Feasible set estimation under functional uncertainty by Gaussian Process modelling

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

Feasible set estimation under functional uncertainty by Gaussian Process modelling

Mohamed Reda El Amri[a,*], Céline Helbert[b], Miguel Munoz Zuniga[c], Clémentine Prieur[d], Delphine Sinoquet[e]

[a]*IFP Energies Nouvelles, Solaize, France (mohamed-reda.el-amri@ifpen.fr)*
[b]*ECL, ICJ, UMR 5208, Université de Lyon, 36 av. G. de Collongue, Ecully, France (celine.helbert@ec-lyon.fr).*
[c]*IFP Energies Nouvelles, Rueil-Malmaison, France (miguel.munoz-zuniga@ifpen.fr).*
[d]*Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, Grenoble, France (clementine.prieur@univ-grenoble-alpes.fr).*
[e]*IFP Energies Nouvelles, Rueil-Malmaison, France (delphine.sinoquet@ifpen.fr).*

Abstract

In this paper we deal with the estimation of a feasible set defined by an inequality constraint on the output of a time-consuming black-box simulator. We focus on the setting where the black-box simulator takes as inputs both a set of scalar controlled variables and a functional uncontrolled variable. We then place ourselves in a probabilistic framework, modelling the functional uncontrolled variable by a random process. The inequality constraint is formulated as the expectation of the output of the simulator conditional on the values taken by the set of controlled variables. We propose an original method to solve the above feasibility problem with a reduced number of evaluations of the costly simulator. A Gaussian Process model of the simulator is learned in the joint space of controlled and uncontrolled input variables, on the basis of a set of simulations which is enriched through a sequential procedure. This procedure aims to reduce the estimation error of the feasible set by evaluating the simulator on new points chosen sequentially in the joint input space according to specific enrichment criteria. It involves as a preliminary step the reduction of the dimension of the uncontrolled input space. A variation of this strategy is also proposed, which increases adaptively the dimension of the reduced space, leading to an improvement in terms of number of calls to the simulator. The procedure we propose is compared with other sampling procedures and another modelling approach on analytical examples. Finally our methodology is implemented on an automotive industrial application. For this application, the feasible set to be recovered is the set of values of controlled variables of a gas after-treatment device leading to the respect of pollutant emission standards of a vehicle under driving profile uncertainties.

Keywords: feasible set estimation; Gaussian Process model; dimension reduction; functional uncertainty.

1. Introduction

In recent years, engineers and scientists are increasingly relying on computer models as surrogates for physical experimentation generally too costly or impossible to execute ([1, 2]). In particular, practitioners using these numerical simulations 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

*Corresponding author
Email address: mohamed-reda.el-amri@ifpen.fr (Mohamed Reda El Amri)

values leading to a prescribed value or range for the output of interest. The problem of estimating such a set is called hereafter feasible set estimation.

In our context, the numerical simulator, denoted f, takes two types of input variables: a set of controlled variables $\mathbf{x} \in \mathbb{X}$, and a set of uncontrolled variables $\mathbf{v} \in \mathcal{V}$. Without considering any assumptions on the set of uncontrolled variables $\mathbf{v}$, robust feasible set estimation consists in seeking the set of controlled variables $\mathbf{x} \in \mathbb{X}$ such that $\sup_{\mathbf{v}\in\mathcal{V}} f(\mathbf{x}, \mathbf{v})$ is smaller than a threshold c. Then, the difficulty of solving this estimation problem strongly depends on the set $\mathcal{V}$.

In our setting, the simulator f takes as inputs a set of scalar controlled variables and a functional uncontrolled variable. Then we place ourselves in a probabilistic framework, modelling the functional uncontrolled variable by a random process $\mathbf{V}$. We are interested in the estimation of the feasible set defined as $\Gamma^* := \{\mathbf{x} \in \mathbb{X} \ , \ g(\mathbf{x}) = \mathbb{E}_{\mathbf{V}}[f(\mathbf{x}, \mathbf{V})] \leq c\}$, with $c \in \mathbb{R}$. It is important to note here that our study is driven by an industrial application on automotive depollution. More precisely, we study an after-treatment device of diesel vehicles, depending on controlled variables, in an uncertain environment corresponding to the uncontrolled driving profile. Knowledge on the driving profile is provided through a finite set of realizations of moderate size (see Section 4.4 for more details). In order to fit this setting, we make the assumption that the process $\mathbf{V}$ is only known through a finite set, denoted Ξ, of its realizations. Note that to estimate the expectation $\mathbb{E}_{\mathbf{V}}[f(\mathbf{x}, \mathbf{V})]$ appearing in the definition of Γ^*, a brute force Monte Carlo is out of reach as each evaluation of the simulator is time consuming. Therefore we propose in the following a more elaborate sampling strategy.

Feasible set estimation has already been carried out in many applications, notably reliability engineering (see, e.g., [1], [2]), climatology (see, e.g., [3], [4]) and many other fields. In the literature, one way to solve the problem is to adopt a sequential sampling strategy based on a Gaussian Process (GP) model for $g : \mathbf{x} \mapsto \mathbb{E}_{\mathbf{V}}[f(\mathbf{x}, \mathbf{V})]$. The underlying idea is that Gaussian Process models, which capture prior knowledge about the regularity of the unknown function, make it possible to assess the estimation error of Γ^* given a set of evaluations of g. More specifically, for the estimation of a feasible set, these sequential strategies are closely related to the field of Bayesian global optimization (see, e.g., [5]). In the case of feasible set estimation, specific Stepwise Uncertainty Reduction (SUR) strategies were introduced in [6]. More recently, a parallel implementation of these strategies has been proposed in [2] and applied to the recovery of a feasible set. Briefly, the strategy SUR gives sequentially the next location in the controlled space where to run the simulator in order to minimize a function (called uncertainty function hereafter) measuring the estimation error of the feasible set.

In the field of robust optimization where uncertainty comes from a real-valued (or vector-valued) random input, various methods exist and aim at optimizing the expectation taken with respect to the probability distribution of the random input (see [7] or [8]). These methods are based on the modelling of f by a Gaussian Process built in the joint space of controlled and uncontrolled variables. Then a "projected" (integrated) Gaussian Process is defined by taking the expectation with respect to the probability distribution of the random input, leading to an approximation of the expected response g. Finally a sequential design of experiments (DoE) is proposed for optimizing the objective function g. In the same spirit, we propose an original method to solve a probabilistic feasible set estimation problem with the aim of reducing at most the number of evaluations of the simulator required. In this work f is approximated by a Gaussian Process model built on $\mathbb{X} \times \mathbb{R}^m$, a finite-dimensional approximation of $\mathbb{X} \times \mathcal{V}$. The choice of the truncation argument m will be discussed further. For the iterative approximation of Γ^*, the sampling strategy in the joint space is based on two steps. Firstly a SUR approach is applied to the "projected" Gaussian Process to determine the next evaluation point $\mathbf{x}_{n+1} \in \mathbb{X}$. Secondly, in the uncontrolled space, the next realization v_{n+1} of the random process $\mathbf{V}$ is chosen such that the standard error of the "projected" process evaluated at $\mathbf{x}_{n+1}$ is minimized.

As already mentioned, our procedure relies on a preliminary step which aims, for a truncation

argument m, at estimating the m-truncated Karhunen-Loève (KL) decomposition of the random process $\mathbf{V}$ from the finite set Ξ of realizations of $\mathbf{V}$. Building the sampling strategy on a GP model built on the finite-dimensional space $\mathbb{X} \times \mathbb{R}^m$ certainly involves a loss of information. However, we propose to mitigate this loss of information in the following manner: once the next point is selected in $\mathbb{X} \times \mathbb{R}^m$, we evaluate the simulator on its closest neighbor in $\mathbb{X} \times \Xi$, where we recall that Ξ is the available finite sample of realizations of $\mathbf{V}$ (for more details, see the description of **step** 8 after the statement of Algorithm 1). By the way, we recover partial knowledge of the full variability of the untruncated process $\mathbf{V}$, leading to a procedure which is robust with respect to the truncation argument m. Also, we propose in Algorithm 2 a variation of our procedure, by increasing adaptively the truncation argument m over the iterates.

Another procedure for solving the feasible set estimation problem described in this introduction was proposed in [9]. The main difference with our work is that the sequential enrichment strategy is defined on $\mathbb{X}$, the space of controlled variables and not on the joint space of controlled and uncontrolled variables. Then, at each iteration, once the new point $\mathbf{x}_{n+1}$ is selected, an accurate approximation of $\mathbb{E}_{\mathbf{V}}[f(\mathbf{x}_{n+1}, \mathbf{V})]$ is computed via quantization. The procedure we propose in the present work outperforms the one in [9] by allowing to reduce even more the number of evaluations of the costly simulator (see Figure 6 Section 4.2).

The article is structured as follows. In Section 2, we recall the problem formulation and we extend the concept of Gaussian Process modelling to the case where one of the inputs is a random process known only through a finite set of realizations. In Section 3, we introduce a new sequential sampling strategy targeted for robust feasible set estimation to choose the next point in the joint space: $(\mathbf{x}_{n+1}, \mathbf{v}_{n+1})$ (Sections 3.1 and 3.2). In Section 3.3, together with numerical implementation details, we summarize our strategy to tackle robust feasible set estimation in two algorithms: with fixed KL parameter m (Algorithm 1) and its adaptive counterpart (Algorithm 2). The results of these algorithms on two analytical test cases are presented in Section 4.1 to 4.3. In particular, concerning the sampling enrichment in uncertain space, we compare our sampling strategy, based on the standard deviation of the "projected" process evaluated at $\mathbf{x}_{n+1}$, with a uniform sampling of v_{n+1} among the finite set of available realizations of the random process $\mathbf{V}$. We also compare our procedure with the one introduced in [9] which combines the fitting of a Gaussian Process model on the controlled space $\mathbb{X}$ with a quantization estimation of the expectation of the output conditional on the values taken by the controlled variables. In Section 4.4, our new procedure is tested on the industrial application of a car pollution control system. Finally further discussion on the modelling assumptions is postponed to Appendix A.

2. Problem formulation

We model the output of the industrial simulator by a function $f : \mathbb{X} \times \mathcal{V} \to \mathbb{R}$ with $\mathbb{X}$ a bounded subset of $\mathbb{R}^p$ being the controlled variable space and $\mathcal{V}$ the functional space in which the random process $\mathbf{V}$, modelling uncertainties, takes its values. We are interested in estimating the feasible set

$$\Gamma^* = \{\mathbf{x} \in \mathbb{X} \ , \ g(\mathbf{x}) \leq c\}, \tag{1}$$

where $c \in \mathbb{R}$ is a threshold and $g : \mathbb{X} \to \mathbb{R}$ such that $g(\mathbf{x}) = \mathbb{E}_{\mathbf{V}}[f(\mathbf{x}, \mathbf{V})]$. An additional constraint is that the random process $\mathbf{V}$ is known only through a finite set of realizations, denoted by Ξ. The implication of this constraint will be specified in Section 3.3. The proposed sequential strategy to estimate Γ^* involves three main ingredients introduced hereafter: dimension reduction to reduce the random process $\mathbf{V}$ to a m-dimensional random vector, Gaussian Process modelling in the joint space $\mathbb{X} \times \mathbb{R}^m$ and a wise selection of next point $(\mathbf{x}_{n+1}, \mathbf{u}_{n+1}) \in \mathbb{X} \times \mathbb{R}^m$ at which to evaluate the simulator. Although the Gaussian Process model is defined on the finite-dimensional truncated space $\mathbb{X} \times \mathbb{R}^m$, robustness with respect to truncation level m is mitigated through the sequential enrichment procedure of the design of experiments as for each selected point, the simulator is

evaluated at point $(\mathbf{x}_{n+1}, \mathbf{v}_{n+1})$ where $\mathbf{v}_{n+1}$ is the realization in Ξ corresponding to the truncated vector $\mathbf{u}_{n+1}$ in a sense to be precised in the following subsections.

