



HAL
open science

Data-driven stochastic inversion under functional uncertainties

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

► **To cite this version:**

Mohamed Reda El Amri, Céline Helbert, Olivier Lepreux, Miguel Munoz Zuniga, Clémentine Prieur, et al.. Data-driven stochastic inversion under functional uncertainties. 2018. hal-01704189v2

HAL Id: hal-01704189

<https://inria.hal.science/hal-01704189v2>

Preprint submitted on 8 Feb 2018 (v2), last revised 2 Feb 2021 (v3)

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

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

1 **DATA-DRIVEN STOCHASTIC INVERSION UNDER FUNCTIONAL**
2 **UNCERTAINTIES***

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

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

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

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

22 **1. Introduction.** In recent years, computer models are omnipresent in engi-
23 neering and sciences, because the corresponding physical experimentation is costly or
24 impossible to execute. Numerical models are adopted to study the behaviour of such
25 physical phenomena. As underlined in [4, 7], practitioners are not only interested in
26 the response of their model for a given set of inputs (forward problem) but also in
27 recovering the set of input values leading to a prescribed value or range for the output
28 of interest. The problem of estimating such set is called hereafter inversion problem.
29 We will consider a system that evolves in an uncertain environment, the uncertain-
30 ties appear for example due to manufacturing tolerances or environmental conditions.
31 The numerical simulator modelling the system, denoted f , takes two types of input
32 variables : a set of control variables $x \in \mathbb{X}$, and a set of uncertain variables $v \in \mathcal{V}$.
33 Robust inversion consists in seeking the set of control variables $x \in \mathbb{X}$ such that
34 $\sup_{v \in \mathcal{V}} f(x, v)$ is smaller than a threshold c . Then, the difficulty of solving the ro-
35 bust inversion problem strongly depends on the uncertainty set \mathcal{V} . In our setting,
36 \mathcal{V} is a functional space, and we consider the inversion problem under uncertainty as
37 a stochastic inversion problem. It means that we assume that the uncertainty has
38 a probabilistic description. Let V denote the random variable, valued in \mathcal{V} , mod-
39 elling the uncertainty. In our framework, we are interested in constructing the set :
40 $\Gamma^* := \{x \in \mathbb{X}, \mathbb{E}_V[f(x, V)] \leq c\}$, 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).

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

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

46 Inverse problems have already been carried out in many applications, notably reli-
 47 ability engineering (see, e.g., [4], [7]), climatology (see, e.g., [5], [14]) and many other
 48 fields. In the literature, one way to solve the problem is to adopt a sequential sampling
 49 strategy based on Gaussian process emulators. The idea is that gaussian process em-
 50 ulators, which capture prior knowledge about the regularity of the unknown function
 51 $g : x \mapsto \mathbb{E}_V[f(x, V)]$, make it possible to assess the uncertainty about Γ^* given a set
 52 of evaluations of g . More specifically, these sequential strategies for the estimation of
 53 an excursion set are closely related to the field of Bayesian global optimization (see,
 54 e.g., [7]). In the case of inversion problems, Stepwise Uncertainty Reduction (SUR)
 55 strategies based on set measures were introduced in [31]. More recently, a parallel im-
 56 plementation of these strategies have been proposed in [8] and applied to the problem
 57 of construction of an excursion set. Briefly, the strategy SUR gives sequentially the
 58 next location where to evaluate the function g in order to minimize an uncertainty
 59 function. The key contribution of the present paper is to propose a data-driven adap-
 60 tation of that procedure in the presence of functional uncertainties.

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

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

$$77 \quad (1) \quad \begin{aligned} \Gamma^* &:= \{x \in \mathbb{X}, \mathbb{E}_V[f(x, V)] \in C\} \\ &:= \{x \in \mathbb{X}, g(x) \in C\} \quad \text{with } C = (\infty, c], c \in \mathbb{R}. \end{aligned}$$

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

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

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

88 quantially expanded by adding new goal oriented points. This procedure is introduced
 89 below.

90 **2.1. Random closed set and bayesian framework.** In a Bayesian frame-
 91 work, we assume that g is a realization of an almost surely continuous Gaussian
 92 process $Y \sim GP(m, k)$ with a mean structure m , defined as, $m(x) = \mathbb{E}[Y_x]$, $x \in \mathbb{X}$,
 93 and a covariance kernel k , defined as, $k(x, x') := Cov(Y_x, Y_{x'})$, $x, x' \in \mathbb{X}$. Due to the
 94 stochastic nature of $(Y_x)_{x \in \mathbb{X}}$, the associated excursion set,

$$95 \quad (2) \quad \Gamma := \{x \in \mathbb{X}, Y_x \in C\}$$

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

103 **2.2. Vorob'ev approach.** Γ defines a random closed set. Several ways to define
 104 the expectation of a random set have been developed in the literature (see e.g. [\[22\]](#)).
 105 In this paper, we focus on Vorob'ev expectation, closely related to quantiles and level-
 106 sets which have been the subject of many developments. Let us define *the coverage*
 107 *probability function* of a random set Γ as

$$108 \quad (3) \quad p(x) = \mathbb{P}(x \in \Gamma), \quad x \in \mathbb{X}.$$

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

$$114 \quad (4) \quad Q_\alpha = \{x \in \mathbb{X} : p(x) \geq \alpha\}.$$

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

$$118 \quad (5) \quad \forall \beta > \alpha^*, \quad \mu(Q_\beta) < \mathbb{E}[\mu(\Gamma)] \leq \mu(Q_{\alpha^*}).$$

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

$$122 \quad (6) \quad d_\mu(A, B) = \mathbb{E}[\mu(A \Delta B)],$$

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

125
 126 In the following, we use the Vorob'ev expectation and deviation to quantify the
 127 variability of Γ conditionally to available observations. Let us denote the initial
 128 design points as $\mathcal{X}_n = (x_1, x_2, \dots, x_n) \in \mathbb{X}^n$ and the responses at these points as

