Bayesian sequential design of computer experiments for quantile set inversion
Romain Ait Abdemalek-Lomenech, Julien Bect, Vincent Chabridon, Emmanuel Vazquez

To cite this version:
Romain Ait Abdemalek-Lomenech, Julien Bect, Vincent Chabridon, Emmanuel Vazquez. Bayesian sequential design of computer experiments for quantile set inversion. 2023. hal-03835704v3

HAL Id: hal-03835704
https://centralesupelec.hal.science/hal-03835704v3
Preprint submitted on 7 Nov 2023

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.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives| 4.0 International License
Bayesian sequential design of computer experiments for quantile set inversion

Romain Ait Abdellmalek-Lomenech*1, Julien Bect1, Vincent Chabridon2, and Emmanuel Vazquez1

1Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 91190 Gif-sur-Yvette, France
2EDF R&D, 6 Quai Watier, 78401 Chatou, France

Abstract

We consider an unknown multivariate function representing a system—such as a complex numerical simulator—taking both deterministic and uncertain inputs. Our objective is to estimate the set of deterministic inputs leading to outputs whose probability (with respect to the distribution of the uncertain inputs) of belonging to a given set is less than a given threshold. This problem, which we call Quantile Set Inversion (QSI), occurs for instance in the context of robust (reliability-based) optimization problems, when looking for the set of solutions that satisfy the constraints with sufficiently large probability. To solve the QSI problem, we propose a Bayesian strategy based on Gaussian process modeling and the Stepwise Uncertainty Reduction (SUR) principle, to sequentially choose the points at which the function should be evaluated to efficiently approximate the set of interest. We illustrate the performance and interest of the proposed SUR strategy through several numerical experiments.

Keywords: Gaussian processes, Active learning, Design of computer experiments, Stepwise Uncertainty Reduction, Set inversion, Uncertainty quantification.

*Corresponding author.

The authors gratefully acknowledge the National French Research Agency (ANR) for funding this work in the context of the SAMOURAI project (ANR-20-CE46-0013).
1 Introduction

When dealing with a numerical model of a physical phenomenon or a system, one is often interested in estimating the set of input parameters leading to outputs in a given range. Such set inversion problems (Jaulin and Walter, 1993) arise in a large variety of frameworks. In particular, “robust” formulations of the set inversion problem, in which some inputs are considered uncertain, have appeared recently in the literature, with applications to nuclear safety (Chevalier, 2013; Marrel et al., 2022), flood defense optimization (Richet and Bacchi, 2019) and pollution control systems (El Amri et al., 2023).

Following Richet and Bacchi (2019), we focus on a robust formulation of the set inversion problem that we call quantile set inversion (QSI). We consider a system modeled by an unknown continuous function $f : X \times S \to \mathbb{R}^q$, where $X$ and $S$ are bounded subsets of $\mathbb{R}^d_X$ and $\mathbb{R}^d_S$, corresponding to the sets of admissible values for the deterministic and uncertain (or stochastic) input variables of the system. We model the uncertain inputs by a random vector $S$ with known distribution $P_S$ on $S$. Then, given a subset $C \subset \mathbb{R}^q$ of the output space and a threshold $\alpha \in (0, 1)$, our objective is to estimate the set

$$\Gamma(f) = \{x \in X : P(f(x, S) \in C) \leq \alpha\}.$$  

(1)

Using the language of machine learning, we can also formulate the QSI problem as that of learning a classifier $X \to \{0, 1\}$ as close as possible to the indicator function $1_{\Gamma(f)}$.

The QSI problem occurs for instance in the context of robust (reliability-based, a.k.a. chance-constrained) optimization problems, when looking for the set of solutions that violate the constraints with sufficiently small probability—where, with our notations, the constraints are violated when $f(x, S)$ belongs to the critical region $C$.

**Remark 1** When $q = 1$ and $C = (T, +\infty)$, the set (1) can be rewritten as

$$\Gamma(f) = \{x \in X : Q_{1-\alpha}(f(x, S)) \leq T\},$$

(2)

where $Q_{1-\alpha}(f(x, S))$ denotes the quantile of order $1-\alpha$ of $f(x, S)$, with $S \sim P_S$. More generally, $\Gamma(f)$ can be seen as a quantile of the random set $\{x \in X : f(x, S) \in C\}$ in the sense of Molchanov (1991)—hence our choice of terminology.
The result of an evaluation of \( f \) at a point \((X_n, S_n)\) is a possibly noisy observation

\[
Z_n^{\text{obs}} = f(X_n, S_n) + \epsilon_n, \tag{3}
\]

the noiseless setting \((\epsilon_n = 0)\) corresponding to the case of a deterministic simulator. When the numerical model \( f \) is computationally expensive, it is important to estimate \( \Gamma(f) \) using only a small number of evaluations of \( f \). With this constraint in mind, we propose in this article a sequential Bayesian strategy based on the \textit{Step-wise Uncertainty Reduction} (SUR) principle (Vazquez and Bect, 2009; Villemonteix et al., 2009; Bect et al., 2012; Chevalier et al., 2014, ...). The starting point of a SUR strategy is to view \( f \) as a sample path of a random process, in practice a Gaussian process (GP). Then, at each step, an evaluation point is chosen by minimization of the expected future uncertainty on the quantity or object of interest—a set in the present case—given past observations.