2.1. Random process finite dimensional representation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We assume that the random process $\mathbf{V}$ belongs to $\mathcal{H} = \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{V})$ with

$$\mathcal{V} = \left\{ \mathbf{v} : [0,T] \to \mathbb{R}, ||\mathbf{v}|| = (< v, v >)^{1/2} = \left(\int_0^T \mathbf{v}(t)^2 \mathbf{dt} \right)^{1/2} < +\infty \right\}.$$

We assume that $\mathbf{V} \in \mathcal{H}$ has zero mean and continuous covariance function $C(t,s)$. Then

$$\forall t \in [0,T],\ \mathbf{V}(t) = \sum_{i=1}^{\infty} U_i \psi_i(t), \tag{2}$$

where $\{\psi_i\}_{i=1}^{\infty}$ is an orthonormal basis of eigenfunctions of the integral operator corresponding to C such that:

$$\lambda_i \psi_i(t) = \int_0^T C(t,s)\psi_i(s)\mathrm{ds}, \tag{3}$$

and with $\{U_i\}_{i=1}^{\infty}$ denoting a set of uncorrelated random variables with zero mean and variance λ_i. Decomposition (2) is known as the Karhunen-Loève (KL) expansion of $\mathbf{V}$ ([10]). In the following we denote the truncated version of $\mathbf{V}$ as $\mathbf{V}_m$:

$$\forall t \in [0,T],\ \mathbf{V}_m(t) = \sum_{i=1}^{m} U_i \psi_i(t), \tag{4}$$

which represents, in the mean square error sense, the optimal m-term approximation of $\mathbf{V}$ ([10]). The value of the parameter m should be chosen such that the approximation is accurate enough. Its influence in practice is discussed in Section 4.2.

2.2. Gaussian Process modelling

We assume that $f(\mathbf{x}, \mathbf{v})$ is a realization of a Gaussian Process $Z_{(\mathbf{x},\mathbf{u})}$ defined on $\mathbb{X} \times \mathbb{R}^m$, where $\mathbf{u} = (< \mathbf{v}, \psi_1 >, \ldots, < \mathbf{v}, \psi_m >)^\top$. Let m_Z be the mean function of $Z_{(\mathbf{x},\mathbf{u})}$ and k_Z its covariance function,

$$\begin{aligned} \mathbb{E}[Z_{(\mathbf{x},\mathbf{u})}] &= m_Z(\mathbf{x},\mathbf{u}), \\ \mathrm{Cov}(Z_{(\mathbf{x},\mathbf{u})}, Z_{(\mathbf{x}',\mathbf{u}')}) &= k_Z((\mathbf{x},\mathbf{u});(\mathbf{x}',\mathbf{u}')). \end{aligned} \tag{5}$$

Let us denote Z^n, the GP Z conditioned on the set of n observations $\mathbf{Z}_n = \{f(\mathbf{x}_1,\mathbf{v}_1), \ldots, f(\mathbf{x}_n,\mathbf{v}_n)\}$ of Z at $(\mathcal{X}_n, \mathcal{U}_n) = \{(\mathbf{x}_1,\mathbf{u}_1), \ldots, (\mathbf{x}_n,\mathbf{u}_n)\}$ where $\mathbf{u}_i = (< \mathbf{v}_i, \psi_1 > , \ldots, < \mathbf{v}_i, \psi_m >)^\top$

$$Z^n_{(\mathbf{x},\mathbf{u})} = [Z_{(\mathbf{x},\mathbf{u})} | Z_{(\mathcal{X}_n,\mathcal{U}_n)} = \mathbf{Z}_n]. \tag{6}$$

The conditional mean and covariance are given by

$$\begin{aligned} \mathbb{E}[Z^n_{(\mathbf{x},\mathbf{u})}] &= m_Z(\mathbf{x},\mathbf{u}) + k_Z((\mathbf{x},\mathbf{u});(\mathcal{X}_n,\mathcal{U}_n))\Sigma^{-1}_{Z,n}(\mathbf{Z} - m_Z(\mathcal{X}_n,\mathcal{U}_n)), \\ \mathrm{Cov}(Z^n_{(\mathbf{x},\mathbf{u})}, Z^n_{(\mathbf{x}',\mathbf{u}')}) &= k_Z((\mathbf{x},\mathbf{u});(\mathbf{x}',\mathbf{u}')) - k_Z((\mathbf{x},\mathbf{u});(\mathcal{X}_n,\mathcal{U}_n))\Sigma^{-1}_{Z,n} k_Z((\mathcal{X}_n,\mathcal{U}_n);(\mathbf{x}',\mathbf{u}')). \end{aligned}$$

where $\Sigma_{Z,n} = k_Z((\mathcal{X}_n,\mathcal{U}_n);(\mathcal{X}_n,\mathcal{U}_n))$. The Gaussian Process $Z_{(\mathbf{x},\mathbf{u})}$ is defined on the finite-dimensional truncated space $\mathbb{X} \times \mathbb{R}^m$. However, it is worth underlying that we recollect partial knowledge of the full variability of the untruncated process $\mathbf{V}$, despite the truncation of the KL expansion, by evaluating the simulator on design points in $\mathbb{X} \times \Xi$. A discussion about this model is proposed in Appendix A.

2.3. Integrated Gaussian Process

Recall that $\Gamma^* = \{\mathbf{x} \in \mathbb{X} \ , \ g(\mathbf{x}) = \mathbb{E}[f(\mathbf{x}, \mathbf{V})] \leq c\}$. Therefore, to model the function g, we introduce the integrated process

$$Y^n_{\mathbf{x}} = \mathbb{E}_{\mathbf{U}}[Z^n_{(\mathbf{x},\mathbf{U})}] = \int_{\mathbb{R}^m} Z^n_{(\mathbf{x},\mathbf{u})} d\rho(\mathbf{u}), \tag{7}$$

where ρ is the probability distribution of $\mathbf{U} = (U_1, \ldots, U_m)^\top$ introduced in (4). The process $Y^n_{\mathbf{x}}$ is a Gaussian Process ([7]) fully characterized by its mean and covariance functions which are given by

$$\mathbb{E}[Y^n_{\mathbf{x}}] = \int_{\mathbb{R}^m} m_Z(\mathbf{x}, \mathbf{u}) + k_Z((\mathbf{x}, \mathbf{u}); (\mathcal{X}_n, \mathcal{U}_n))\Sigma^{-1}_{Z,n}(\mathbf{Z} - m_Z(\mathcal{X}_n, \mathcal{U}_n))d\rho(\mathbf{u}), \tag{8}$$

and

$$\mathrm{Cov}(Y^n_{\mathbf{x}}, Y^n_{\mathbf{x}'}) = \iint_{\mathbb{R}^m} k_Z((\mathbf{x}, \mathbf{u}); (\mathbf{x}', \mathbf{u}')) - k_Z((\mathbf{x}, \mathbf{u}); (\mathcal{X}_n, \mathcal{U}_n))\Sigma^{-1}_{Z,n} k_Z((\mathcal{X}_n, \mathcal{U}_n); (\mathbf{x}', \mathbf{u}'))d\rho(\mathbf{u})d\rho(\mathbf{u}'). \tag{9}$$

3. Data driven infill strategy for robust feasible set estimation

In this section we propose a two-step infill strategy in the joint space. The first step consists in choosing a point in the controlled space while the second one aims at enriching the design with a new point in the uncertain space.

3.1. Minimization of the Vorob'ev deviation: choice of next $\boldsymbol{x}$

The objective of the first step is to wisely choose the points in the controlled space $\mathbb{X}$ in order to accurately estimate the set $\Gamma^* = \{\mathbf{x} \in \mathbb{X} \ , \ g(\mathbf{x}) = \mathbb{E}_{\mathbf{V}}[f(\mathbf{x}, \mathbf{V})] \leq c\}$. For this purpose, we consider the statistical model of the non-observable function g given by $Y^n_{\mathbf{x}}$ introduced in Section 2.3. In the following, we assume that the Gaussian Process $(Z_{(\mathbf{x},\mathbf{u})})_{(\mathbf{x},\mathbf{u})\in\mathbb{X}\times\mathbb{R}^m}$ is separable with continuous mean function m_Z and Matérn (5/2 or 3/2) covariance function k_Z. Then the feasible set defined as $\Gamma = \{\mathbf{x} \in \mathbb{X} \ , \ Y^n_{\mathbf{x}} \leq c\}$ is a random closed set (see, e.g., [11] p.4, 23).

From the assumption that g is a realization of $Y^n_{\mathbf{x}}$, the true unknown set Γ^* can be seen as a realization of the random closed set Γ. The book of [11] gives many possible definitions for the variance of a random closed set. In the present work we focus on the Vorob'ev deviation ([12, 13]) and we adapt the Stepwise Uncertainty Reduction (SUR) strategy introduced in [5] which aims at decreasing an uncertainty function defined as the Vorob'ev deviation ([12, 13]) of the random set.More precisely the uncertainty function at step n is defined as

$$\mathcal{H}^{\mathrm{uncert}}_n = \mathbb{E}[\mu(\Gamma \triangle Q_{n,\alpha^*_n}) \mid Z_{(\mathcal{X}_n,\mathcal{U}_n)} = \mathbf{Z}_n],$$

where μ is the Lebesgue measure on $\mathbb{X}$, Δ the symmetric difference operator between two sets, the Vorob'ev quantiles are given by $Q_{n,\alpha} = \{\mathbf{x} \in \mathbb{X} \ , \ \mathbb{P}(Y^n_{\mathbf{x}} \leq c) \geq \alpha\}$, and the Vorob'ev expectation Q_{n,α^*_n} can be determined by tuning α to a level α^* such that $\mu(Q_{n,\alpha^*_n}) = \mathbb{E}[\mu(\Gamma) \mid Z_{(\mathcal{X}_n,\mathcal{U}_n)} = \mathbf{Z}_n]$. Let

$$\mathcal{H}^{\mathrm{uncert}}_{n+1}(\mathbf{x}) = \mathbb{E}[\mu(\Gamma \triangle Q_{n+1,\alpha^*_{n+1}}) \mid Z_{(\mathcal{X}_n,\mathcal{U}_n)} = \mathbf{Z}_n, Y^n_{\mathbf{x}}].$$

The objective of the SUR strategy is thus to enrich the current design with a new point $\mathbf{x}_{n+1}$ satisfying

$$\begin{aligned} \mathbf{x}_{n+1} &\in \mathrm{argmin}_{\mathbf{x}\in\mathbb{X}} \ \mathbb{E}_{n,x}[\mathcal{H}^{\mathrm{uncert}}_{n+1}(\mathbf{x})] \\ &:= \mathrm{argmin}_{\mathbf{x}\in\mathbb{X}} \ \mathcal{J}_n(\mathbf{x}), \end{aligned} \tag{10}$$

where $\mathbb{E}_{n,\mathbf{x}}$ denotes the expectation with respect to $Y^n_{\mathbf{x}}$. For the computation of $\mathcal{J}_n(\mathbf{x})$, we use the formula given in [14, Eq. (4.43) in Section 4.2] (see also [5, Section 3.1]).

The enrichment of the DoE consists in selecting a couple $(\mathbf{x}_{n+1}, \mathbf{u}_{n+1})$ in the joint space $\mathbb{X} \times \mathbb{R}^m$. $\mathbf{x}_{n+1}$ has just been defined by (10), it remains now to choose a new point u_{n+1} in the uncertain space.

3.2. Minimization of the variance: choice of next $\boldsymbol{u}$

The process Y^n approximates the expectation $\mathbb{E}_{\mathbf{V}}[f(\cdot, \mathbf{V})]$. It can be seen as a projection of Z^n from the joint space onto the controlled space. We propose to sample the point $\mathbf{u}_{n+1}$ in the uncertain space in order to reduce at most the one-step-ahead variance at point $\mathbf{x}_{n+1}$, $\mathbb{VAR}(Y^{n+1}_{\mathbf{x}_{n+1}})$, whose expression is obtained from Eq. (9). More precisely,

$$\mathbf{u}_{n+1} = \text{argmin}_{\tilde{\mathbf{u}} \in \mathbb{R}^m} \mathbb{VAR}(Y^{n+1}_{\mathbf{x}_{n+1}}), \tag{11}$$

with

$$\begin{aligned} \mathbb{VAR}(Y^{n+1}_{\mathbf{x}_{n+1}}) &= \vartheta(\tilde{\mathbf{u}}), \\ &= \iint_{\mathbb{R}^m} k_Z((\mathbf{x}_{n+1}, \mathbf{u}); (\mathbf{x}_{n+1}, \mathbf{u}'))d\rho(\mathbf{u})d\rho(\mathbf{u}') \\ &- \iint_{\mathbb{R}^m} k_Z((\mathbf{x}_{n+1}, \mathbf{u}); (\mathcal{X}_{n+1}, \mathcal{U}_{n+1}))\Sigma^{-1}_{Z,n+1} k_Z((\mathcal{X}_{n+1}, \mathcal{U}_{n+1}); (\mathbf{x}_{n+1}, \mathbf{u}'))d\rho(\mathbf{u})d\rho(\mathbf{u}'), \end{aligned} \tag{12}$$

where $\Sigma_{Z,n+1} = k_Z((\mathcal{X}_{n+1}, \mathcal{U}_{n+1}); (\mathcal{X}_{n+1}, \mathcal{U}_{n+1}))$ and $(\mathcal{X}_{n+1}, \mathcal{U}_{n+1}) = (\mathcal{X}_n, \mathcal{U}_n) \cup \{(\mathbf{x}_{n+1}, \tilde{\mathbf{u}})\}$.