129 $\mathbf{g}_{\mathcal{X}_n} = (g(x_1), g(x_2), \dots, g(x_n)) \in \mathbb{R}^n$. We note $Y_{\mathcal{X}_n} = (Y_{x_1}, Y_{x_2}, \dots, Y_{x_n})$ the random
 130 vector associated to the random process Y considered at \mathcal{X}_n . The main object of
 131 interest is then the conditional probability distribution of the random closed set Γ
 132 given the n observations. We know that the posterior distribution of the process Y
 133 given the n available observations remains Gaussian and is characterized by the pos-
 134 terior mean $m_n(x) = \mathbb{E}[Y_x \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n}]$, $x \in \mathbb{X}$, and the posterior covariance kernel,
 135 $k_n(x, x') := \text{Cov}(Y_x, Y_{x'} \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n})$. The *coverage probability function* (3) and
 136 *Vorob'ev Quantile* (4) given the n observations can easily be computed as follows :

$$137 \quad (7) \quad p_n(x) = \mathbb{P}(x \in \Gamma \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n}) = \mathbb{P}(Y_x \leq c \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n}) = \Phi\left(\frac{c - m_n(x)}{\sqrt{k_n(x, x)}}\right),$$

$$Q_{n, \alpha} = \{x \in \mathbb{X} : p_n(x) \geq \alpha\} = \{x \in \mathbb{X} : m_n(x) + \Phi^{-1}(\alpha)\sqrt{k_n(x, x)} \leq c\},$$

138 where $\Phi(\cdot)$ denotes the cumulated distribution function (c.d.f.) of the standard
 139 Gaussian distribution. The Vorob'ev expectation Q_{n, α_n^*} can be determined by tun-
 140 ing α to a level α_n^* such that $\mu(Q_{n, \alpha_n^*}) = \mathbb{E}[\mu(\Gamma) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n}]$, knowing that
 141 $\mathbb{E}[\mu(\Gamma) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n}] = \int_{\mathbb{X}} p_n(x) \mu(dx)$, this can be done through simple dichotomy or
 142 more advanced technique as, e.g., brent's method.
 143 Once the Vorob'ev expectation is determined, the computation of Vorob'ev deviation
 144 $\mathbb{E}[\mu(\Gamma \Delta Q_{n, \alpha_n^*}) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n}]$ can be expressed as a function of the posterior coverage
 145 probability function p_n and does not require simulations of Γ . Indeed,

$$146 \quad (8) \quad \begin{aligned} \mathbb{E}_n[\mu(\Gamma \Delta Q_{n, \alpha_n^*})] &= \mathbb{E}_n\left[\mu(\Gamma \cap Q_{n, \alpha_n^*}^c) + \mu(Q_{n, \alpha_n^*} \cap \Gamma^c)\right] \\ &= \mathbb{E}_n\left[\int_{\mathbb{X}} 1_{\{x \in \Gamma, x \notin Q_{n, \alpha_n^*}\}} + 1_{\{x \notin \Gamma, x \in Q_{n, \alpha_n^*}\}} \mu(dx)\right] \\ &= \int_{Q_{n, \alpha_n^*}^c} \mathbb{E}_n[1_{\{x \in \Gamma\}}] \mu(dx) + \int_{Q_{n, \alpha_n^*}} \mathbb{E}_n[1_{\{x \notin \Gamma\}}] \mu(dx) \\ &= \int_{Q_{n, \alpha_n^*}^c} p_n(x) \mu(dx) + \int_{Q_{n, \alpha_n^*}} (1 - p_n(x)) \mu(dx), \end{aligned}$$

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

153 **2.3. SUR strategies.** The principle of stepwise uncertainty reduction (SUR)
 154 (see, e.g., [4]; [8]) is to define an uncertainty measure, depending on the objective to
 155 be fulfilled, and to sequentially choose the points that decrease most this uncertainty.
 156 In other words, the aim of the SUR strategy is to construct a sequence of evaluation
 157 locations in order to reduce the *expected* uncertainty on a quantity of interest.
 158 Here, we work in the particular setting where g is a sample path of a random process
 159 Y . The uncertainty function for an estimate of Γ is defined as a function $\mathcal{H}^{\text{uncert}}$ that
 160 associates to any finite sequence of observations $(\mathcal{X}_n, \mathbf{g}_{\mathcal{X}_n})$ a real value representing
 161 the uncertainty on the estimation of Γ . When n observations are available, we denote
 162 by $\mathcal{H}_n^{\text{uncert}}$ the uncertainty at step n . We assume that we have r evaluations left. The
 163 objective of the SUR strategy is to find r optimal locations x_{n+1}, \dots, x_{n+r} such that
 164 the uncertainty $\mathcal{H}_{n+r}^{\text{uncert}}$ is as small as possible.

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

$$167 \quad (9) \quad \mathcal{H}_n^{\text{uncert}} = \mathbb{E}[\mu(\Gamma \Delta Q_{n, \alpha_n^*}) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n}]$$

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

$$170 \quad (10) \quad \mathcal{H}_{n+1}^{\text{uncert}}(x) = \mathbb{E}[\mu(\Gamma \Delta Q_{n+1, \alpha_n^*}) \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n}, Y_x]$$

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

$$174 \quad (11) \quad \begin{aligned} x_{n+1} &= \operatorname{argmin}_{x \in \mathbb{X}} \mathbb{E}_{n,x}[\mathcal{H}_{n+1}^{\text{uncert}}(x)] \\ &= \operatorname{argmin}_{x \in \mathbb{X}} \mathcal{J}_n(x), \end{aligned}$$

175 where $\mathbb{E}_{n,x}$ denotes the expectation with respect to $Y_x \mid Y_{\mathcal{X}_n} = \mathbf{g}_{\mathcal{X}_n}$.
 176 After having evaluated the function g at the optimal location x_{n+1} , we update the
 177 parameters of the posterior mean and covariance, and we restart until the evaluation
 178 budget r is exhausted. Such strategy is called *one-step lookahead*, which means that
 179 we select the next evaluation point as if it were the last one.