The structure of the article is as follows: in the next two sections, we present our framework and a brief overview of the literature on Bayesian set inversion strategies, with a particular emphasis on SUR approaches. The core contribution of this article is given in Section 4, which presents the construction of a SUR sampling criterion for the QSI problem. Section 5 demonstrates the performance of our approach on numerical examples, including an application to history matching. In Section 6, we summarize our conclusions and provide perspectives for further research.

\textbf{Nota bene.} The authors have become aware, at the occasion of the SIAM Conference on Uncertainty Quantification (UQ22) in Atlanta, of related research work conducted by Charlie Sire (IRSN, France) and co-authors (Sire, 2022). The research presented in this article has been carried out independently of theirs.

\section{Framework and notations}

In the following, we consider a function \( f : U \to \mathbb{R} \), where \( U = X \) or \( U = X \times S \), depending on whether there are stochastic input variables or not. We adopt a Bayesian approach to sequentially choose the evaluation points \( U_1, U_2, \ldots \in U \) of \( f \) and estimate \( \Gamma(f) \) from evaluation results. It is assumed that we observe, at each selected point \( U_n \), a response \( Z_n^{\text{obs}} = f(U_n) + \epsilon_n \), where the \( \epsilon_n \) are independent zero-mean Gaussian random variables, with zero variance in the case of a deterministic
simulator. As a prior for the unknown function $f$, we consider a GP model (see, e.g., Santner et al., 2019; Rasmussen and Williams, 2006)—in other words, we assume that $f$ is a sample path of a GP. We denote by $\xi$ this process, and by $\mu$ and $k$ its mean and covariance functions.

Denote by $P_n$ the conditional probability given $I_n = \{(U_1, Z_{\text{obs}}^1), \ldots, (U_n, Z_{\text{obs}}^n)\}$, the information available at step $n$. Bayesian strategies employ a sampling criterion, also referred to as an acquisition function, denoted as $J_n$ or $G_n$. This criterion is based on the distribution of $\xi$ under $P_n$, and it is used at each step $n$ to sequentially select the next evaluation point from $U$. Specifically, we choose $U_{n+1}$ as an element in $U$ that minimizes $J_n$ or maximizes $G_n$:

$$U_{n+1} \in \arg\min_{u \in U} J_n(u) \quad \text{or} \quad U_{n+1} \in \arg\max_{u \in U} G_n(u).$$

In the following sections, two families of such criteria are reviewed.

**Notations.** $E_n = E(\cdot | I_n)$ denotes the conditional expectation associated with $P_n$. $\mu_n(u)$ and $\sigma_n(u)$ stand for the conditional (posterior) mean and standard deviation of $\xi(u)$. $p_n(u) = P_n(\xi(u) \in C)$ is the conditional (posterior) probability that $\xi(u)$ belongs to $C$.

### 3 Overview of Bayesian strategies for set inversion

#### 3.1 Maximal uncertainty sampling

We review in this section a first family of sampling criteria, which corresponds to the general idea of *maximal uncertainty sampling*, i.e., sampling at the location $x \in X$ where the uncertainty about $\mathbb{1}_C(\xi(x))$ and/or $\xi(x)$ is maximal. The literature on such criteria only deals, to the best of our knowledge, with the deterministic case $U = X$, when $f$ is a real-valued function ($q = 1$), and when $C = (T, +\infty)$, for a given $T \in \mathbb{R}$. In this setting, the set inversion problem reduces to the estimation of the set

$$\gamma(f) = \{x \in X : f(x) \leq T\}. \quad (4)$$

A natural approach to this problem is to select the point at which the probability of misclassification is maximal (Bryan et al., 2005), leading to the sampling criterion $G_n(x) = \min (p_n(x), 1 - p_n(x))$. Maximizing this criterion leads to selecting the next
point \( X_{n+1} \) such that \( p_n(X_{n+1}) \) is as close as possible to \( 1/2 \). Several equivalent criteria lead to the same choice of sampling point, including the variance or the entropy of the indicator \( 1_C(\xi(x)) \), or the sampling criterion proposed by Echard et al. (2011).