3.3. Implementation

We present in this section two algorithms. Algorithm 1 is the global algorithm summarizing the strategy we propose for robust feasible set estimation. The truncation argument is fixed once for all in the algorithm. Algorithm 2 is a variation of Algorithm 1 with a data driven procedure for increasing the truncation argument m over the iterations. This last algorithm allows to further reduce the number of calls to the simulator.

The setting of our procedure is driven by our industrial application where the random process $\mathbf{V}$ is known only through a finite set of realizations $\Xi = \{\breve{\mathbf{v}}_1, \ldots, \breve{\mathbf{v}}_N\}$. In this framework, points a) and b) below detail the computation of KL decomposition and the minimization of the one-step-ahead variance.

a) Computational method for functional PCA. We consider the empirical version of $C(s,t)$ defined as $C^N(s,t) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \breve{\mathbf{v}}_i(s)\breve{\mathbf{v}}_i(t)$. The eigenvalue problem defined by Eq. (3) is then solved by discretizing the trajectories $\{\breve{\mathbf{v}}_i\}_{i=1,\ldots,N}$ on $[0,T]$ and replacing C by C^N. Denoting by $\hat{\psi}_i$, $i = 1, \ldots, m$, the m first estimated eigenfunctions, we define

$$\mathcal{G}_m = \{\breve{\mathbf{u}}_1, \ldots \breve{\mathbf{u}}_N\} \tag{13}$$

where $\breve{\mathbf{u}}_i = (< \breve{\mathbf{v}}_i, \hat{\psi}_1 >, \ldots, < \breve{\mathbf{v}}_i, \hat{\psi}_m >)^\top$.

b) Minimization of the one-step-ahead variance. Since $\mathbf{V}$ is known through a finite set Ξ, Eq. (11) is solved on the finite set $\mathcal{G}_m$.

We now state the global algorithm we propose for robust feasible set estimation (Algorithm 1) and its variation with a data driven increase of the truncation argument (Algorithm 2).

3.3.1. A first algorithm for robust feasible set estimation

In this section we provide a global algorithm for the implementation of our methodology. Then we comment some of its steps.

Algorithm 1 Robust feasible set estimation via joint space modelling

Require: The truncation argument m, the initial DoE of n_0 points $(\mathcal{X}_n, \mathcal{U}_n)$ in $\mathbb{X} \times \mathcal{G}_m$, and a maximal simulation budget

Set $n = n_0$.
Calculate $\mathbf{Z}$ the simulator responses at the design points $(\mathcal{X}_n, \mathcal{U}_n)$
while $n \leq$ budget **do**
  Fit the GP model Z^n
  Induce the integrated GP $Y^n_{\mathbf{x}}$
  $\mathbf{x}_{n+1} \leftarrow$ sampling criterion $\mathcal{J}_n$
  $\mathbf{u}_{n+1} \leftarrow \text{argmin}_{\tilde{\mathbf{u}} \in \mathcal{G}_m} \mathbb{VAR}(Y^{n+1}_{\mathbf{x}_{n+1}})$
  Simulation at $(\mathbf{x}_{n+1}, \mathbf{v}_{n+1})$, where $\mathbf{v}_{n+1} \in \Xi$ is the curve corresponding to $\mathbf{u}_{n+1}$
  Update DoE : $(\mathcal{X}_{n+1}, \mathcal{U}_{n+1}) = (\mathcal{X}_n, \mathcal{U}_n) \cup \{(\mathbf{x}_{n+1}, \mathbf{u}_{n+1})\}$
  Update $\mathbf{Z} = \mathbf{Z} \cup \{f(\mathbf{x}_{n+1}, \mathbf{v}_{n+1})\}$
  Set $n = n + 1$
end while
Fit the final GP model Z^n
Approximate Γ^* by the Vorob'ev expectation

step 1 Let $\mathbb{U}$ be the smallest m-rectangle containing $\mathcal{G}_m$, $\mathbb{U} = \prod_{i=1}^m [\min(< \Xi, \hat{\psi}_i >), \max(< \Xi, \hat{\psi}_i >)])$. For the initial DoE, we first build a Random Latin Hypercube Design of n points $(\mathcal{X}_n, \bar{\mathcal{U}}_n)$ in the joint space $(\mathbb{X}, \mathbb{U})$. Then the set of points $\mathcal{U}_n$ is determined such that for $i = 1, ..., n$, $\mathbf{u}_i \in \mathcal{G}_m$ is the closest point from $\bar{\mathbf{u}}_i \in \bar{\mathcal{U}}_n$ (with respect to the euclidean norm in $\mathbb{R}^m$).

step 4 The covariance kernel of the GP is chosen as a Matèrn-5/2 covariance and we add a noise modelled with a constant variance term. The homoscedastic modelling of the noise is discussed in Appendix A. The mean function of the GP is modelled by a constant function. All types of parameters (mean, correlation lengths, variance and noise) are estimated by maximum likelihood [15].

step 5 In the framework where the uncertain vector $\mathbf{U}$ is Gaussian as well as the covariance kernel, closed form solutions of the integrals in (8) and (9) are given in [7]. In our framework, the integrals in (8) and (9) are approximated by Monte Carlo.

step 6 $\mathbf{x}_{n+1}$ is obtained by solving (10) with a continuous global optimization algorithm: GENetic Optimization Using Derivatives (GENOUD) [16].

step 7 Once more the integrals in (12) are approximated by Monte Carlo. More details on the estimation of (12) can be found in [7]. Here the minimization problem is solved by an exhaustive search on the finite set $\mathcal{G}_m$ defined in (13).

step 8 The simulator is evaluated at point $(\mathbf{x}_{n+1}, \mathbf{v}_{n+1})$ where $\mathbf{v}_{n+1}$ is the curve of the initial set of curves Ξ corresponding to the truncated vector of coefficients $\mathbf{u}_{n+1}$. Note that evaluating the simulator at a curve in the initial set of realizations whose coordinate in the uncertain space is $\mathbf{u}_{n+1}$ and not a projected curve on the basis composed with first eigenfunctions brings robustness with respect to the truncation argument.

3.3.2. A variant of the proposed algorithm with a varying size of the reduced uncertain space

One limitation of our methodology is the prior choice of the truncation argument m. It can be based on a sufficient level of explained variance. But, depending on the stochastic process involved, this parameter can be high (up to 60 for the industrial application). Increasing m

implies increasing the dimension of the GP space. A high number of design points is then needed to produce an accurate response surface, which is very costly in simulation calls. To overcome the simulation extra cost, another variant of our strategy is hereafter introduced. The approach consists in augmenting the uncertain space sequentially when needed. More precisely, a first Gaussian Process is defined in the $p+m$ dimensional space, with m chosen small. Once the enrichment strategy (given by Algorithm 1) no longer provides information - a rough or bias approximation of the feasible set is achieved - the dimension of the uncertain space is increased and the GP is updated in the $p+m+1$ dimensional space. It is important to underline that this approach does not require additional calls to the numerical simulator. This alternative strategy is summarized by Algorithm 2:

Algorithm 2 Robust feasible set estimation via sequential joint space modelling

Require: The initial truncation argument m and the DoE of n points $(\mathcal{X}_n, \mathcal{U}_n)$ in $\mathbb{X} \times \mathcal{G}_m$

1: Set $n = n_0$.
2: Calculate $\mathbf{Z}$ the simulator responses at the design points $(\mathcal{X}_n, \mathcal{U}_n)$
3: **while** $n \leq$ budget **do**
4: $m \leftarrow$ Update.Dimension()
5: Fit the GP model Z^n
6: Induce the integrated GP $Y^n_{\mathbf{x}}$
7: $\mathbf{x}_{n+1} \leftarrow$ sampling criterion $\mathcal{J}_n$
8: $\mathbf{u}_{n+1} \leftarrow \mathtt{arg} \min_{\tilde{\mathbf{u}} \in \mathcal{G}_m} \mathbb{VAR}(Y^{n+1}_{\mathbf{x}_{n+1}})$
9: Simulator response at $(\mathbf{x}_{n+1}, \mathbf{v}_{n+1})$, where $\mathbf{v}_{n+1} \in \Xi$ is the curve corresponding to $\mathbf{u}_{n+1}$
10: Update DoE : $(\mathcal{X}_{n+1}, \mathcal{U}_{n+1}) = (\mathcal{X}_n, \mathcal{U}_n) \cup \{(\mathbf{x}_{n+1}, \mathbf{u}_{n+1})\}$
11: Update $\mathbf{Z} = \mathbf{Z} \cup \{f(\mathbf{x}_{n+1}, \mathbf{v}_{n+1})\}$
12: Set $n = n + 1$
13: **end while**
14: Fit the GP model Z^n
15: Approximate Γ^* by the Vorob'ev expectation

In **step** 4 of Algorithm 2, the uncertain space dimension is updated based on a stagnation criterion of the Vorob'ev Deviation. More precisely, the dimension is updated from m to $m+1$ if the following stopping criterion is verified

$$\forall\ 0 \leq j < n_0^{\mathrm{SUR}}\ ,\ e^{\mathrm{SUR}}_{n-j} \leq \epsilon^{\mathrm{SUR}}, \tag{14}$$

where $e^{\mathrm{SUR}}_n = \mid \mathcal{H}^{\mathrm{uncert}}_n - \mathcal{H}^{\mathrm{uncert}}_{n-1} \mid$ is the absolute error between Vorob'ev deviations estimated for two successive iterations. The condition in Eq. (14) tests if the variation of the Vorob'ev deviation is smaller than a tolerance ϵ^{SUR} on n_0^{SUR} consecutive steps, with n_0^{SUR} to be tuned by the practitioner.

4. Numerical experiments

In the following, we first introduce in Section 4.1 two analytical examples on which we will test our strategy. Then in Section 4.2 we present the results obtained by implementing Algorithm 1. In Section 4.3 we present the improvements we get by increasing adaptively the truncation argument in the KL decomposition over iterations (Algorithm 2). Finally Section 4.4 is devoted to the implementation of our strategy on an industrial application of automotive depollution.

4.1. Two analytical examples - set-up

To illustrate the behaviour of the proposed method, we consider two analytical examples. We suppose that a sample Ξ of size $N = 200$ realizations of the random process $\mathbf{V}$ is available and its probability distribution is unknown. To highlight the robustness of our method regarding the random distribution of the uncertainties, we consider two types of random processes: max-stable process and the well-known brownian motion. Regarding the max-stable process, we consider the Schlather model with powered exponential correlation function, i.e., $\ell(h) = \exp\{-(h/\lambda)^\kappa\}$, where $\kappa = 1$ and $\lambda = 10$. This process is also known as the extremal Gaussian process [17]. The function `rmaxstab` from the R-package `SpatialExtremes` is used to generate a sample of realizations. As Algorithm 1 depends on the truncation argument m, different values are tested (see Table 1) to better understand the effect of the uncertain space dimension.

	$m = 2$	$m = 4$	$m = 8$
$\mathbf{V}$: brownian motion	90.1%	95.2%	97.6%
$\mathbf{V}$: max-stable process	58.8%	63.3%	70%

Table 1: Variance explained by the truncated KL decomposition according to m for two different random processes.