180 In practice, it is possible to run the model evaluations in parallel. Thus, we aim
 181 at finding an optimal **batch** of q points to reduce the uncertainty. Such strategy
 182 is called *batch sequential one-step lookahead*. In this setting, the sampling criterion
 183 \mathcal{J}_n is the expected uncertainty at next step assuming the batch of q points $\mathbf{x} =$
 184 $(x^{(1)}, x^{(2)}, \dots, x^{(q)}) \in \mathbb{X}^q$ is evaluated.

$$185 \quad (12) \quad \mathcal{J}_n(\mathbf{x}) = \mathbb{E}_{n,\mathbf{x}}[\mathcal{H}_{n+q}^{\text{uncert}}(\mathbf{x})],$$

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

$$188 \quad (13) \quad (x_{n+1}, x_{n+2}, \dots, x_{n+q}) = \operatorname{argmin}_{\mathbf{x} \in \mathbb{X}^q} \mathcal{J}_n(\mathbf{x}).$$

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

$$192 \quad (14) \quad \begin{aligned} \mathcal{J}_n(\mathbf{x}) &= \int_{\mathbb{X}} \left(2\Phi_2 \left(\left(\left[\Phi^{-1}(\alpha_n^*) - a_{n+q}(u) \right]; \begin{bmatrix} 1 + \gamma_{n+q}(u) & -\gamma_{n+q}(u) \\ -\gamma_{n+q}(u) & \gamma_{n+q}(u) \end{bmatrix} \right) \right) \right. \\ &\quad \left. + p_n(u) + \Phi \left(\frac{a_{n+q}(u) - \Phi^{-1}(\alpha_n^*)}{\sqrt{\gamma_{n+q}(u)}} \right) \right) d\mu(u) \end{aligned}$$

193 where

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

194

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

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

199

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

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

214 **3.1. Dimension reduction and density estimation (Fpca).** Let V_t be a
 215 real-valued random process, which is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and
 216 indexed on the bounded interval $I \subset \mathbb{R}$. Without loss of generality, we will consider a
 217 centred stochastic process with finite variance. In this work, the stochastic process V_t
 218 is observable through a finite sample of N realizations. In this context, uncertainty
 219 quantification methodology followed by most of authors (e.g. [27], [13]) can be sepa-
 220 rated into two steps, hereafter called the Fpca method.

221 First, the variable is decomposed on a truncated basis of functions. There are dif-
 222 ferent prominent examples of fixed functional basis, we can cite B-Spline basis or
 223 wavelet basis (see [20] for more details about wavelets and wavelet transforms). The
 224 Karhunen-Loève (KL) expansion is another way of representing a stochastic process.
 225 It is based on the spectral expansion of the covariance function of the process. This
 226 covariance function, which can be viewed as a kernel, is defined as :

$$227 \quad (15) \quad C(t, s) = \text{Cov}(V_t, V_s) = \mathbb{E}[V_t V_s],$$

228 and the corresponding integral operator is :

$$229 \quad (16) \quad \mathcal{T}_C \zeta(t) = \int_I C(t, s) \zeta(s) ds.$$

230 We suppose that the covariance function is continuous in t, s . By Mercer's theorem,
 231 the corresponding integral operator \mathcal{T}_C has an orthonormal basis of eigenfunctions
 232 $\{\psi_i(t)\}$. We define :

$$233 \quad (17) \quad \eta_i = \int_I V_t \psi_i(t) dt,$$

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

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

$$245 \quad (18) \quad V_t \simeq \sum_{i=1}^m \eta_i \psi_i(t),$$

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

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

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

260 **3.2. Scenario modelling.** Another method, suggested by Lilburne et al. [19],
 261 is to represent the uncertainty of the functional variable by directly sampling from the
 262 set of N realizations of V , denoted by $\Xi = \{v_1, \dots, v_N\}$. The functional probability
 263 distribution is approximated by a uniform discrete distribution over Ξ . This method
 264 was applied in [19] to calculate the first and total-order sensitivity indices for spatial
 265 models for simulating nitrate transport from paddock to groundwater. Ruffo et al.
 266 [29] also used this approach to perform sensitivity analysis of a model for oil reservoir
 267 production forecasting. This approach is likewise used in many other fields (see, e.g.,
 268 [15]; [18]; [1]; [21]).

269 **3.3. Weighted scenario modelling.** The two approaches, described previ-
 270 ously, suffer from various pathologies. The first strategy [Subsection 3.1] depends
 271 on the approach used for the estimation of the KL coefficients distribution : non-
 272 parametric (distribution-free) model which makes no assumption about the probabil-
 273 ity distribution, or parametric model like GMM. To preserve the underlying curve
 274 structure, one should choose a large value for the truncation argument. However,
 275 parametric or non-parametric approaches for density estimation suffer from the curse
 276 of dimensionality. The second strategy [Subsection 3.2] has the merit of keeping the
 277 shape of curves, but the realizations are randomly selected from the initial set Ξ .
 278 However, sampling a few representative curves among Ξ is a difficult task. This is
 279 the reason why we propose here an optimal space filling strategy instead of the crude
 280 method of Lilbrune et al. [19]. Our aim is to explain at best the variability of V
 281 with a reduced sample of realizations. This approach can be related to randomized

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

$$286 \quad (19) \quad \Xi = \{v_{ij}\}_{i=1:N}^{j=1:f} \longrightarrow \mathcal{G} = \{\eta_i\}_{i=1:N}$$

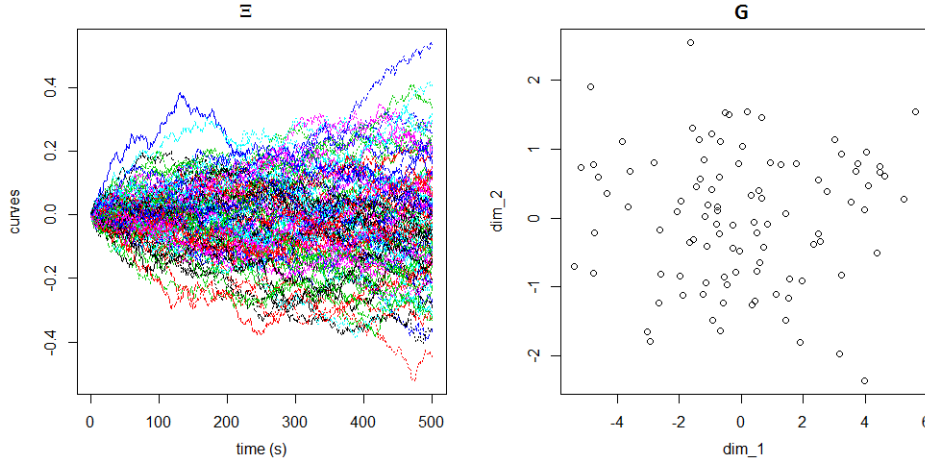


FIG. 1. Ξ the sample of 100 realizations of V (left) and their representatives in the space of coefficients η (right), where $m = 2$.

