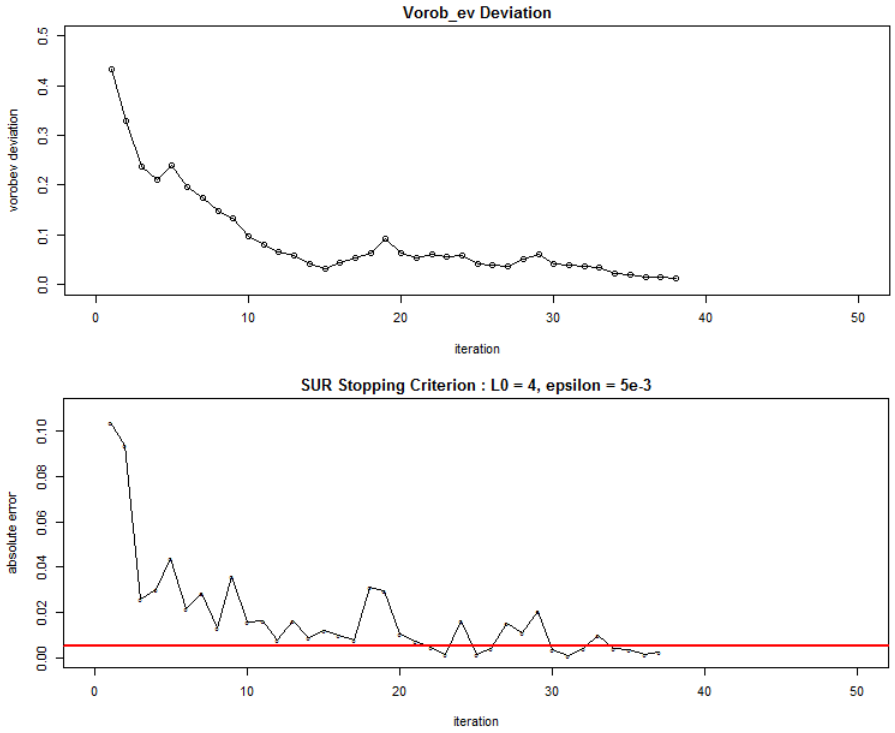

FIG. 7. *Automotive test case. Top : Decrease of the Vorobev deviation when new points are added. Bottom : Evolution of the absolute error (27) and the red line represents the stopping criterion.*

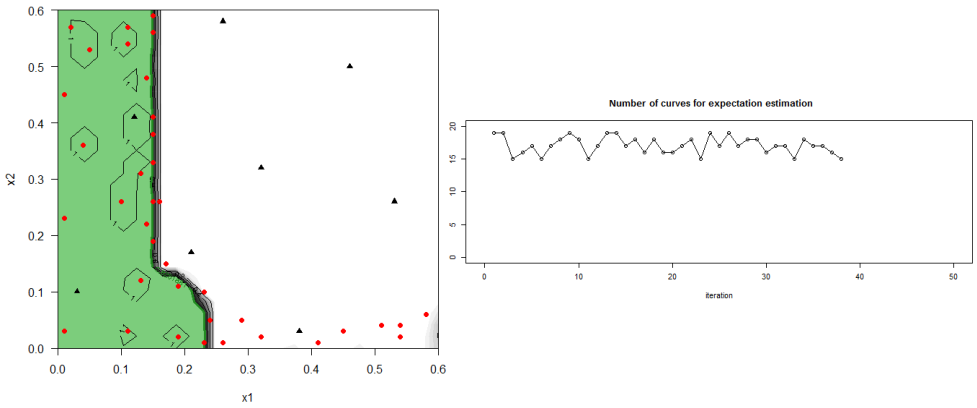

FIG. 8. *Automotive test case. Left : Coverage probability function, estimate set (green) after 37 added points and 815 function evaluations, initial DoE (black triangles), the added points (red circles). Right : needed driving cycles to estimate expectation at each added point.*

6. Conclusions. In this paper, a new method of inversion under uncertainty was proposed for problems where some of the input parameters are functional random variables with unknown probability distribution (only a sample is available). The objective is to recover the set of control variables leading to ensure some constraints by taking into account the uncertainties. The method is composed of two steps : a

sequential strategy to catch the excursion set, and the modelling of functional uncertainties. To solve the first issue a kriging model in the control input space is built. It makes possible to assess the uncertainty on the set of interest given a sample of evaluations. Then a sequential strategy (SUR) proposed by [4] and based on the kriging model is used to sequentially and efficiently choose new evaluation points to improve the excursion set estimation. For the second issue, we consider the expectation to model uncertainties and we propose a sequential approach to estimate the expectation in each point proposed by SUR. Each curve is represented by its coefficients in a truncated KL decomposition. The chosen points in the KL coefficients finite set, each one corresponding to a curve, are sequentially added and chosen to approximate a maximin space filling design. This methodology leads to an efficient estimation of the expectation. As illustrated on the application on an analytical test case with two control inputs and a functional random one. The results indicate significant enhancement in term of precision and number of calls to the simulator. We also applied this method to the automotive test case which motivated this research work. In this case, the obtained result agrees with the intuitions made from physics behind the simulator. The paper focuses on the mean of $f(x, \mathrm{V})$. In practice, other functionals of this distribution may be of great importance. For example, practitioners may be interested in ensuring a certain level of reliability, leading to consider a probabilistic constraint. The proposed method could be adapted to that case by seeing the probability as an expectation, at least for moderate risk levels.

Acknowledgments. This work was supported by IFPEN within the framework of the OQUAIDO consortium.