For all analytical examples, we consider a Gaussian Process prior $Z_{(\mathbf{x},\mathbf{u})}$ with constant mean and Matèrn covariance kernel with $\nu = 5/2$. Random Latin Hypercube Designs (RLHD) are used as initial DoEs in all the experiments. The number of points of the initial DoE is 20 for the first analytical example and 30 for the second one. RLHD induce variability in the behaviour of the algorithms. To account for this variability, the performance is averaged over 30 (respectively 10) independent runs for brownian motion (respectively max-stable process).

Analytical example 1. We consider an additive function, sum of the two-dimensional Bohachevsky function and a random term, defined as

$$f : (\mathbf{x}, \mathbf{V}) \mapsto \left(x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) - 0.4\cos(4\pi x_2) + 0.7\right) + \int_0^T e^{\mathbf{V}_t} \mathrm{dt},$$

where $\mathbf{x} \in \mathbb{X} = [-100, 100]^2$. The objective is to approximate the set $\Gamma^* = \{\mathbf{x} \in \mathbb{X} \ , \ g(\mathbf{x}) = \mathbb{E}_{\mathbf{V}}[f(\mathbf{x}, \mathbf{V})] \leq 3500\}$ for the two different types of random processes (brownian motion and max-stable process).

Analytical example 2. For the second example we define a function that is not separable with respect to the controlled variables $\mathbf{x}$ and the random process $\mathbf{V}$. The function involves the maximum and the minimum of $(V_t)_{t\geq 0}$, so catching the whole variability of $\mathbf{V}$ becomes important. The function f is given by

$$f : (\mathbf{x}, \mathbf{V}) \mapsto \max_t \mathbf{V}_t |0.1\cos(x_1 \max_t \mathbf{V}_t)\sin(x_2)(x_1 + x_2 \min_t \mathbf{V}_t)^2| \int_0^T (30 + \mathbf{V}_t)^{\frac{x_1 x_2}{20}} \mathrm{dt},$$

where the controlled variables lie in $\mathbb{X} = [1.5, 5] \times [3.5, 5]$. The objective is to approximate the set $\Gamma^* = \{\mathbf{x} \in \mathbb{X} \ , \ g(\mathbf{x}) = \mathbb{E}_{\mathbf{V}}[f(\mathbf{x}, \mathbf{V})] \leq c\}$, when $c = 1.2$ for the brownian motion and $c = 0.9$ for the max-stable process.

Note that for both examples the reference solution Γ^* is obtained from a 30×30 grid experiment, where at each grid point the expectation is empirically approximated using the whole sample Ξ. In the following, we measure the performance of the different strategies with the ratio of the volume of the symmetric difference between the reference set Γ^* and the estimated one Q_{n,α^*} to the volume of the reference set: $\mu(\Gamma^* \triangle Q_{n,\alpha_n^*})/\mu(\Gamma^*)$ to which we will refer as the quality-ratio.

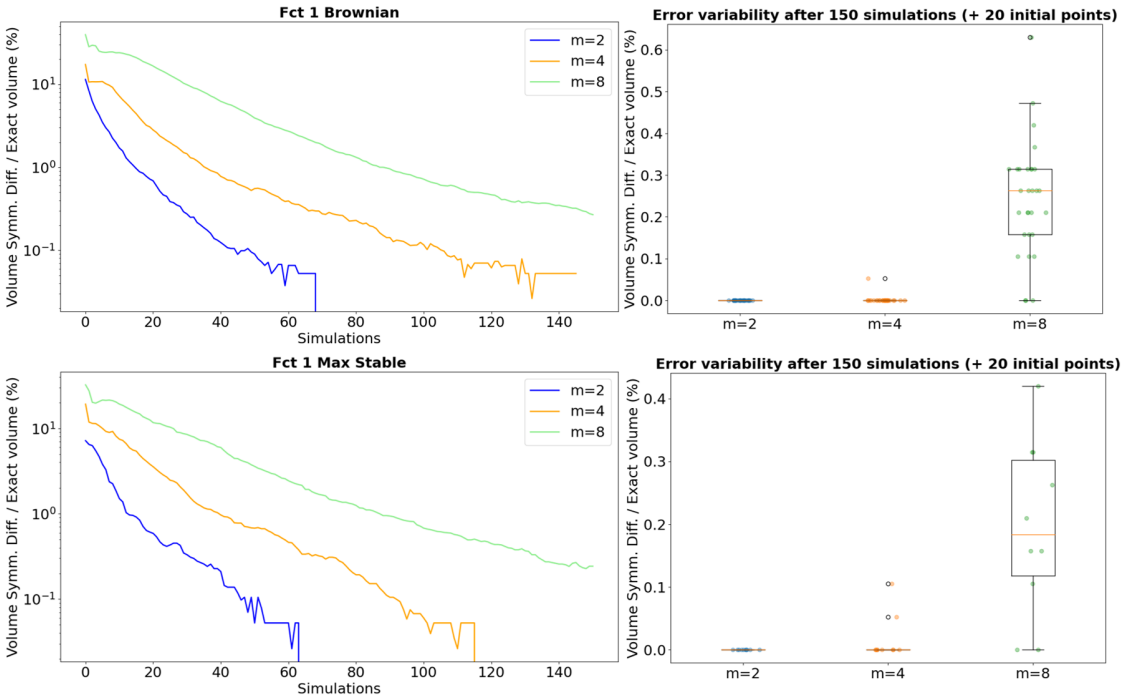

Figure 1: Analytical example 1 with brownian motion (top) and with max-stable process (bottom). Convergence of Algorithm 1 for $m = \{2, 4, 8\}$. Left: quality-ratio mean as function of the number of simulator calls in log scale. The mean is taken over the independent runs of initial RLHD. Right: quality-ratios associated with the random initial DoEs at the maximal simulation budget.

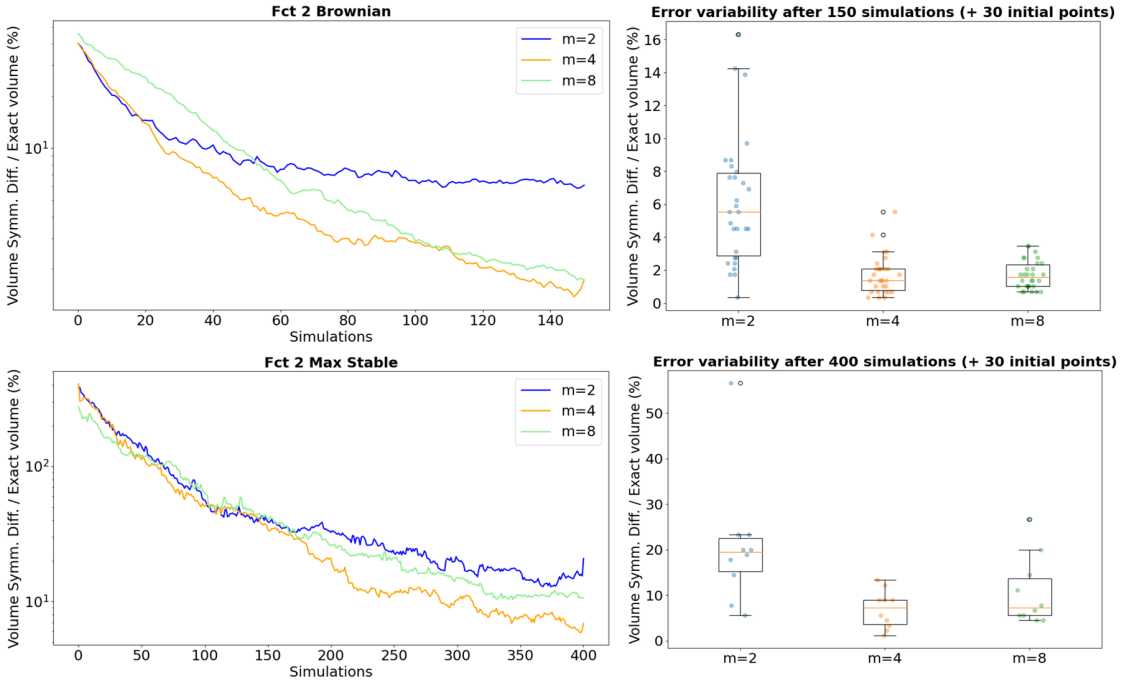

Figure 2: Analytical example 2 with brownian motion (top) and with max-stable process (bottom). Convergence of Algorithm 1 for $m = \{2, 4, 8\}$. Left: quality-ratio mean as function of the number of simulator calls in log scale. The mean is taken over the independent runs of initial RLHD. Right: quality-ratios associated with the random initial DoEs at the maximal simulation budget.

In Figures 1 and 2, we show the evolution of the averaged quality-ratio with respect to the number of simulations involved in the Algorithm 1 on the two analytical examples with the two types of functional uncertainties (brownian and max-stable processes). The average is taken over the repeated runs of the complete approach corresponding to the 30 random initial designs (10 for the max-table process), and for 3 values of the truncation argument m.

For the first analytical example, the smaller values of m, the faster is the convergence. This observation can be explained by the fact that, in higher dimensional joined space (due to larger values of m), much more evaluation points are necessary to learn an accurate GP model (more hyper-parameters to determine). It is worth noting that even for 90% (for brownian motion) or 58.8% (for max-stable process) of explained variance with $m = 2$ the proposed algorithm provides an efficient estimate of the true set Γ^*. Indeed, on stage 8 in Algorithm 1 the full curve $\mathbf{v}_{n+1} \in \Xi$ associated to $\mathbf{u}_{n+1}$ is recovered, such that the information lost after the dimension reduction is reduced, thereby further robustifying the method.

Regarding the second analytical example, the output depends on local behaviours of the stochastic process. The truncation argument $m = 2$ is too small to catch these dependencies, the function is sensitive to higher KL order. For the brownian motion, more than 95% of variance is explained with $m = 4$. It seems sufficient to obtain an accurate approximation of Γ^*. The improvement between $m = 2$ and $m = 4$ is noticeable. The improvement is not as important when the uncertainties are driven by a max-stable process since the percentage of explained variance increases slowly. Better results should be observed with $m = 8$. It is not the case because a higher dimension leads to difficulties in the estimation of the GP except by increasing consequently the number of observation points. Figure A.14 in Appendix shows the evolution of the feasible domain estimation with respect to the iterations of Algorithm 1 for the second analytic case and the brownian motion, and for different truncation levels.

As shown in Figure 3, the higher the dimension of the GP space is, the longer the internal computations last. Moreover, the computational time needed to provide the next evaluation point increases with the number of simulator calls, and thus with the number of iterations, the cost of kriging being directly linked to the learning sample size. For example in the case of $m = 8$ (resp. $m = 2$), iteration 80 requires 203 (resp. 126) seconds to provide the next evaluation point whereas iteration 150 requires 275 (resp. 164) seconds.

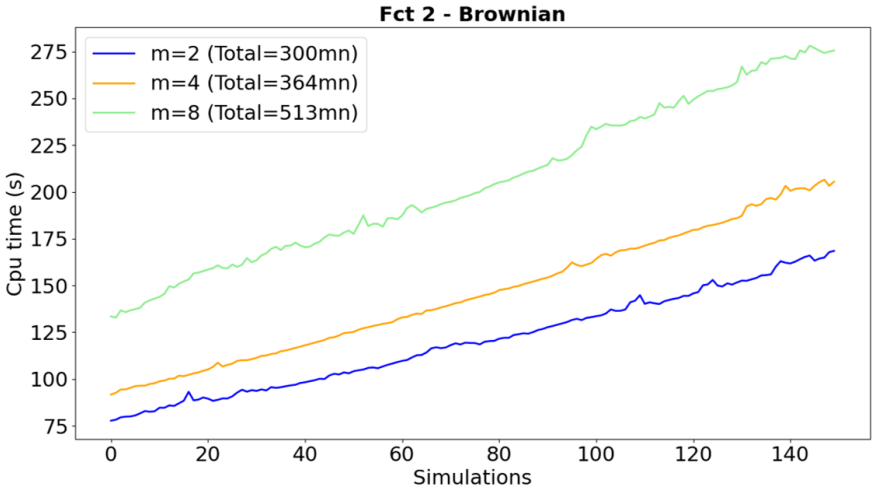

Figure 3: The computational time (sec.) needed to provide the next evaluation point as a function of iterations for the second analytic example with brownian motion. The values are averaged computational times for 5 runs of each strategy: $m = 2, 4, 8$.