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

295 **3.3.1. Space-Filling Design and quality criterion.** Let us define by $D =$
 296 $\{d_1, \dots, d_l\}$ a collection of l points. We denote

$$297 \quad (20) \quad \text{dist}_{ij} = \|d_i - d_j\|,$$

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

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

$$309 \quad (21) \quad \max \min_{i \neq j} \|\eta_i - \eta_j\|.$$

310 **3.3.2. Sequential Space-filling Design.** Finding the design D is a compu-
 311 tationally difficult problem. We could adapt the optimal design algorithms used in
 312 the literature such as simulated annealing (see Morris et al. [23]) and stochastic evo-
 313 lutionary algorithm (see Jin et al. [16]) for our purpose. Different from theirs, we
 314 propose a one-point-at-time greedy algorithm for the generation of our design. The
 315 sequential construction is described below

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

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

$$322 \quad (23) \quad \forall l \geq 2, \quad D_l = \{d_i\}_{i=1:l} \longrightarrow \tilde{D}_l = \{v_{(i)}\}_{i=1:l}$$

323 **Figure 2** shows a construction of 10 points design applied to the example of **Figure 1**.
 324 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.

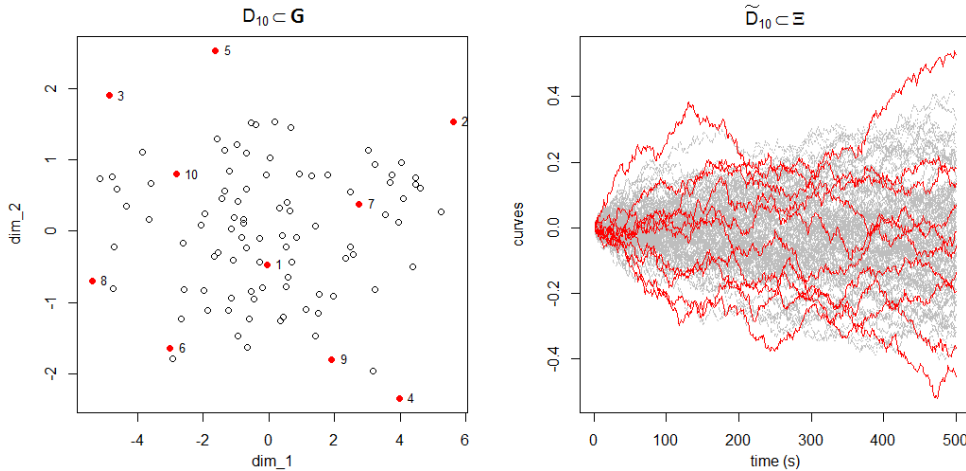


FIG. 2. Sequential design of 10 points (red points) in the space of the coefficients \mathcal{G} (left) and their corresponding red curves in Ξ (right).

325

326 **3.3.3. Weight calculus.** The proposed method produces a weighted mean that
 327 has less variability than the arithmetic mean of a simple random sample. To do so,
 328 one must calculate the weight associated to each selected curve. We propose to use
 329 the concept of Voronoi cell: we associate to each design point a region consisting of all

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

$$331 \quad (24) \quad \forall l \geq 2, \forall i \in \{1, \dots, l\}, w_i^{(l)} = |S_{li}| \text{ and } |\cdot| \text{ denotes the cardinality of set} \\ S_{li} = \{\eta \in \mathcal{G} \mid \forall h \in \{1, \dots, l\}, d(\eta, d_i) \leq d(\eta, d_h)\}$$

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

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

$$345 \quad (25) \quad e_l(x_{n+1}) = |\hat{m}_{l-1}(x_{n+1}) - \hat{m}_l(x_{n+1})|$$

346 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 $|\cdot|$ the absolute value
347 function. In the following, the *stopping criterion* must satisfy the condition,

$$348 \quad (26) \quad e_{l-l_0}(x_{n+1}) \leq e_{l-l_0+1}(x_{n+1}) \leq \dots \leq e_l(x_{n+1}) \leq \epsilon$$

349 it ensures that the quantities e_l are smaller than a tolerance ϵ on the l_0 previous steps
350 in the estimation. These two parameters are set by practitioners. It allows to use
351 fewer curves without sacrificing estimation accuracy.

352 The strategy SUR could be stopped if the allocated number of simulations is reached.
353 However, we define in this work an additional stopping criterion based on the Vorob'ev
354 deviation and close to the one defined for the expectation estimate. Thus, the strategy
355 is carried out until the following stopping criterion is verified

$$356 \quad (27) \quad e_{l-l_0}^{\text{SUR}} \leq e_{l-l_0+1}^{\text{SUR}} \leq \dots \leq e_l^{\text{SUR}} \leq \epsilon,$$

357 where $e_i^{\text{SUR}} = |\mathbb{E}_{i-1}[\mu(\Gamma \Delta Q_{i-1, \alpha_{i-1}^*})] - \mathbb{E}_i[\mu(\Gamma \Delta Q_{i, \alpha_i^*})]|$ is the absolute error between
358 two successive Vorob'ev deviations. The condition (27) tests if all the quantities are
359 smaller than a tolerance ϵ on l_0 consecutive steps.

360 The global methodology to perform robust inversion in presence of functional uncer-
361 tainty proposed in this paper is summarized in [Algorithm 1](#).