Some sampling criteria operate a trade-off between the posterior variance of \( \xi \) and its proximity to the threshold \( T \). This is the case, for instance, for the family of criteria defined by \( G_n(x) = E_n \left[ \max \left( 0, (\kappa \sigma_n(x))^{\delta} - |\xi(x) - T|^\delta \right) \right] \), with \( \kappa > 0 \), introduced separately by Bichon et al. (2008) (\( \delta = 1 \)) and Ranjan et al. (2008) (\( \delta = 2 \)). Similarly, Bryan et al. (2005) proposed the straddle heuristic, in which the sampling criterion is \( G_n(x) = 1.96 \sigma_n(x) - |\mu_n(x) - T| \).

### 3.2 Stepwise uncertainty reduction

SUR strategies (see Bect et al., 2019, and references therein) are a special case of the Bayesian approach in which the evaluation points are sequentially chosen by minimizing the expected future uncertainty about the object of interest. More precisely, a SUR strategy starts by defining a measure of uncertainty \( H_n \), at each step \( n \), that depends on the currently available information \( I_n \). Then, a sampling criterion \( J_n \) is built by considering the expectation of \( H_{n+1} \) conditional on \( I_n \), for a given choice of \( U_{n+1} = u \):

\[
J_n(u) = E_n \left[ H_{n+1} \mid U_{n+1} = u \right].
\]

Notice that \( H_{n+1} \) depends on the unknown outcome of the evaluation at \( u \) and that \( J_n(u) \) is an expectation over this random outcome. Instead of minimizing the criterion \( J_n \), one can also maximize the information gain \( G_n(u) = H_n - J_n(u) \).

We now give more details and first focus on the case of deterministic inversion (\( U = X \)). Several approaches have been developed in the past years. For instance, Bect et al. (2012) suggest the integrated probability of misclassification:

\[
H_n = \int_X \min(p_n(x), 1 - p_n(x)) \, dx,
\]

and the integrated posterior variance of \( 1_{\gamma(\xi)}(x) \): \( H_n = \int_X p_n(x)(1 - p_n(x)) \, dx \) as uncertainty measures.

Picheny et al. (2010) propose a targeted Integrated Mean Square Error (tIMSE) based on the uncertainty measure: \( H_n = \int_X \sigma_n^2(x) W_n(x) \, dx \), where \( W_n(x) = E_n[g(\mu_n(x) - T)] \), with \( g \) a kernel (typically Gaussian or uniform).
For additional examples of uncertainty measures and corresponding SUR criteria applicable to the deterministic set inversion problem, refer to Chevalier et al. (2013), Chevalier (2013), Marques et al. (2018), Azzimonti et al. (2021), and Duhamel et al. (2023).

To conclude this section, let us mention two formulations of the set inversion problem, which are related to—but distinct from—the QSI problem. First, Chevalier (2013) considers the task of estimating the set \( \{ x \in X : \max_{s \in S} f(x, s) \leq T \} \). In this setting, the proposed uncertainty measure is \( H_n = \int_X p_n^\forall(x)(1 - p_n^\forall(x))dx \), where \( p_n^\forall(x) = \mathbb{P}_n(\max_{s \in S} \xi(x, s) \leq T) \). Second, in the work of El Amri et al. (2023), the objective instead is to estimate the set \( \{ x \in X : \mathbb{E}(f(x, S)) \leq T \} \). To this end, the authors propose a hybrid SUR strategy to choose, sequentially, the deterministic component \( x \) and stochastic component \( s \) of each new evaluation point.

4 Construction of a SUR strategy for QSI

4.1 Sampling criterion

Our objective is now to estimate the set \( \Gamma(f) \) defined by (1) using evaluation results modeled by (3). In the following, we construct a sampling criterion for the QSI problem. Consider the random process

\[
\tau(x) = \int_{S} 1_{C}(\xi(x, s)) \, d\mathbb{P}_S(s) = \mathbb{P}_S(\{s \in S : \xi(x, s) \in C\}),
\]

which corresponds, for each \( x \in X \), to the (random) volume with respect to \( \mathbb{P}_S \) of the excursion of the process \( \xi(x, \cdot) \) in \( C \). Notice that \( \Gamma(\xi) \) can be written as

\[
\Gamma(\xi) = \{ x \in X : \tau(x) \leq \alpha \}.
\]

Assume that a sequence \( (\hat{\Gamma}_n)_{n \geq 1} \) of estimators of \( \Gamma(\xi) \) has been chosen. Then we propose to use as uncertainty measure the expected volume of the symmetric difference (see Figure 1) between \( \Gamma(\xi) \) and its estimator:

\[
\mathcal{H}_n = \mathbb{E}_n \left[ \lambda(\Gamma(\xi) \Delta \hat{\Gamma}_n) \right], \quad n \geq 1,
\]

where \( \lambda \) is the usual (Lebesgue) volume measure on \( \mathbb{R}^d \).
The SUR strategy derived from (9) consists in minimizing, at each step, the criterion
\[ J_n(x, s) = \mathbb{E}_n [\mathcal{H}_{n+1} \mid (X_{n+1}, S_{n+1}) = (x, s)]. \]  
(10)
Several approaches using the symmetric difference have been used in different contexts (see, e.g., Chevalier, 2013; Azzimonti et al., 2021).