To highlight the interest of the sampling criterion (11), we compared our approach to the one where $\mathbf{u}_{n+1}$ is chosen accordingly to a uniform distribution. The results obtained for both analytical cases with the brownian motion are shown on Figure 4. We note that our criterion leads to a faster decrease of the quality-ratio and to a much smaller error variability, in comparison to a uniform sampling in the uncertain space. As shown in Figure 5, the points selected with the criterion based on (11) seem to concentrate in interest areas, in comparison to the points selected uniformly. Thus, the guided sampling in the uncertain space, that reduces the variance of the expectation estimation, leads to a faster convergence towards the feasible domain.

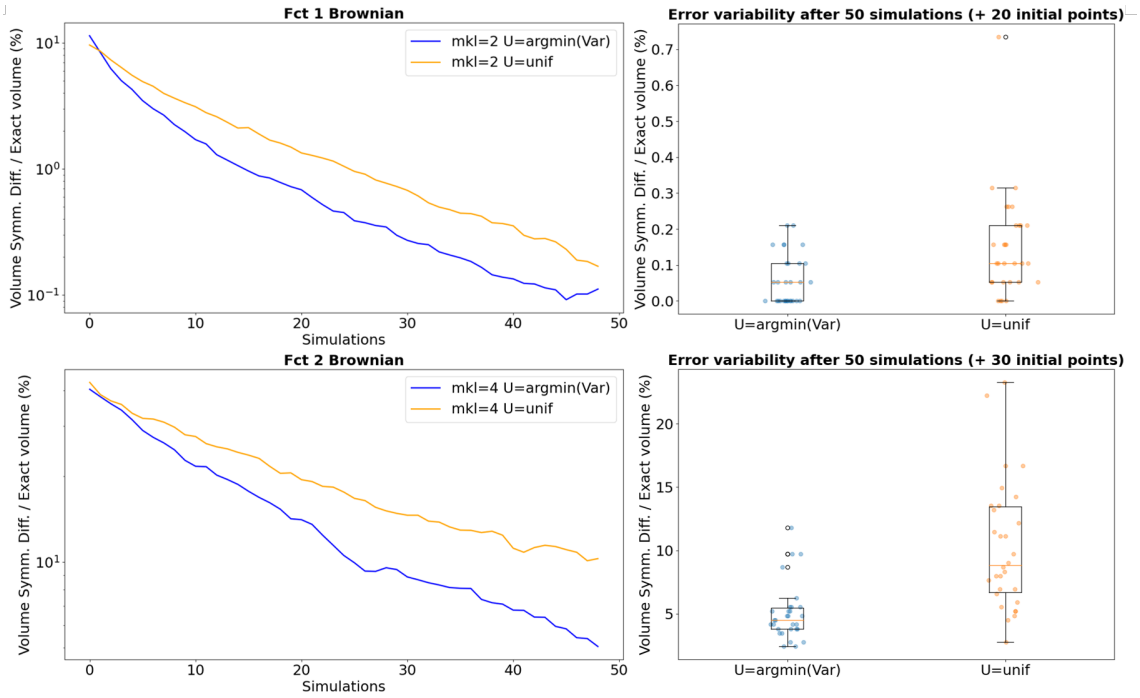

Figure 4: Analytical examples with the uncertainty modelled by a brownian motion. Decrease of the mean quality-ratio as function of the number of simulator calls (left) and boxplots for quality-ratios associated with the different random initial DoEs at the maximal simulation budget (right) for the first analytical case (top) and the second one (bottom), comparing criterion based on (11) with the uniform sampling in the uncertain space.

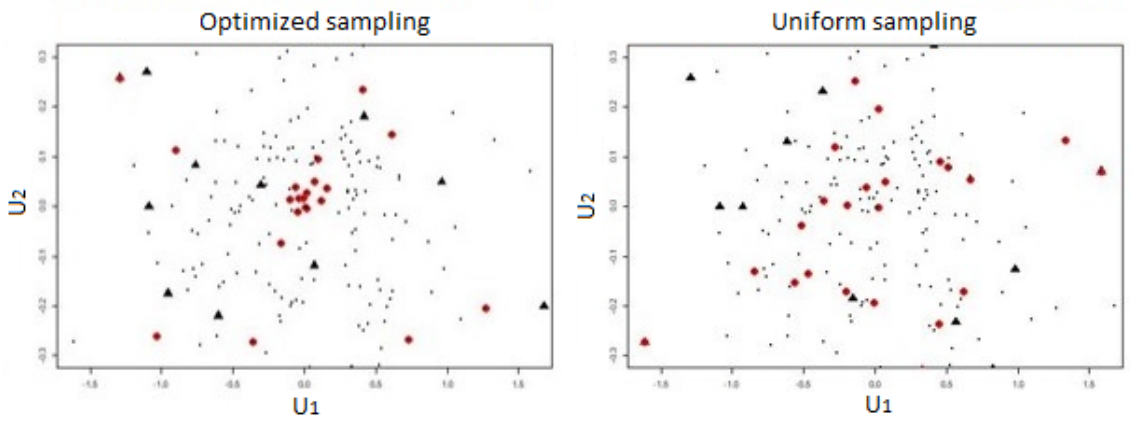

Figure 5: Analytical example 1 with the uncertainty modelled by a brownian motion. Black triangles correspond to the coefficients of the initial RLHD plotted in the uncertain truncated space ($m = 2$). Red points are the added points based on our criterion (left) and uniformly sampled (right) up to 50 simulations.

Finally we compare our joint modelling based method with the approach introduced in [9] which combines GP modelling in the controlled space with quantization to estimate the expectation in the uncontrolled space. Even without taking into account the costs induced by the initial designs in the controlled space (RLHD of size 9), the current approach based on joint GP modelling performs better regardless of the truncation argument m (Figure 6). Adding the costs induced by the

quantization on the initial sample points would disadvantage even more the approach introduced in [9] as the size of the quantizer is around 20 for each of the 9 points of the initial design, i.e. 180 simulator calls in average.

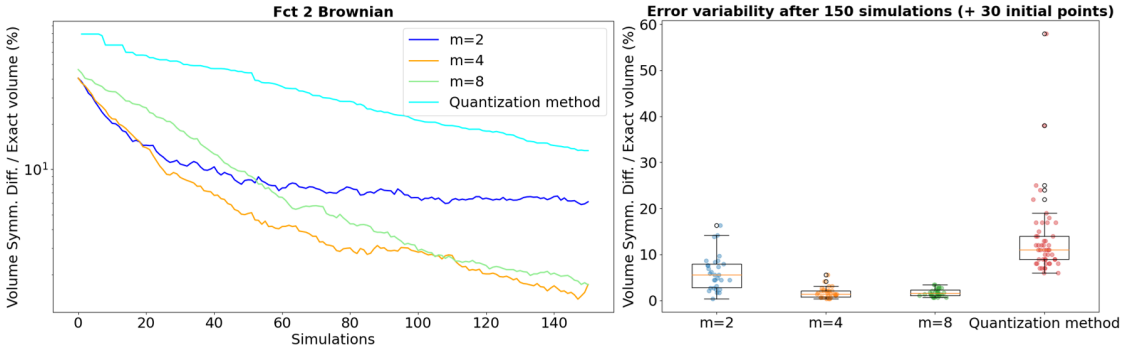

Figure 6: Comparison between the current approach based on joint GP modelling and the approach introduced in [9] combining GP modelling in the controlled space with quantization in the uncertain space. The costs induced by the initial designs (RLHD of size 9) are not compatibilized. The uncertainty is modelled by a brownian motion.

4.3. Evaluation of the adaptive strategy of Algorithm 2

We now evaluate the adaptive strategy presented in Algorithm 2. We focus on the second analytical function of Section 4.1 which is the most complex one. A small initial value of m is chosen, $m = 2$, and it is then increased when the variation of the Vorob'ev deviation remains smaller than a given threshold $\epsilon = 0.005$ during $n_0 = 4$ consecutive iterations (see Eq. (14) in Section 3.3). It allows to increase the dimension of the KL reduced space only when it is necessary to obtain a better accuracy. As illustrated on Figure 7 it allows to save simulations and reduce computational time. The accuracy reached with this strategy is similar to the one obtained with the strategy with fixed $m = 8$ but with a gain of $\approx 12\%$ in terms of computational time (Figure 8). We notice that the first iterations are performed with $m = 2$ and only the last iterations with $m = 8$.

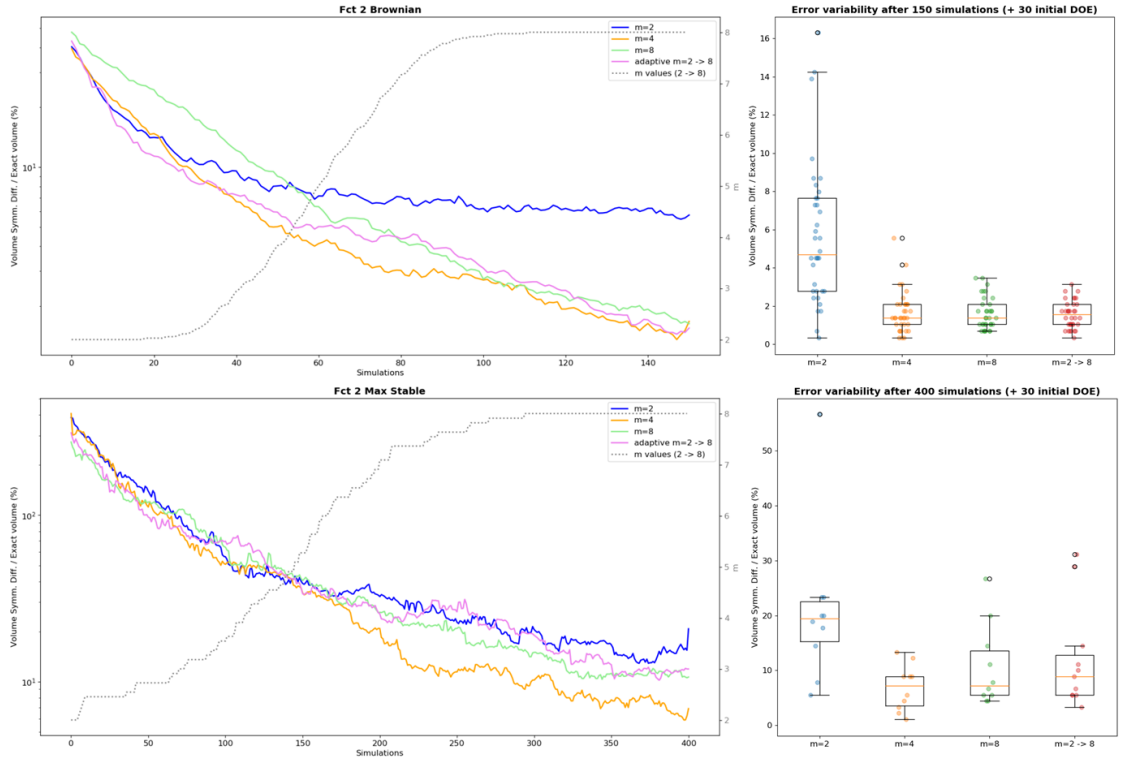

Figure 7: Analytical example 2 with brownian motion (top) and with max-stable process (bottom). Convergence of Algorithm 1 for $m = \{2, 4, 8\}$ and for adaptive choice of m value (Algorithm 2). Left: mean quality-ratio as function of the number of simulator calls in log scale. The dashed grey curve is the mean of m values in the case of an adaptive choice of its value. The mean is taken over the independent runs of initial RLHD. Right: boxplots for the quality-ratios associated with the different random initial DoEs at the maximal simulation budget.