Algorithm 1

```

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

```

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

369 **5.1. Analytical example.** In this example, we define the function f as follows:
(28)

$$370 f : (x, V) \mapsto |0.1 \cos(x_1 \max_{t \in T} V_t) \sin(x_2) \cdot (x_1 + x_2 \min_{t \in T} V_t)^2| \cdot \int_T (30 + V_t)^{\frac{x_1 \cdot x_2}{20}} dt \cdot \max_{t \in T} V_t$$

371 where the control variable x lies in $\mathbb{X} = [1.5, 5] \times [3.5, 5]$, and V is a standard
372 Brownian motion with state space in \mathbb{R} and $T = [0, 1]$. We suppose that a sample of
373 N realizations of V is available, denoted by Ξ , and these realizations are discretized
374 uniformly on 100 points of T . The objective is to construct the set $\Gamma^* := \{x \in$
375 $\mathbb{X}, g(x) = \mathbb{E}_V[f(x, V)] \leq c\}$, where $c = 1.2$.

376 Here we consider a Gaussian process prior $(Y_x)_{x \in \mathbb{X}} \sim GP(m, k)$, with constant mean
377 function and Matérn covariance kernel with $\nu = 5/2$. The initial DoE consists of a
378 9 points LHS design optimized by maximin criterion. The covariance kernel hyper-
379 parameters are estimated by Maximum Likelihood Estimation (MLE). Figure 3 shows
380 the initial design of experiments and the target set Γ^* obtained from a 30×30 grid
381 experiment, where at each grid point the expectation is approximated by a Monte
382 Carlo Method over 5000 realizations of V . We aim at estimating the set Γ^* using the
383 SUR strategy to choose the next evaluation point as defined in Section 2, and the
384 methods presented in Section 3 to provide an estimation of the expectation.

385 We proceed to add one point at each iteration of the SUR strategy until the
386 condition (27) for $(l_0, \epsilon) = (4, 5 \cdot 10^{-3})$ is reached. The covariance parameters are re-
387 estimated at each step by MLE. Since this criterion is based on the Vorob'ev deviation,
388 the objective is to reduce the uncertainty on the estimate until stability. For the
389 sequential estimation of the expectation, we test the sensitivity to the parameters
390 (l_0, ϵ) of criterion (26) (see Table 1).

391 The estimation of the expectation at the proposed point by SUR is carried out with
392 the 3 methods (Fpca, Scenario, W.Scenario) detailed in Section 3. As presented in
393 Section 4, the estimation is done sequentially and it depends on the *stopping criteria*

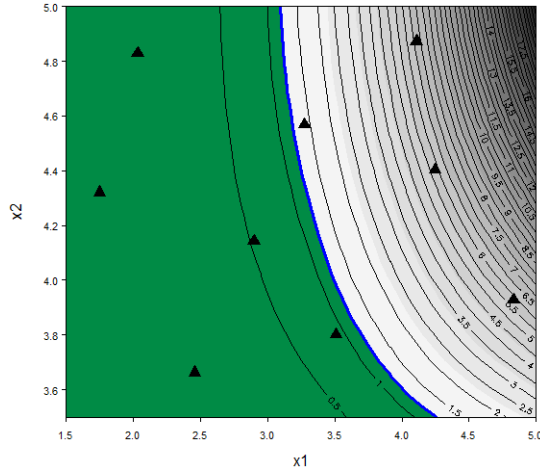


FIG. 3. Analytical example. Contour plot of the function g , the set of interest (green) with boundary (blue line), the initial design of experiments (black triangles).

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

l_0	4	2	3	4
ϵ	10^{-2}	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$

TABLE 1

Analytical example. Estimation of expectation stopping criteria parameters

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

(l_0, ϵ)	$\mu(\Gamma^* \Delta Q_{n_{\text{last}}, \alpha_{n_{\text{last}}}^*}) / \mu(\Gamma^*)$			Number of calls to f		
	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)

412 **Table 2** indicates that the three methods are sensitive to the parameters l_0 and
 413 ϵ . Bigger is the parameter l_0 , i.e., one seeks a stability of the estimation, smaller is
 414 the error but higher is the number of calls to the function. The weighted scenario
 415 method performs well both in terms of error and number of calls to the function f .
 416 The cumulative number of calls to f has been improved by a factor greater than 4 in
 417 comparison with the two other methods, even if in term of error the three methods
 418 remain close and of reasonable quality.
 419

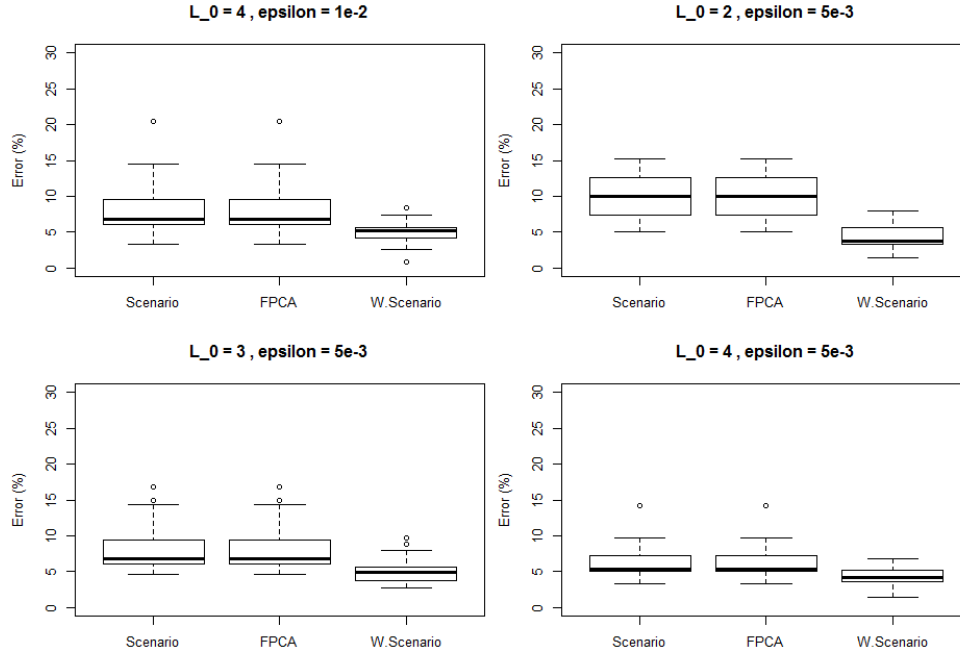


FIG. 4. Analytical example. Boxplots of the error estimation for different values l_0 , ϵ and methods (30 runs for each boxplot)

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

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

429

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

	$\mu(\Gamma^* \Delta Q_{n_{\text{last}}, \alpha_{n_{\text{last}}}^*}) / \mu(\Gamma^*)$			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)*

435 We note that for a larger sample size, the recovering error is smaller. This can
 436 be explained by the fact that with a large sample size, the available information on
 437 variable V enables an effective estimation of the expectation.