When considering the Bayes-optimal estimator
\[ \hat{\Gamma}_n = \left\{ x \in \mathcal{X} : \pi_n(x) > \frac{1}{2} \right\}, \]  
(11)
where \( \pi_n(x) = \mathbb{P}_n(\tau(x) \leq \alpha) \), the uncertainty measure \( \mathcal{H}_n \) can be expressed as the integrated probability of misclassification associated to the classifier \( \mathbb{1}_{\hat{\Gamma}_n} \), leading to
\[ \mathcal{H}_n = \int_{\mathcal{X}} \min(\pi_n(x), 1 - \pi_n(x)) \, dx. \]  
(12)
The proof of this simple result can be found in the supplementary material. As a consequence, the SUR strategy derived from (9) consists in minimizing at each step the criterion
\[ J_n(\tilde{x}, \tilde{s}) = \mathbb{E}_n [\mathcal{H}_{n+1} \mid (X_{n+1}, S_{n+1}) = (\tilde{x}, \tilde{s})] \]
\[ = \int_{\mathcal{X}} \mathbb{E}_n [\min(\pi_{n+1}(x), 1 - \pi_{n+1}(x)) \mid (X_{n+1}, S_{n+1}) = (\tilde{x}, \tilde{s})] \, dx. \]  
(13)
Here, notice that we use the new notations \( \tilde{x} \) and \( \tilde{s} \), since \( x \) is now used as integration variable.

Other choices for the uncertainty measure \( \mathcal{H}_n \) are possible. Notably, we could use any increasing transformation of \( \min(\pi_n(x), 1 - \pi_n(x)) \) in the integral (12) to
define the measure. In particular, we can construct a variance-based measure $H_n = \int_X \pi_n(x)(1-\pi_n(x)) \, dx$, and an entropy-based one $H_n = -\int_X \pi_n(x) \log_2(\pi_n(x)) \, dx - \int_X (1-\pi_n(x)) \log_2(1-\pi_n(x)) \, dx$.

For the sake of clarity, in the following, we focus on the misclassification-based QSI-SUR strategy described by (13). However, a brief comparative benchmark using the different possible uncertainty measures will be given in Section 5.4.

**Remark 2** It is instructive to compare the different misclassification-based and variance-based criteria to those proposed by Bect et al. (2012), and to notice the formal resemblance, if replacing $\pi_n$ by $p_n$ and $\Gamma(\xi)$ by $\gamma(\xi)$.

### 4.2 Approximation of the criterion

In this section, we discuss the numerical approximation of $J_n(\hat{x}, \hat{s})$, defined by (13), at a given point $(\hat{x}, \hat{s}) \in X \times S$. We emphasize that the same methodology could be applied to the variance-based and entropy-based criteria.

Two difficulties arise for the numerical evaluation of $J_n(\hat{x}, \hat{s})$. The first and obvious one lies in the approximation of the integral over $X$. A possible solution, which is applied in this work, is to estimate the integral (with respect to the uniform distribution on $X$) using importance sampling. This allows to non-uniformly sample an approximation grid $\tilde{X}$ for our integral in order, for instance, to concentrate the sampled points in uncertain areas of $\Gamma(\xi)$. More precisely, given a random a finite collection $\tilde{X}$ of elements of $X$ sampled from a density $p_X$, we use an importance sampling approximation:

$$J_n(\hat{x}, \hat{s}) \approx \sum_{x \in \tilde{X}} \frac{1}{p_X(x)} E_n \left[ \min(\pi_{n+1}(x), 1-\pi_{n+1}(x)) \mid (X_{n+1}, S_{n+1}) = (\hat{x}, \hat{s}) \right]. \quad (14)$$

The second difficulty comes from the fact that the integrand

$$E_n \left[ \min(\pi_{n+1}(x), 1-\pi_{n+1}(x)) \mid (X_{n+1}, S_{n+1}) = (\hat{x}, \hat{s}) \right] \quad (15)$$

does not admit—to the best of our knowledge—a closed-form expression as in Chevalier et al. (2014), because $\tau$ is not a GP. We propose to estimate (15) using quantization of the distribution $P_S$ together with Monte Carlo simulations of the process $\xi$, in the spirit of Villemonteix et al. (2009) and Bect et al. (2012).
More specifically, consider a finite subset \( \tilde{S} \) of \( S \), and a family \((w_S(s))_{s \in \tilde{S}}\) of positive real numbers such that \( P_{\tilde{S}} = \sum_{s \in \tilde{S}} w_S(s) \delta_s \) is a “good” approximation of \( P_S \), where \( \delta_s \) denotes the Dirac measure at \( s \). This can be achieved, for instance, by defining \( \tilde{S} \) as a collection of \( n_S \) i.i.d. samples from \( P_S \) and fixing \( w_S(s) = \frac{1}{n_S} \) for all \( s \in \tilde{S} \). For other methods, the reader can refer to Graf and Luschgy (2000) for more information about quantization.