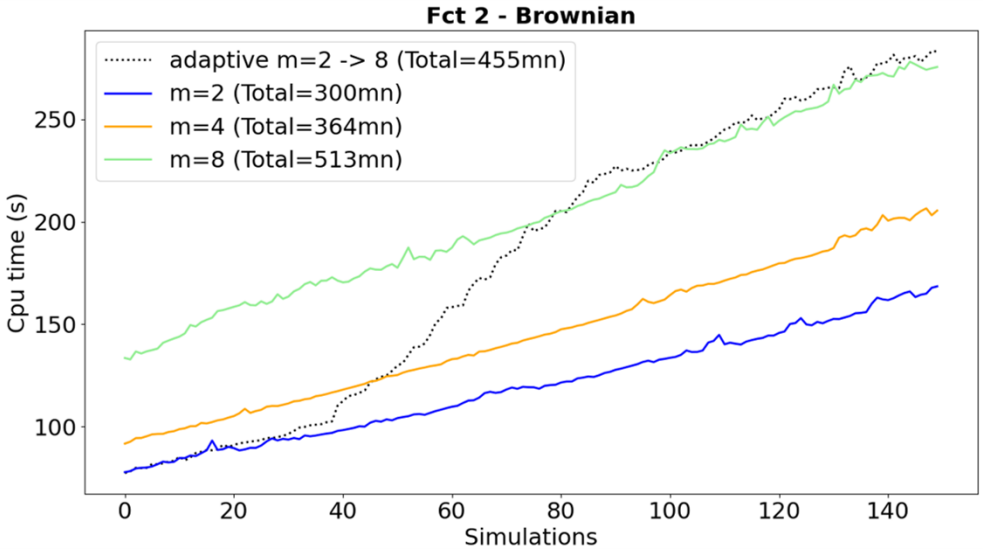

Figure 8: The computational time (sec.) needed to provide the next evaluation point as a function of iterations for the second analytic example with brownian motion. The values are averaged computational times for 5 runs of each strategy: $m = 2, 4, 8$ and adaptive choice of m value.

4.4. Application to a pollution control system SCR

In this section we test the proposed method on an automotive test case from IFPEN. The problem concerns an after-treatment device of diesel vehicles, called Selective Catalytic Reduction (SCR). This latter consists of a basic process of chemical reduction of nitrogen oxides (NOx) to diatomic nitrogen (N2) and water (H2O) by the reaction of NOx and ammonia NH3. 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 (unreacted ammonia) downstream of the catalyst, whereas under-dosing causes insufficient NOx 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. This latter takes as input the vehicle driving cycle profile and provides the time-series of corresponding exhaust emissions as output. A realistic SCR control law is used in this simulator (see [18] for more details). In this study, the two controlled variables are parameters of the SCR control law and lie in $\mathbb{X} = [0, 0.6]^2$. The random process describes the evolution of vehicle speed on $I = [0, 5400s]$ and is known through an available sample of 100 real driving cycles. A few samples are represented in Figure 9.

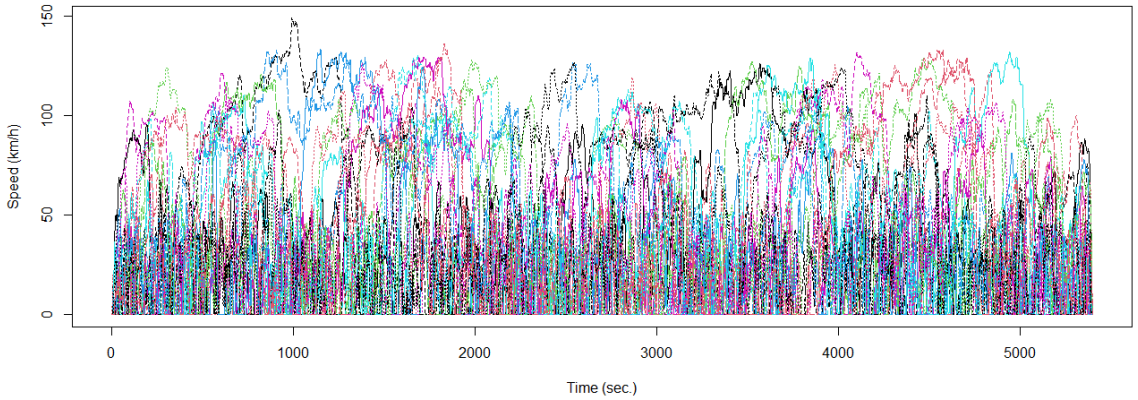

Figure 9: Seven real-driving cycles extracted from the available sample of size 100.

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

$$f : \begin{cases} \mathbb{X} \times \mathcal{V} & \to \mathbb{R} \\ (\mathbf{x}, \mathbf{V}) & \mapsto f(\mathbf{x}, \mathbf{V}) = \max_{t \in I} \mathtt{NH}_3(t) \end{cases} \tag{15}$$

We are interested in recovering the set $\Gamma^* = \{\mathbf{x} \in \mathbb{X}, g(\mathbf{x}) = \mathbb{E}_{\mathbf{V}}[f(\mathbf{x}, \mathbf{V})] \leq c\}$, with $c = 30ppm$. Conducting this study on a full grid would consists in covering the space $[0, 0.6]^2$ with a fine mesh and evaluating the code 100 times at each point. Knowing that each simulation takes about two minutes, such study would require many hours of computational time, and thus using meta-models allows to tackle this computational issue.

Truncation argument	$m = 15$	$m = 17$	$m = 19$	$m = 21$	$m = 23$	$m = 25$	$m = 64$
Explained variance	63.11%	66.64%	69.64%	72.45%	74.86%	76.87%	95.32%

Table 2: SCR pollution control system. The Variance explained by the KL decomposition according to m for the random cycle process $\mathbf{V}$.

In this industrial case, the explained variance grows very slowly as shown in the Table 2. Therefore, to represent at best the driving cycle $\mathbf{V}$ one should consider a high-dimensional threshold ($m = 64$). In order to avoid the curse of dimensionality involved by a high choice of m, we will adopt the adaptive strategy introduced in section 3.3.2 and tested in subsection 4.3 on a toy problem. More precisely, we start modeling the numerical simulator in a reduced space by taking $m = 15$. We increase the dimension of the uncertain space as soon as the condition in Eq. (14) is satisfied. We set the condition parameters at $\epsilon^{\text{SUR}} = 5 \times 10^{-3}$ and $n_0^{\text{SUR}} = 4$. We consider a Gaussian Process prior $Z_{(\mathbf{x},\mathbf{u})}$, with constant mean function and Matérn covariance kernel with $\nu = 5/2$. The initial DoE consists of a $n = 60$ points LHS design optimized with respect to the maximin criterion. The covariance kernel hyper-parameters are estimated by maximizing the likelihood. Figure 10 shows the so-called excursion probability function defined by $\mathbf{x} \mapsto \mathbb{P}(Y_{\mathbf{x}}^n \leq c)$, with $Y_{\mathbf{x}}^n$ the integrated Gaussian Process $Y_{\mathbf{x}}$ conditionally to the n available observations. The initial estimate of Γ^* is given by the green set with blue boundary. In the following and as for the analytical examples, we proceed to add one point at each iteration of the SUR strategy.

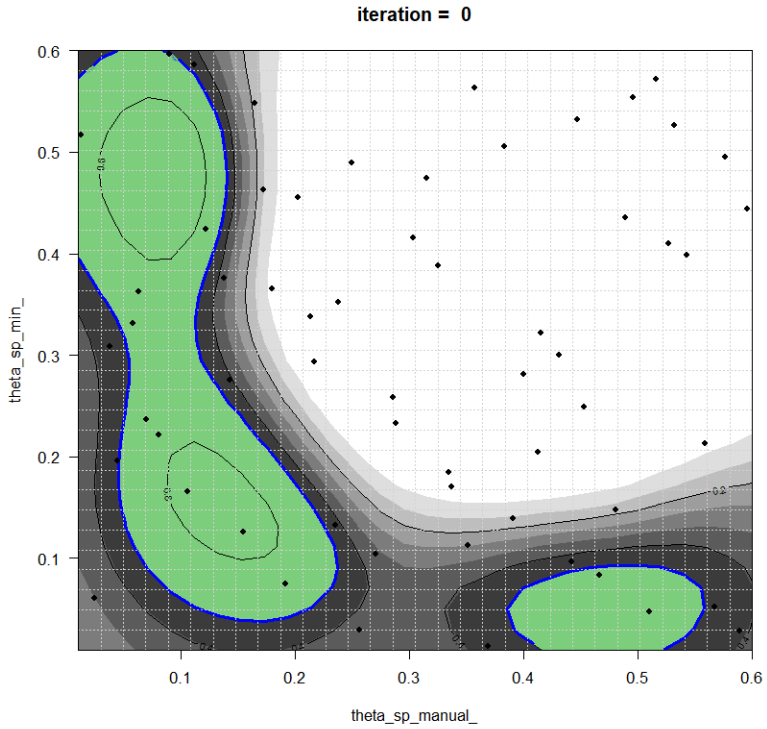

Figure 10: SCR pollution control system. The initial DoE (black triangles) and the initial estimate set (green). The contour plot in grey represents the excursion probability: darker corresponding to higher probability for the integrated process to be under the threshold.

Figure 11 shows the evolution of the Vorobev deviation and the dimension during the iterations. Initially, we reduce the Vorobev deviation based on GP modeling in a 17-dimensional joint space. At iteration 754, we increase the dimension m since we no longer improve the set estimation. This step of increasing the dimension allows us to reduce bias and to bring more information from the additional dimension. Fixing the maximal budget of iterations to 1000, we observe a stagnation of the Vorobev deviation from dimension $m = 24$. The estimated set at different iterations is given in Figure A.15 in Appendix. From Figure 12, we note that the SUR algorithm heavily visits the boundary region of Γ^* and explore also other potentially interesting regions. Actually, after 1000 iterations the whole domain $\mathbb{X}$ has an excursion probability close to either 0 or 1.

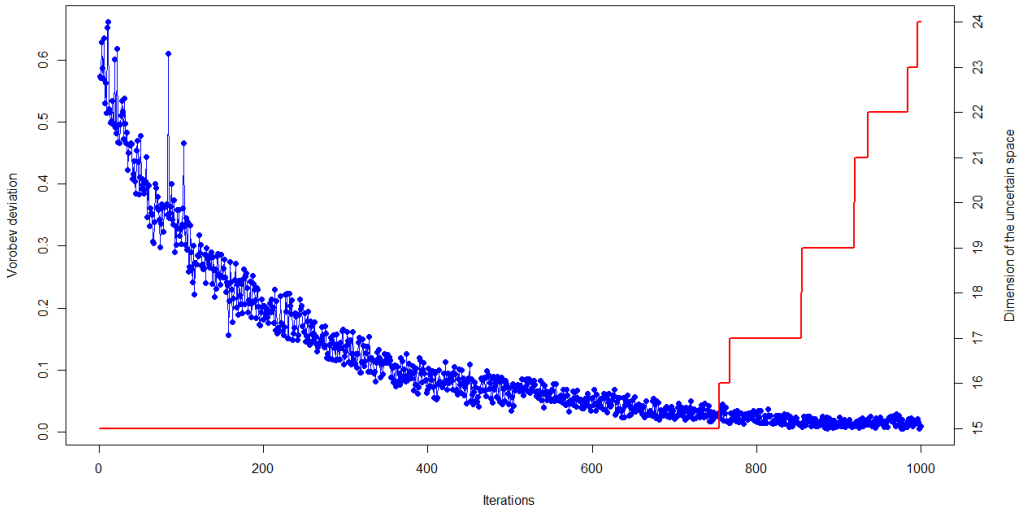

Figure 11: SCR pollution control system. The Vorob'ev deviation and the uncertain space dimension in function of the number of simulations.

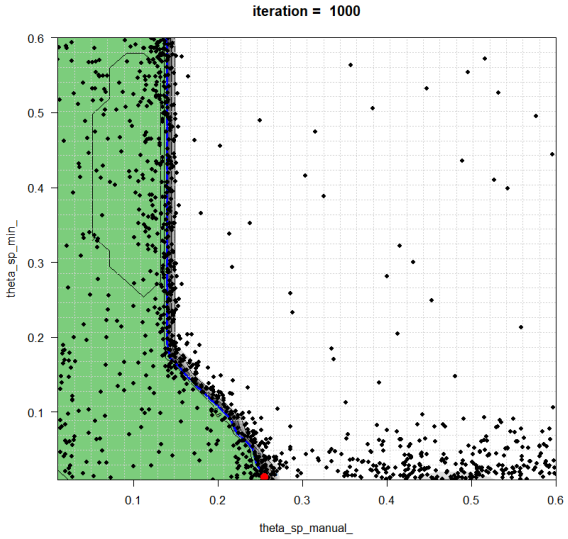

Figure 12: SCR pollution control system. The coverage probability function, the last proposed point by the algorithm (red point) and the estimate set (green set) after 1000 added points (black points). The contour plot in grey represents the excursion probability.