438 We know that the Weighted Scenario and the probabilistic modelling (Fpca) depend
 439 on the truncation argument. To better understand the effect of the number of dimen-
 440 sions m , we fix the stopping criteria for the SUR strategy and expectation estimation,
 441 and we consider different values of $m = \{2, 3, 4, 5, 6\}$. Each cell of **Table 4** represents
 442 the result averaged over 14×20 independent runs. For each m , we generate 14 samples
 443 Ξ of size $N=200$, and for each of them we perform 20 runs of each method.

	$\mu(\Gamma^* \Delta Q_{n_{\text{last}}, \alpha_{n_{\text{last}}}^*}) / \mu(\Gamma^*)$		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)*

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

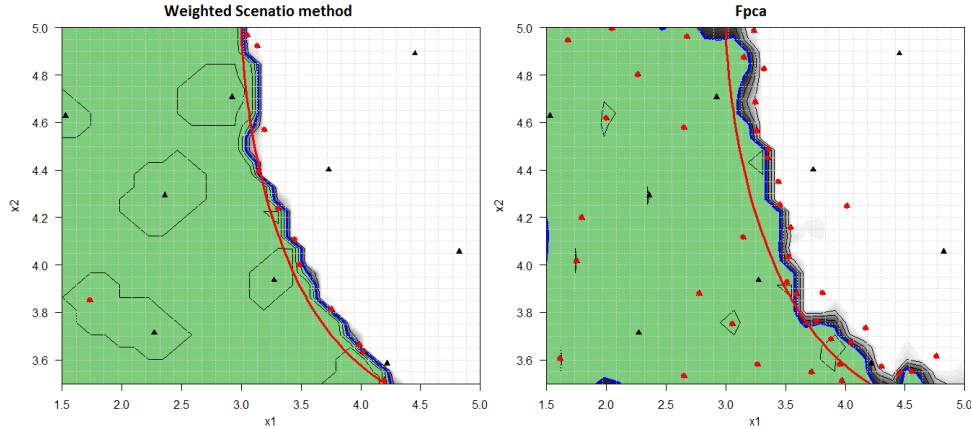


FIG. 5. Analytical example. Results at the last iteration in the case $m = 4$. Coverage function, boundary of the true set (red), estimate set (green). The initial design of experiments (black triangles), the added points (red circles)

452 However the errors in Table 4 seem to be bounded below. To go below that bound,
 453 we probably need to increase the size of Ξ . Figure 5 shows the coverage functions
 454 of the random set Γ obtained at the last iteration for the two methods and there
 455 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

456 **5.2. IFPEN test case.** In this section we test the proposed method on a au-
 457 tomotive test case from the IFPEN. The problem concerns an after-treatment device
 458 of diesel vehicles, called Selective Catalytic Reduction (SCR). This latter consists on
 459 a basic process of chemical reduction of nitrogen oxides (NO_x) to diatomic nitrogen
 460 (N_2) and water (H_2O) by the reaction of NO_x and ammonia NH_3 . The reaction it-
 461 self occurs in the SCR catalyst. Ammonia is provided by a liquid-reductant agent
 462 injected upstream of the SCR catalyst. The amount of ammonia introduced into the
 463 reactor is a critical quantity: overdosing causes undesirable ammonia slip downstream
 464 of the catalyst, whereas under-dosing causes insufficient NO_x reduction. In practice,
 465 ammonia slip is restricted to a prescribed threshold.
 466 We use an emission-oriented simulator developed by IFPEN, which models the vehi-
 467 cle, its engine and the exhaust after-treatment system. It takes the vehicle driving
 468 cycle profile as input and provides the time-series of corresponding exhaust emissions
 469 as output. A realistic SCR control law is used in this simulator. See [6] for an example
 470 of such a control law. In this study, we choose two control variables as input and a
 471 functional one considered as random. The control variables are parameters of the SCR
 472 control law. They set the targeted level of NH_3 storage in the catalyst and then are
 473 indirectly related to the NH_3 injected. They lie in $\mathbb{X} = [0, 0.6]^2$. The functional ran-
 474 dom variable describes the evolution of vehicle speed on I , with $I = [0, 5400s]$. These
 475 functional uncertainties come from an available sample of 100 real driving cycles. A

476 subset of that sample is represented in [Figure 6](#).

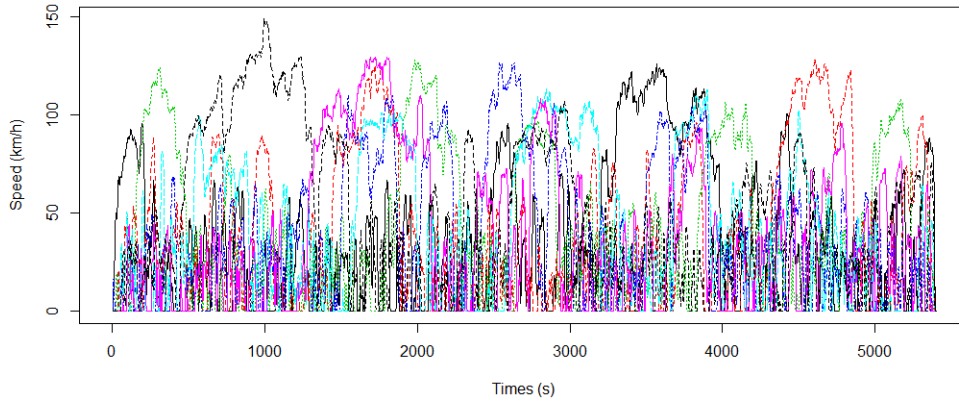


FIG. 6. Automotive test case. Sample of 7 real driving cycles.