Moreover, let \( \{z_1, \ldots, z_N\} \) and \((w_\xi(z_i))_{i \in \{1, \ldots, N\}}\) be such that \( \sum_{i=1}^N w_\xi(z_i) \delta_{z_i} \) is a quantization of the distribution of \( \xi(\tilde{x}, \tilde{s}) \) given \( I_n \), and recall that \( \tilde{X} \subset X \) is the finite subset used for the approximation of the integral over \( X \) arising in \( J_n(\tilde{x}, \tilde{s}) \). Thanks to the Gaussian nature of \( \xi \), assuming that \( \tilde{X} \times \tilde{S} \) is not too large, we can easily simulate \( M \) sample paths \( \{\xi_{i,1}, \ldots, \xi_{i,M}\} \) of \( \xi \) over \( \tilde{X} \times \tilde{S} \), under the conditional distribution \( P_n(\cdot | \xi(\tilde{x}, \tilde{s}) = z_i) \).

Given a point \( x \in \tilde{X} \), set

\[
\hat{\pi}_{i+1}^i(x) = \frac{1}{M} \sum_{m=1}^M 1_{[0,\alpha]} \left( \sum_{s \in \tilde{S}} w_S(s) 1_C(\xi_{i,m}(x,s), s) \right). \tag{16}
\]

Then, for a sufficiently large \( M \) and a “good” quantization \( P_{\tilde{S}} \), we have

\[
\hat{\pi}_{i+1}^i(x) \approx P(\tau(x) \leq \alpha | I_n, \xi(\tilde{x}, \tilde{s}) = z_i). \tag{17}
\]

As a consequence, it is possible to use

\[
\tilde{J}_n(\tilde{x}, \tilde{s}) = \sum_{i=1}^N w_\xi(z_i) \min(\hat{\pi}_{i+1}^i(x), 1 - \hat{\pi}_{i+1}^i(x)) \tag{18}
\]

as an approximation of \( (15) \).

Finally, combining (14) and (18), the criterion \( J_n(\tilde{x}, \tilde{s}) \) is then approximated by

\[
\tilde{J}_n(\tilde{x}, \tilde{s}) = \sum_{x \in \tilde{X}} \frac{1}{p_X(x)} \tilde{J}_n^x(\tilde{x}, \tilde{s}). \tag{19}
\]

Algorithm 1 summarizes the numerical procedure.

**Remark 3** For a better numerical efficiency, the simulations of the sample paths of \( \xi \) under \( P_n(\cdot | \xi(\tilde{x}, \tilde{s}) = Z_i) \) are preferably carried out using reconditioning of sample paths. A description of this procedure is given by Villemonteix et al. (2009), Section 5.1.
Algorithm 1 Procedure to approximate $J_n$

For each candidate point $(\tilde{x}, \tilde{s})$:
1. Generate $\{z_1, \ldots, z_N\}$ and $(w_\xi(z_i))_{i \in \{1, \ldots, N\}}$, a quantization of the distribution of $\xi(\tilde{x}, \tilde{s})$ with respect to $P_n$.
2. For $i = 1, \ldots, N$:
   2.1 Generate $M$ sample paths of $\xi$ on $\tilde{X} \times \tilde{S}$ according to $P_n(\cdot | \xi(\tilde{x}, \tilde{s}) = z_i)$.
   2.2 Compute $\bar{\pi}^i_{n+1}(x)$ using (16), for all $x \in \tilde{X}$.
3. Compute the approximation $\bar{J}_n(\tilde{x}, \tilde{s})$ of the criterion according to (19)

5 Numerical experiments

5.1 Artificial examples

To assess the performances of the sampling criterion in Section 4, we focus first on three artificial examples, with scalar output values ($q = 1$) and noise-free observations.

The first test function (see Figure 2), defined on $X = [0; 10]$ and $S = [0; 15]$, is a modified Branin-Hoo function:

$$f_1(x, s) = \frac{1}{12} b(x, s) + 3 \sin \left(\frac{x^3}{3}\right) + \sin \left(\frac{s^3}{3}\right),$$

(20)

where $b$ is the Branin-Hoo function $b(x, s) = \left(s - \frac{5.1x^2}{4\pi^2} + \frac{5x}{\pi} - 6\right)^2 + 10 \left(1 - \frac{1}{8\pi}\right) \cos(x) + 10$ (Branin and Hoo, 1972). We take $C = (-\infty; T]$ with $T = 7.5$, $\alpha = 0.05$ and $P_S$ the Beta distribution with parameters (7.5, 1.9), rescaled from $[0, 1]$ to $S$. The associated set $\Gamma(f_1)$ is represented in Figure 3.