5. Conclusion

The aim of this paper is to propose a new feasible set estimation procedure for an automative control system in presence of functional uncontrolled variables modelled by a random process. Our procedure outperforms the one recently introduced in [9], as it requires less evaluations of the high-fidelity and expensive-to-evaluate model used to simulate the behaviour of the automative control system to achieve a similar accuracy.

Our procedure is based on a new enrichment strategy whose main ingredient is to fit a Gaussian Process model to the initial expensive-to-evaluate code in the joint input soace of controlled and uncontrolled variables. In our framework, the uncontrolled variable is modelled by a random process $\mathbf{V}$ to which a preliminary step of dimension reduction is applied. Moreover our knowledge of $\mathbf{V}$ is limited to a finite set of realizations Ξ. Note that our approach guarantees robustness with respect to the order of reduction. Indeed, partial knowledge of the full variability of $\mathbf{V}$ is recovered by evaluating the simulator on design points in $\mathbb{X}\times\Xi$. More precisely, at each step of the enrichment procedure, the simulator is evaluated in the point in $\mathbb{X} \times \Xi$ whose projection corresponds to the point obtained by optimizing the criterion in the truncated joint input space. The enrichment procedure guides the sampling toward informative regions for the feasible set, allowing by the way an accurate estimation with less evaluations of the expensive-to-evaluate code. We also propose a variation of our strategy, in which the order of reduction of $\mathbf{V}$ is increased adaptively. This approach consists in increasing the order of reduction sequentially, only when necessary, leading to even more computational savings.

Two bi-dimensional analytical examples are considered for which a reference solution can be computed. Then our procedure is validated by computing the quality-ratio defined as the ratio of the volume of the symmetric difference between the reference set and the estimated one to the volume of the true set. On these analytical examples, our procedure outperforms the one in [9] by achieving similar accuracy with much less evaluations of the expensive-to-evaluate code. Finally, we apply our procedure to an industrial problem related to the pollution control system of an automotive. A feasible set solution is found within a reasonable number of simulations.

The paper focuses on a formulation of the excursion set involving a unique constraint and where the uncertainty is summarized via its expectation. Nevertheless, as perspective, other reliability measures may also be of great interest. For example, one may be interested in ensuring a certain level of reliability with high probability or in considering multiple constraints, e.g., on the mean and the variance.

References

[1] Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. Statistics and Computing, 22(3):773–793, 2012.

[2] Clément Chevalier, Julien Bect, David Ginsbourger, Emmanuel Vazquez, Victor Picheny, and Yann Richet. Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. Technometrics, 56(4):455–465, 2014.

[3] David Bolin and Finn Lindgren. Excursion and contour uncertainty regions for latent gaussian models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77(1): 85–106, 2015.

[4] Joshua P French, Stephan R Sain, et al. Spatio-temporal exceedance locations and confidence regions. The Annals of Applied Statistics, 7(3):1421–1449, 2013.

[5] Clément Chevalier and David Ginsbourger. Fast computation of the multi-points expected improvement with applications in batch selection. In International Conference on Learning and Intelligent Optimization, pages 59–69. Springer, 2013.

[6] Emmanuel Vazquez and Julien Bect. A sequential bayesian algorithm to estimate a probability of failure. IFAC Proceedings Volumes, 42(10):546–550, 2009.

[7] Janis Janusevskis and Rodolphe Le Riche. Simultaneous kriging-based estimation and optimization of mean response. Journal of Global Optimization, 55(2):313–336, 2013.

[8] Brian J Williams, Thomas J Santner, and William I Notz. Sequential design of computer experiments to minimize integrated response functions. Statistica Sinica, pages 1133–1152, 2000.

[9] Mohamed Reda El Amri, Céline Helbert, Olivier Lepreux, Miguel Munoz Zuniga, Clémentine Prieur, and Delphine Sinoquet. Data-driven stochastic inversion via functional quantization. Statistics and Computing, 30(3):525–541, 2020. doi: 10.1007/s11222-019-09888-8. URL https://doi.org/10.1007/s11222-019-09888-8.

[10] Olivier Le Maître and Omar M. Knio. Spectral Methods for Uncertainty Quantification. Scientific Computation. Springer, Dordrecht, 2010. doi: 10.1007/978-90-481-3520-2.

[11] Ilya Molchanov. Theory of random sets. Springer Science & Business Media, 2006.

[12] O Yu Vorob'ev. Srednemernoje modelirovanie (mean-measure modelling), 1984.

[13] Oleg Yu. Vorobyev and Natalia A. Lukyanova. A mean probability event for a set of events. Mpra paper, University Library of Munich, Germany, 2013. URL https://EconPapers.repec.org/RePEc:pra:mprapa:48101.

[14] Clément Chevalier. Fast uncertainty reduction strategies relying on Gaussian process models. Theses, Universität Bern, September 2013. URL https://theses.hal.science/tel-00879082.

[15] Olivier Roustant, David Ginsbourger, and Yves Deville. Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. Journal of Statistical Software, 51(1):1–55, 2012. URL http://www.jstatsoft.org/v51/i01/.

[16] Walter Mebane Jr. and Jasjeet Sekhon. Genetic optimization using derivatives: The rgenoud package for r. Journal of Statistical Software, 42(11):1–26, 2011. ISSN 1548-7660. doi: 10.18637/jss.v042.i11. URL https://www.jstatsoft.org/v042/i11.

[17] Martin Schlather. Models for stationary max-stable random fields. Extremes, 5(1):33–44, 2002.

[18] Anthony Bonfils, Yann Creff, Olivier Lepreux, and Nicolas Petit. Closed-loop control of a scr system using a nox sensor cross-sensitive to nh3. IFAC Proceedings Volumes, 45(15):738–743, 2012.

Appendix A. Discussion on the GP model on the finite-dimensional truncated space

We discuss here the assumption stated in Section 2.2 that $f(\mathbf{x}, \mathbf{v})$ is a realization of a Gaussian Process $Z_{(\mathbf{x},\mathbf{u})}$ defined on the truncated space $\mathbb{X} \times \mathbb{R}^m$. Considering a m-truncation of the random process KL expansion, we reduce the hyperspace on which the GP is defined. Let us consider two truncation arguments m and $L > m$, with L large enough to ensure that the part of variance explained by the KL terms indexed by $i > L$ is negligible. For a given realization $\mathbf{v}$ of $\mathbf{V}$, let us introduce the notation $(\mathbf{u}, \tilde{\mathbf{u}}) \in \mathbb{R}^m \times \mathbb{R}^{L-m}$ where $\mathbf{u} = (< \mathbf{v}, \hat{\psi}_1 >, \ldots, < \mathbf{v}, \hat{\psi}_m >)^\top$ and $\tilde{\mathbf{u}} = (< \mathbf{v}, \hat{\psi}_{m+1} >, \ldots, < \mathbf{v}, \hat{\psi}_L >)^\top$. In that setting $f(\mathbf{x}, \mathbf{V})$ can be expressed as

$$f(\mathbf{x}, \mathbf{V}) = f(\mathbf{x}, \hat{\mathbf{V}}_L) + \epsilon_T = f\big(\mathbf{x}, (\mathbf{U}, \tilde{\mathbf{U}})\hat{\mathbf{\Phi}}_{\mathbf{L}}\big) + \epsilon_T$$

where $\hat{\mathbf{V}}_L$ is the empirical version (estimated from C^N) of the KL approximation of $\mathbf{V}$ given by (4) (replacing m by L), $\hat{\mathbf{\Phi}}_L = (\hat{\psi}_1, ..., \hat{\psi}_L)^\top$ and ϵ_T is the error associated to the KL truncation and empirical approximation, supposed small by construction.

Then, the best L^2-approximation of $f\big(\mathbf{x}, (\mathbf{U}, \tilde{\mathbf{U}})\hat{\mathbf{\Phi}}_L\big)$ by a measurable function of $\mathbf{U}$ only is the conditional expectation $\mathbb{E}_{\tilde{\mathbf{U}}}\big[f\big(\mathbf{x}, (\mathbf{U}, \tilde{\mathbf{U}})\hat{\mathbf{\Phi}}_L\big)|\mathbf{U}\big]$. We thus write:

$$f(\mathbf{x}, \mathbf{V}) = \mathbb{E}_{\tilde{\mathbf{U}}}\big[f\big(\mathbf{x}, (\mathbf{U}, \tilde{\mathbf{U}})\hat{\mathbf{\Phi}}_L\big)|\mathbf{U}\big] + \epsilon_P + \epsilon_T$$

with ϵ_P the L^2-projection error. We can further approximate the conditional expectation by

$$f\big(\mathbf{x}, (\mathbf{U}, \tilde{\mathbf{u}}(\mathbf{U}))\hat{\mathbf{\Phi}}_L\big) + \epsilon_E$$

where $\tilde{\mathbf{u}}(\mathbf{U})$ is one realization of $\tilde{\mathbf{U}}|\mathbf{U}$ and ϵ_E accounts for the expectation approximation. The latter approximation is motivated by the fact that, since $\mathbf{V}$ is only known through a finite sample, we only have access to one $\tilde{\mathbf{u}}(\mathbf{u})$ realization for each $\mathbf{u}$ corresponding to $\mathbf{v}$ in the initial finite set Ξ. Thus we can write:

$$f(\mathbf{x}, \mathbf{V}) = f\big(\mathbf{x}, (\mathbf{U}, \tilde{\mathbf{u}}(\mathbf{U}))\hat{\mathbf{\Phi}}_L\big) + \epsilon \tag{A.1}$$

with $\epsilon = \epsilon_T + \epsilon_P + \epsilon_E$. According to this last equation, the modelling assumption in Section 2.2 should include a noise term. However, the estimation of this heteroscedastic noise comes with an extra estimation cost and as it can be seen in Figure A.13, no significant model improvement is observed. Indeed in Figure A.13, for $m = 2$, we present the evolution of the symmetric difference for the noisy GP model $Z_{(\mathbf{x},\mathbf{u})}$ introduced from equation (A.1) when the noise ϵ is Gaussian and heteroscedastic with a variance function of $(\mathbf{x}, \mathbf{u})$:

$$\tau^2(\mathbf{x}, \mathbf{u}) = Var_{\tilde{U}}\big[f\big(\mathbf{x}, \big(\mathbf{u}, \tilde{\mathbf{U}}(\mathbf{u})\big)\hat{\mathbf{\Phi}}_L\big)\big|\mathbf{U} = \mathbf{u}\big].$$

Moreover, supposing $\mathbf{V}$ Gaussian or "nearly Gaussian", that is assuming that $\tilde{\mathbf{U}}$ can be considered in first approximation as independent of $\mathbf{U}$, then $\tau^2(\mathbf{x}, \mathbf{u})$ can be estimated by

$$\hat{\tau}^2(\mathbf{x}, \mathbf{u}) = \sum_{k=1}^{l} w_k\big[f\big(\mathbf{x}, \mathbf{V}_k^{Quant}\big) - \sum_{j=1}^{l} w_j f\big(\mathbf{x}, \mathbf{V}_j^{Quant}\big)\big]^2$$

where $l = 5$ and the $\mathbf{V}_k^{Quant}$ are greedy functional quantizers and w_k associated weights (see [9] for more details). These quantizers are built from a set of N curves $\{(\mathbf{u}, \tilde{\mathbf{u}}_k)\hat{\mathbf{\Phi}}_L, k = 1, ..., N\}$ where $\tilde{\mathbf{u}}_k$ are independent samples of $\tilde{\mathbf{U}}$ which in practice are uniformly sampled in the finite set $\bar{\mathcal{G}}_{m,L} = \{\bar{\mathbf{u}}_1, ..., \bar{\mathbf{u}}_N\}$ where $\bar{\mathbf{u}}_i = (< \breve{v}_i, \hat{\psi}_{m+1} >, ..., < \breve{v}_i, \hat{\psi}_L >)$. Numerically we select 20 $(\mathbf{x}, \mathbf{u})$-points from the initial DoE set of size $n = 30$ and estimate the corresponding $\hat{\tau}^2$. To avoid further estimation of τ^2 at new locations (the remaining DoE points and during the infill strategy), we build a second GP model of $\log(\hat{\tau}^2)$ based on the 20 initial estimations. Finally the noisy GP model Z is built using as noise variance $\exp\left(\hat{\log}(\hat{\tau}^2)\right)$. Overall we need additional $l \times 20 = 100$ costly evaluations of f to estimate the heteroscedastic noise.