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

$$479 \quad \begin{aligned} f &: \mathbb{X} \times \mathcal{V} \rightarrow \mathbb{R} \\ (x, V) &\mapsto f(x, V) = \max_{t \in I} \text{NH}_3^{\text{slip}}(t) \end{aligned}$$

480 We are interested in recovering the set $\Gamma^* = \{x \in \mathbb{X}, g(x) = \mathbb{E}_V[f(x, V)] \leq$
481 $c\}$, with $c = 30\text{ppm}$. Conducting these studies with full grid simulations requires
482 many hours of computational time, and the use of meta-models allows to tackle this
483 computational issues.

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

488 As for the analytical example, we proceed to add one point at each iteration for
489 the SUR strategy until the stopping criterion with $(l_0, \epsilon) = (4, 5 \cdot 10^{-3})$ is verified.
490 Concerning the expectation estimation, we set the stopping criterion parameters at
491 $(l_0, \epsilon) = (4, 10^{-2})$ and the truncation argument is set at $m=20$ in order to explain
492 80% of the variance. The algorithm was stopped at the 37-point design because
493 the Vorob'ev deviation appears to have stabilized, in other words, the absolute error
494 between the Vorob'ev deviations of the points 33-37 are smaller than 0.005, as shown
495 in [Figure 7](#). We note that for each additional point, the new observed response affects
496 the estimation of the excursion set and its uncertainty. Thus, although the Vorob'ev
497 deviation generally decreases, it is not a monotonic decreasing. The stopping criterion
498 is constructed to check the stability of convergence by taking into account the last
499 four iterations.

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

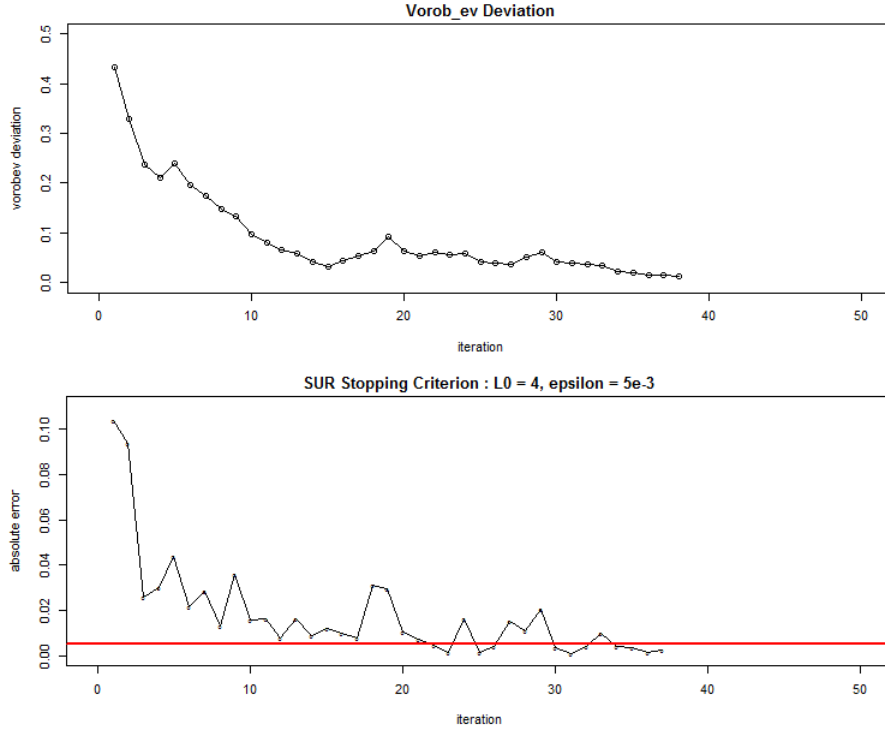


FIG. 7. Automotive test case. Top : Decrease of the Vorobev deviation when new points are added. Bottom : Evolution of the absolute error (27) and the red line represents the stopping criterion.

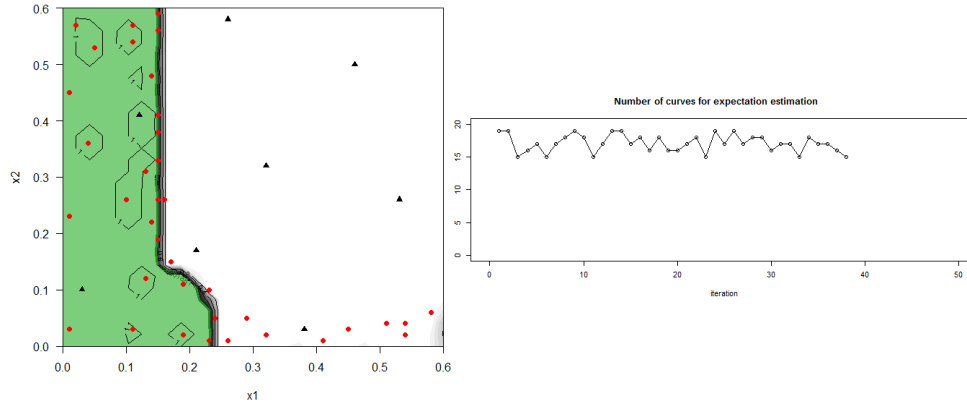


FIG. 8. Automotive test case. Left : Coverage probability function, estimate set (green) after 37 added points and 815 function evaluations, initial DoE (black triangles), the added points (red circles). Right : needed driving cycles to estimate expectation at each added point.