The second test function is defined on $X = [-2; 2]^2$ and $S = [-1; 1]^2$ by

$$f_2((x_1, x_2), (s_1, s_2)) = \frac{1}{2} c(x_1, s_1) + \frac{1}{2} c(x_2, s_2),$$

(21)

with $c$ the “six-hump camel” function $c(x, s) = \left(4 - 2.1x^2 + \frac{x^4}{4}\right)x^2 + xs + (4s^2 - 4)s^2$ (Dixon and Szegö, 1978). We take $C = (-\infty; T]$ with $T = 1.2$, $\alpha = 0.15$ and $P_S$ the uniform distribution on $S$. The corresponding set $\Gamma(f_2)$ is shown in Figure 3.

As a third test function $f_3$, we consider the Hartmann4 function, in the form of Picheny et al. (2013), defined on $X = [0; 1]^2$ and $S = [0; 1]^2$ equipped with the uniform distribution. We define $\alpha = 0.6$ and $C = (-\infty, T]$, with $T = 1.1$. 

11
Figure 2: Representation of $f_1$ (right) and probability density function associated to $P_S$ (left). The black curve represents the contour of $\gamma(f)$.

Figure 3: Representation of $x \mapsto \mathbb{P}\{s \in S : f(x, s) \in C\}$ and $\Gamma(f)$ for the artificial test functions $f_1$ (left) and $f_2$ (right).
5.2 Application to history matching

The objective of this application is to retrieve the set of plausible deterministic input variables of a numerical simulator, given real-life measurements. Such problem can be seen as a particular case of a “history matching” problem (see, e.g., Williamson et al., 2013). More precisely, we consider an uncertain Mogi model (Mogi, 1958), whose aim is to simulate the displacement at the surface of a volcano caused by an underground magma reservoir, while taking into account the mechanical property of the soil (Durrande and Le Riche, 2017).

Formally, the simulator can be seen as a function $v : X \times S \rightarrow \mathbb{R}^{220}$, with inputs $(x_1, x_2, x_3, x_4, x_5) \in X = [0, 1]^5$ representing the normalized latitude, longitude, elevation, radius and overpressure of the magma source and $(s_1, s_2) \in S = [0, 1]^2$ representing uncertain perturbations of the shear modulus and Poisson ratio of the material. The uncertain variables are assumed independent and identically distributed, following a Beta distribution with parameters $(2, 2)$.

Given the real measurements $(y_i)_{i \in \{1, \ldots, 220\}}$ and considering the mean absolute error $V(x, s) = \frac{1}{220} \sum_{i=1}^{220} |v_i(x, s) - y_i|$, our objective is to retrieve the set of plausible parameters for the Mogi model. More specifically, in this application, a vector of parameters $x \in X$ is considered to be “plausible” if it yields an error strictly less than 0.015 with a probability larger than 0.1. To exhibit the direct link between this history matching problem and the QSI framework, notice that it can be equivalently reformulated as the problem of estimating the set $\Gamma(V)$, with critical region $C = [0.015, +\infty)$ and $\alpha = 0.9$.

5.3 Implementation details

We use for all the examples a GP prior with an unknown constant mean and an anisotropic Matérn covariance function. At each step of the algorithm, the parameters of the Gaussian prior are estimated using the restricted maximum likelihood (ReML) method (see, e.g., Stein, 1999), with the constraint that the regularity parameter $\nu$ of the kernel belongs to $\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, +\infty\}$, where the limit case $\nu \rightarrow +\infty$ corresponds to the Gaussian kernel.

Note that in what follows, all sampling are assumed to be without replacement. At each step, we first sample $n_S$ points according to $P_S$ as a discretization grid $\tilde{S}$ of $S$. Then, a set of $n_X$ points is uniformly sampled in $X$ (with $n_X = 500d_X$ in
Table 1: Summary of the simulation parameters used in the numerical experiments. Note that \( \text{card}(\tilde{X} \times \tilde{S}) = n_\pi n_S = 1500 \) for all the examples (cf. Remark 4).