REFERENCES

[1] F. Anstett-Collin, J. Goffart, T. Mara, and L. Denis-Vidal, *Sensitivity analysis of complex models: coping with dynamic and static inputs*, Reliability Engineering & System Safety, 134 (2015), pp. 268–275.

[2] D. Azzimonti, J. Bect, C. Chevalier, and D. Ginsbourger, *Quantifying uncertainties on excursion sets under a gaussian random field prior*, SIAM/ASA Journal on Uncertainty Quantification, 4 (2016), pp. 850–874.

[3] J. Bect, F. Bachoc, and D. Ginsbourger, *A supermartingale approach to gaussian process based sequential design of experiments*, arXiv preprint arXiv:1608.01118, (2016).

[4] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez, *Sequential design of computer experiments for the estimation of a probability of failure*, Statistics and Computing, 22 (2012), pp. 773–793.

[5] D. Bolin and F. Lindgren, *Excursion and contour uncertainty regions for latent gaussian models*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77 (2015), pp. 85–106.

[6] A. Bonfils, Y. Creff, O. Lepreux, and N. Petit, *Closed-loop control of a scr system using a nox sensor cross-sensitive to nh3*, IFAC Proceedings Volumes, 45 (2012), pp. 738–743.

[7] C. Chevalier, *Fast uncertainty reduction strategies relying on Gaussian process models*, PhD thesis, Citeseer, 2013.

[8] C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, and Y. Richet, *Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set*, Technometrics, 56 (2014), pp. 455–465.

[9] C. Chevalier, X. Emery, and D. Ginsbourger, *Fast update of conditional simulation ensembles*, Mathematical Geosciences, 47 (2015), pp. 771–789.

[10] C. Chevalier, V. Picheny, and D. Ginsbourger, *Kriginv: An efficient and user-friendly implementation of batch-sequential inversion strategies based on kriging*, Computational statistics & data analysis, 71 (2014), pp. 1021–1034.

[11] A. Cuevas and R. Fraiman, *A plug-in approach to support estimation*, The Annals of Statistics, (1997), pp. 2300–2312.

[12] A. Cuevas, W. González-Manteiga, and A. Rodríguez-Casal, *Plug-in estimation of general level sets*, Australian & New Zealand Journal of Statistics, 48 (2006), pp. 7–19.
[13] F. Ferraty and P. Vieu, *Nonparametric functional data analysis: theory and practice*, Springer Science & Business Media, 2006.
[14] J. P. French, S. R. Sain, et al., *Spatio-temporal exceedance locations and confidence regions*, The Annals of Applied Statistics, 7 (2013), pp. 1421–1449.
[15] T. Harris and W. Yu, *Variance decompositions of nonlinear-dynamic stochastic systems*, Journal of Process Control, 20 (2010), pp. 195–205.
[16] R. Jin, W. Chen, and A. Sudjianto, *An efficient algorithm for constructing optimal design of computer experiments*, Journal of Statistical Planning and Inference, 134 (2005), pp. 268–287.
[17] M. E. Johnson, L. M. Moore, and D. Ylvisaker, *Minimax and maximin distance designs*, Journal of statistical planning and inference, 26 (1990), pp. 131–148.
[18] A. Ligmann-Zielinska and P. Jankowski, *Exploring normative scenarios of land use development decisions with an agent-based simulation laboratory*, Computers, Environment and Urban Systems, 34 (2010), pp. 409–423.
[19] L. Lilburne and S. Tarantola, *Sensitivity analysis of spatial models*, International Journal of Geographical Information Science, 23 (2009), pp. 151–168.
[20] S. Mallat, *A wavelet tour of signal processing*, Academic press, 1999.
[21] A. Marrel, N. Saint-Geours, and M. De Lozzo, *Sensitivity analysis of spatial and/or temporal phenomena*, Handbook of Uncertainty Quantification, (2016), pp. 1–31.
[22] I. Molchanov, *Theory of random sets*, Springer Science & Business Media, 2006.
[23] M. D. Morris and T. J. Mitchell, *Exploratory designs for computational experiments*, Journal of statistical planning and inference, 43 (1995), pp. 381–402.
[24] S. Nanty, C. Helbert, A. Marrel, N. Pérot, and C. Prieur, *Sampling, metamodeling, and sensitivity analysis of numerical simulators with functional stochastic inputs*, SIAM/ASA Journal on Uncertainty Quantification, 4 (2016), pp. 636–659.
[25] S. Nanty, C. Helbert, A. Marrel, N. Pérot, and C. Prieur, *Uncertainty quantification for functional dependent random variables*, Computational Statistics, 32 (2017), pp. 559–583.
[26] L. Pronzato and W. G. Müller, *Design of computer experiments: space filling and beyond*, Statistics and Computing, 22 (2012), pp. 681–701.
[27] J. O. Ramsay, *Functional data analysis*, Wiley Online Library, 2006.
[28] O. Roustant, D. Ginsbourger, and Y. Deville, *Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization*, (2012).
[29] P. Ruffo, L. Bazzana, A. Consonni, A. Corradi, A. Saltelli, and S. Tarantola, *Hydrocarbon exploration risk evaluation through uncertainty and sensitivity analyses techniques*, Reliability Engineering & System Safety, 91 (2006), pp. 1155–1162.
[30] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, *Design and analysis of computer experiments*, Statistical science, (1989), pp. 409–423.
[31] E. Vazquez and J. Bect, *A sequential bayesian algorithm to estimate a probability of failure*, IFAC Proceedings Volumes, 42 (2009), pp. 546–550.