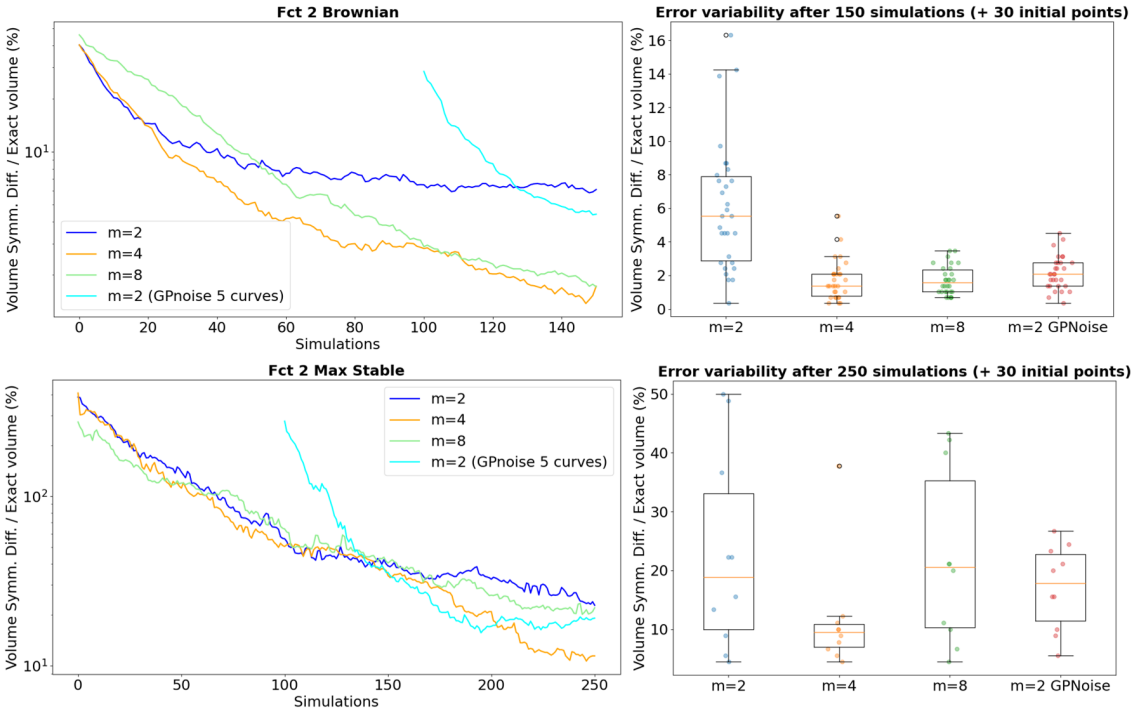

Figure A.13: Function 2 with brownian (top) and max-stable processes (bottom) with a comparison with the heteroscedastic GP model. Convergence of Algorithm 1 for $m = \{2, 4, 8\}$. Left: mean of the symmetric difference vs. number of simulator calls. The mean is taken over the independent runs of initial RLHD. The additional curve (cyan) corresponds to $m = 2$ with the heteroscedastic model, it is translated to take into account the extra-cost of 100 simulations for the noise estimation. Right: symmetric differences associated with the random initial DoEs at the maximal simulation budget.

In Figure A.13 we notice that compared to the homoscedastic model with $m = 2$, the model with heteroscedastic noise achieves a faster symmetric difference volume reduction but the overcost, for the variance estimation, makes this approach interesting only for a large simulation budget: at least 130 simulations. For the brownian case, on function 2, the homoscedastic models with higher m still perform better for a budget up to 150 than the heteroscedastic one. A model with a small m, that is to say with a rough truncation error, involves a larger bias. Nevertheless, refining the heteroscedastic noise estimation should bring the method to a similar level but much further on the axis corresponding to the number of simulations. But on function 2 with a max-stable process, the heteroscedastic model slightly outperforms the homoscedastic models ($m = 2, 4, 8$) when approaching the 150 simulations (Figure A.13). We can understand this improvement by the fact that even with higher m a homoscedastic model does not make up for a wider truncation

error which is better approximated by a heteroscedastic model. Note that it is possible to relax the "nearly Gaussian" hypothesis on $\mathbf{V}$. In that case the same kind of heteroscedastic variance estimator could be used but would require an empirical estimation of the conditional distribution of $\tilde{\mathbf{U}}|\mathbf{U}$ which seems difficult in the context of our partial knowledge of $\mathbf{V}$ imposing on us to work on finite predefined sets $\mathcal{G}$ and $\bar{\mathcal{G}}_{m,L}$.

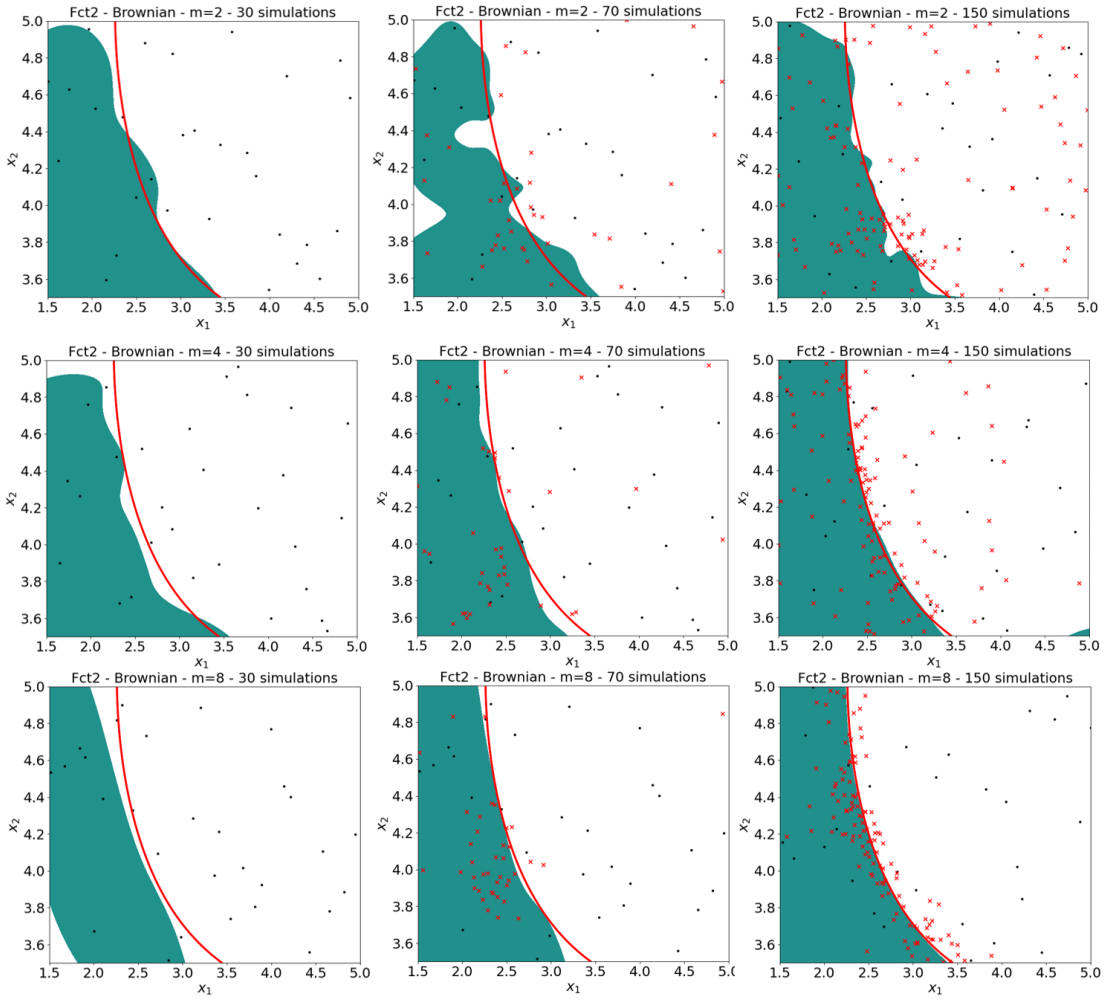

Figure A.14: Feasible domain estimation for analytical example 2 with brownian motion in green and its boundary in red for 3 different iterations (30, 70 and 150 from left to right) and for the 3 values of $m = 2$, 4 and 8 (from top to bottom). The black dots are the $\mathbf{x}$ coordinates of the points in the initial design of experiments, the red crosses are the additional points chosen by the algorithm.

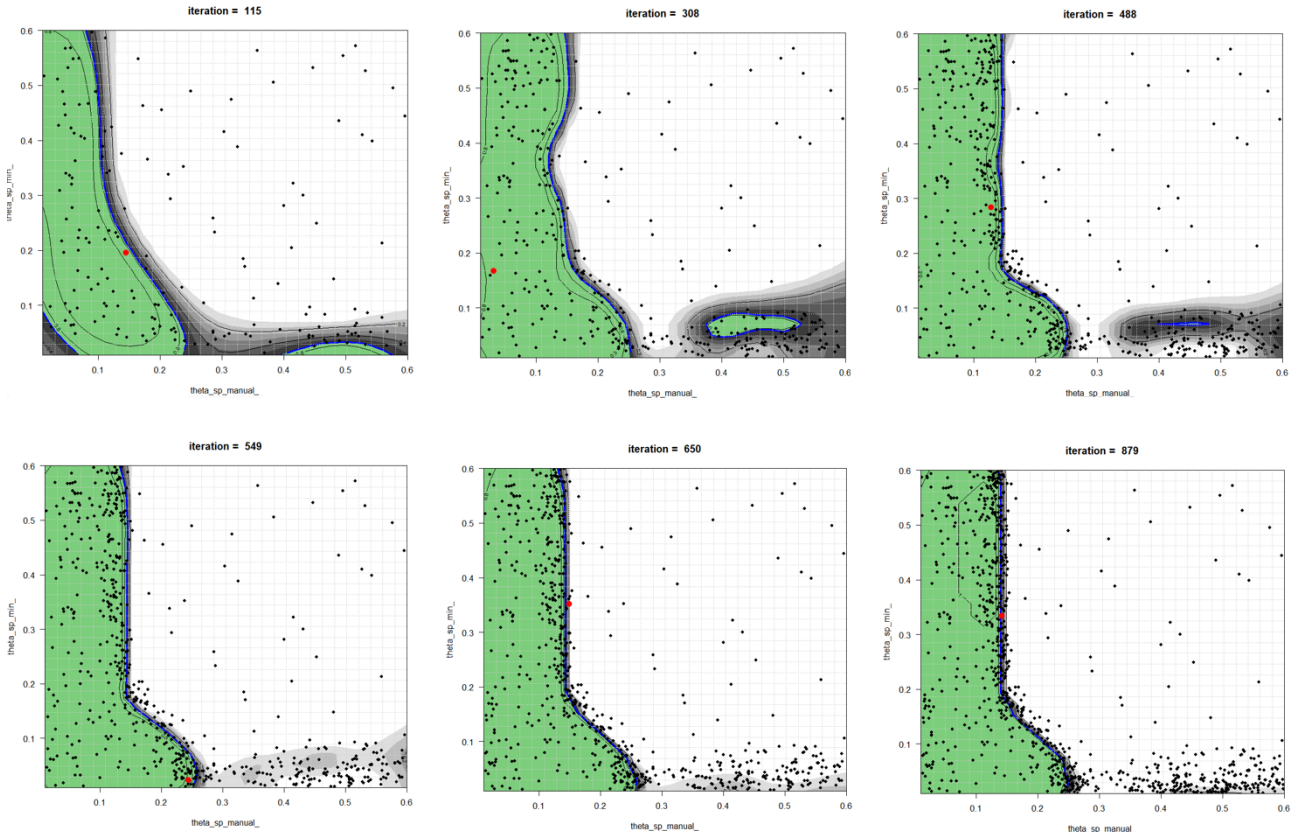

Figure A.15: SCR pollution control system. The estimated feasible domain at 6 different iterations.