<table>
<thead>
<tr>
<th>Test case</th>
<th>( \alpha )</th>
<th>( d_X + d_S )</th>
<th>( n_0 )</th>
<th>( n_X )</th>
<th>( n_S )</th>
<th>( n_\pi )</th>
<th>( n_c )</th>
<th>( N )</th>
<th>( M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_1 )</td>
<td>0.05</td>
<td>1 + 1</td>
<td>20</td>
<td>500</td>
<td>100</td>
<td>15</td>
<td>200</td>
<td>15</td>
<td>250</td>
</tr>
<tr>
<td>( f_2 )</td>
<td>0.15</td>
<td>2 + 2</td>
<td>40</td>
<td>1000</td>
<td>25</td>
<td>60</td>
<td>200</td>
<td>15</td>
<td>250</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>0.60</td>
<td>2 + 2</td>
<td>40</td>
<td>1000</td>
<td>25</td>
<td>60</td>
<td>200</td>
<td>15</td>
<td>250</td>
</tr>
<tr>
<td>“Volcano”</td>
<td>0.90</td>
<td>5 + 2</td>
<td>70</td>
<td>2500</td>
<td>75</td>
<td>20</td>
<td>200</td>
<td>15</td>
<td>250</td>
</tr>
</tbody>
</table>

Remark 4 Due to the expensive nature of the simulation of conditional GPs, the simulation parameters described in this section must be calibrated according to the
computational budget at hand. In particular \( n_\pi \) and \( n_S \) must be chosen such that \( \text{card}(\tilde{X} \times \tilde{S}) = n_\pi n_S \) is sufficiently low to accommodate a given computation time.

5.4 Results

We compare the results obtained with the proposed QSI-SUR sampling criterion with several other Bayesian methods. Due to the absence in the literature of strategies dealing with the exact problem considered in this work, we focus on methods that aim at approximating the set \( \gamma(f) = \{(x, s) \in X \times S : f(x, s) \in C\} \). It is clear indeed that a perfect knowledge of \( \gamma(f) \) implies a perfect knowledge of \( \Gamma(f) \).

More precisely, we compare our results to those obtained with the misclassification probability criterion (see Section 3.1) and with the SUR criteria corresponding to the uncertainty measure (6) for the random set \( \gamma(\xi) \). As a baseline method, we also include the results obtained with random sampling of the evaluation points. More details about the implementation of these methods can be found in the supplementary material. The different algorithms are repeated 100 times on all the test cases, using different pseudo\(^1\)-maximin LHS as initial designs.

For performances evaluation, we compare the proportions of misclassified points obtained on a prediction grid composed of the product of a Sobol’ sequence of \( 2^{13} \) points in \( X \) and the inversion (with respect to the cumulative distribution function of \( P_S \)) of a Sobol sequence of \( 2^8 \) points in \([0, 1]^d_S\). Considering that the Bayes-optimal estimator (11) is too expensive to compute on such a large grid (using heavy Gaussian sample paths simulations on the prediction grid), we use instead the estimator \( \hat{\Gamma}_n = \{x \in X : E_n(\tilde{\tau}(x)) \leq \alpha\} \), where \( \tilde{\tau}(x) \) is the approximation of \( \tau(x) \) defined by average over the \( 2^8 \) selected points of \( S \).

From the results in Figure 4 and Figure 5, observe that the new sampling criteria tends to perform better on the QSI problem than the state-of-the-art methods focusing on the set \( \gamma(f) \) in the joint space \( X \times S \). Moreover, as illustrated in Figure 6 and in the complementary results of the supplementary material, the performances appear to never lag behind even in the worst cases.

These differences in performance can be explained, from a heuristic viewpoint, by the fact that, in the joint space \( X \times S \), the new criterion tend to concentrate

\(^1\)We call “pseudo-maximin” the best LHS, in the sense of the minimal distance between points, in a collection of 1000 independent LHSs.
Figure 4: Median of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test functions $f_1$ (left) and $f_2$ (right).

Figure 5: Median of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the function $f_3$ (left) and the “volcano” test case (right).
Figure 6: Quantiles of order 0.75 (left) and 0.95 (right) of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_1$.

the evaluations around specific zones of $\gamma(f)$ which are particularly relevant for the approximation of $\Gamma(f)$. This phenomenon, related to the geometry of $\gamma(f)$, can be visualized on Figure 7. It is important to notice, however, that in some cases (illustrated here by the function $f_3$), the performances of the two kinds of methods are similar.

From the results shown on Figure 8, we can observe that the performances of the misclassification-based and variance-based QSI-SUR criteria are similar. Such results are expected, given the fact that the variance and the entropy are both increasing transformation of the misclassification probability.

More results regarding these numerical experiments can be found in the supplementary material, including figures similar to Figure 6 and Figure 8 for $f_2$, $f_3$ and “volcano”.

6 Conclusion

This article presents a SUR strategy for a particular set inversion problem, in a framework where a function admits deterministic and uncertain input variables, that we called Quantile Set Inversion (QSI). The practical interest of the proposed method is illustrated on several problems, on which methods that do not take advantage of the
Figure 7: Examples of sequential designs (red dots) obtained after $n = 30$ iterations on the function $f_1$, with the four sampling criteria. Black dots represent the initial design points, black curve the boundary of $\gamma(f)$, and red curve the boundary of $\gamma(\mu_n)$. 
specificity of the QSI problem tend to be outperformed. However, this gain in performance comes at the cost of a high numerical complexity, in relation with the heavy use of conditioned Gaussian trajectory simulations. Future work will concentrate on making the method applicable to harder test problems (higher input dimensions, smaller probabilities, etc.) and on demonstrating its practical relevance to real-life problems.