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

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

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

537

REFERENCES

- 538 [1] F. ANSTETT-COLLIN, J. GOFFART, T. MARA, AND L. DENIS-VIDAL, *Sensitivity analysis of*
 539 *complex models: coping with dynamic and static inputs*, Reliability Engineering & System
 540 Safety, 134 (2015), pp. 268–275.
- 541 [2] D. AZZIMONTI, J. BECT, C. CHEVALIER, AND D. GINSBOURGER, *Quantifying uncertainties on*
 542 *excursion sets under a gaussian random field prior*, SIAM/ASA Journal on Uncertainty
 543 Quantification, 4 (2016), pp. 850–874.
- 544 [3] J. BECT, F. BACHOC, AND D. GINSBOURGER, *A supermartingale approach to gaussian process*
 545 *based sequential design of experiments*, arXiv preprint arXiv:1608.01118, (2016).
- 546 [4] J. BECT, D. GINSBOURGER, L. LI, V. PICHENY, AND E. VAZQUEZ, *Sequential design of computer*
 547 *experiments for the estimation of a probability of failure*, Statistics and Computing, 22
 548 (2012), pp. 773–793.
- 549 [5] D. BOLIN AND F. LINDGREN, *Excursion and contour uncertainty regions for latent gaussian*
 550 *models*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77
 551 (2015), pp. 85–106.
- 552 [6] A. BONFILS, Y. CREFF, O. LEPREUX, AND N. PETIT, *Closed-loop control of a scr system using*
 553 *a nox sensor cross-sensitive to nh3*, IFAC Proceedings Volumes, 45 (2012), pp. 738–743.
- 554 [7] C. CHEVALIER, *Fast uncertainty reduction strategies relying on Gaussian process models*, PhD
 555 thesis, Citeseer, 2013.
- 556 [8] C. CHEVALIER, J. BECT, D. GINSBOURGER, E. VAZQUEZ, V. PICHENY, AND Y. RICHET, *Fast*
 557 *parallel kriging-based stepwise uncertainty reduction with application to the identification*
 558 *of an excursion set*, Technometrics, 56 (2014), pp. 455–465.
- 559 [9] C. CHEVALIER, X. EMERY, AND D. GINSBOURGER, *Fast update of conditional simulation en-*
 560 *sembles*, Mathematical Geosciences, 47 (2015), pp. 771–789.
- 561 [10] C. CHEVALIER, V. PICHENY, AND D. GINSBOURGER, *Kriginv: An efficient and user-friendly*
 562 *implementation of batch-sequential inversion strategies based on kriging*, Computational
 563 statistics & data analysis, 71 (2014), pp. 1021–1034.
- 564 [11] A. CUEVAS AND R. FRAIMAN, *A plug-in approach to support estimation*, The Annals of Statis-
 565 tics, (1997), pp. 2300–2312.

- 566 [12] A. CUEVAS, W. GONZÁLEZ-MANTEIGA, AND A. RODRÍGUEZ-CASAL, *Plug-in estimation of gen-*
567 *eral level sets*, Australian & New Zealand Journal of Statistics, 48 (2006), pp. 7–19.
- 568 [13] F. FERRATY AND P. VIEU, *Nonparametric functional data analysis: theory and practice*,
569 Springer Science & Business Media, 2006.
- 570 [14] J. P. FRENCH, S. R. SAIN, ET AL., *Spatio-temporal exceedance locations and confidence regions*,
571 The Annals of Applied Statistics, 7 (2013), pp. 1421–1449.
- 572 [15] T. HARRIS AND W. YU, *Variance decompositions of nonlinear-dynamic stochastic systems*,
573 Journal of Process Control, 20 (2010), pp. 195–205.
- 574 [16] R. JIN, W. CHEN, AND A. SUDJANTO, *An efficient algorithm for constructing optimal design of*
575 *computer experiments*, Journal of Statistical Planning and Inference, 134 (2005), pp. 268–
576 287.
- 577 [17] M. E. JOHNSON, L. M. MOORE, AND D. YLVIKAKER, *Minimax and maximin distance designs*,
578 Journal of statistical planning and inference, 26 (1990), pp. 131–148.
- 579 [18] A. LIGMANN-ZIELINSKA AND P. JANKOWSKI, *Exploring normative scenarios of land use devel-*
580 *opment decisions with an agent-based simulation laboratory*, Computers, Environment and
581 Urban Systems, 34 (2010), pp. 409–423.
- 582 [19] L. LILBURNE AND S. TARANTOLA, *Sensitivity analysis of spatial models*, International Journal
583 of Geographical Information Science, 23 (2009), pp. 151–168.
- 584 [20] S. MALLAT, *A wavelet tour of signal processing*, Academic press, 1999.
- 585 [21] A. MARREL, N. SAINT-GEOURS, AND M. DE LOZZO, *Sensitivity analysis of spatial and/or*
586 *temporal phenomena*, Handbook of Uncertainty Quantification, (2016), pp. 1–31.
- 587 [22] I. MOLCHANOV, *Theory of random sets*, Springer Science & Business Media, 2006.
- 588 [23] M. D. MORRIS AND T. J. MITCHELL, *Exploratory designs for computational experiments*, Jour-
589 nal of statistical planning and inference, 43 (1995), pp. 381–402.
- 590 [24] S. NANTY, C. HELBERT, A. MARREL, N. PÉROT, AND C. PRIEUR, *Sampling, metamodeling, and*
591 *sensitivity analysis of numerical simulators with functional stochastic inputs*, SIAM/ASA
592 Journal on Uncertainty Quantification, 4 (2016), pp. 636–659.
- 593 [25] S. NANTY, C. HELBERT, A. MARREL, N. PÉROT, AND C. PRIEUR, *Uncertainty quantification for*
594 *functional dependent random variables*, Computational Statistics, 32 (2017), pp. 559–583.
- 595 [26] L. PRONZATO AND W. G. MÜLLER, *Design of computer experiments: space filling and beyond*,
596 Statistics and Computing, 22 (2012), pp. 681–701.
- 597 [27] J. O. RAMSAY, *Functional data analysis*, Wiley Online Library, 2006.
- 598 [28] O. ROUSTANT, D. GINSBOURGER, AND Y. DEVILLE, *Dicekriging, diceoptim: Two r packages*
599 *for the analysis of computer experiments by kriging-based metamodeling and optimization*,
600 (2012).
- 601 [29] P. RUFFO, L. BAZZANA, A. CONSONNI, A. CORRADI, A. SALTELLI, AND S. TARANTOLA, *Hydro-*
602 *carbon exploration risk evaluation through uncertainty and sensitivity analyses techniques*,
603 Reliability Engineering & System Safety, 91 (2006), pp. 1155–1162.
- 604 [30] J. SACKS, W. J. WELCH, T. J. MITCHELL, AND H. P. WYNN, *Design and analysis of computer*
605 *experiments*, Statistical science, (1989), pp. 409–423.
- 606 [31] E. VAZQUEZ AND J. BECT, *A sequential bayesian algorithm to estimate a probability of failure*,
607 IFAC Proceedings Volumes, 42 (2009), pp. 546–550.