Acknowledgements. The authors are grateful to Valérie Cayol, Rodolphe Le Riche, Nicolas Durrande and Victor Picheny for sharing their R implementation of the Mogi model used in Section 5.

References


Gaussian process based sequential design of experiments. *Bernoulli* 25(4A), 2883–2919.


parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics* 41, 1703–1729.
SM1 Proof of the expression of $\mathcal{H}_n$

Let us remark that

$$\Gamma(\xi)\Delta\hat{\Gamma}_n = \left\{ x \in \Gamma(\xi) : x \notin \hat{\Gamma}_n \right\} \cup \left\{ x \in \hat{\Gamma}_n : x \notin \Gamma(\xi) \right\}.$$ 

In consequence, by defining the classifier $c_n(x) = 1_{\hat{\Gamma}_n}(x)$ and by Fubini theorem:

$$E_n \left[ \lambda \left( \Gamma(\xi) \Delta\hat{\Gamma}_n \right) \right] = \int_X E_n \left[ 1_{\Gamma(\xi)\Delta\hat{\Gamma}_n}(x) \right] dx$$

$$= \int_X E_n \left[ 1_{\{c_n=0\}}(x) 1_{\Gamma(\xi)}(x) \right]$$

$$+ \int_X E_n \left[ 1_{\{c_n=1\}}(x) (1 - 1_{\Gamma(\xi)}(x)) \right] dx$$

$$= \int_X 1_{\{c_n=0\}}(x) \pi_n(x) dx + \int_X 1_{\{c_n=1\}}(x) (1 - \pi_n(x)) dx.$$

It suffices to observe that, if $\hat{\Gamma}_n = \{ x \in X : \pi_n(x) > \frac{1}{2} \}$, then $c_n(x) = 1_{\{\pi_n(x) > \frac{1}{2}\}}(x)$, and for all $x \in X$:

$$1_{\{c_n=0\}}(x)\pi_n(x) + 1_{\{c_n=1\}}(x) (1 - \pi_n(x))$$

$$= 1_{\{\pi_n \leq \frac{1}{2}\}}(x)\pi_n(x) + 1_{\{\pi_n > \frac{1}{2}\}}(x) (1 - \pi_n(x))$$

$$= \min(\pi_n(x), 1 - \pi_n(x))$$

to obtain the simplified expression of $\mathcal{H}_n$.

SM2 Implementation details

In order to have coherent implementations of the comparatives methods regarding the QSI-SUR sampling criterion, adequate procedures are adopted. For clarity, we
base the values of the simulation parameters involved on those used for the QSI-SUR implementation, of which we also recall the implementation.

In the following, all sampling are assumed to be without replacement.

- **Random sampling**: at each step, $f$ is evaluated at a random point sampled according to $\mathcal{U}(X) \otimes P_S$.

- **Misclassification**: at each step, we sample $n_\pi \times n_S$ points in $X \times S$ according to $\mathcal{U}(X) \otimes P_S$. We evaluate the function $f$ at the point with the highest misclassification probability $\min(p_n(x,s), 1 - p_n(x,s))$.

- **Joint-SUR**: at each step, we sample $n_X \times n_S$ points in $X \times S$ with respect to $\mathcal{U}(X) \otimes P_S$. For each sampled point, we evaluate $p_n(x,s)$. The point with the highest probability of misclassification in $C$ is kept and $n_X \times n_S - 1$ points are sampled according to $p_{X \times S} \propto \min(p_n(x,s), 1 - p_n(x,s))$, giving us the discretization grid $\tilde{D}$. The joint-SUR criterion is then optimized using a search over a set of candidate points composed of the $n_c$ first sampled points of $\tilde{D}$.

Notice that, similarly to the QSI-SUR implementation, an importance sampling approximation based on the distribution $p_{X \times S}$ is used for the evaluation of the integral on $X \times S$ involved in the sampling criterion.

- **QSI-SUR**: at each step, we first sample $n_S$ points according to $P_S$ as discretization grid $\tilde{S}$. We then uniformly sample a set of $n_X$ points in $X$. For each of those points, we approximate $\pi_n(x)$ using Monte Carlo simulation of $\xi(x, \cdot)$ on $\tilde{S}$ (with respect to $P_n$). Finally, the point with the maximum misclassification probability is kept and we sample $n_\pi - 1$ points according to $p_X(x) \propto \min(\pi_n(x), 1 - \pi_n(x))$. This procedure gives us a product set $\tilde{X} \times \tilde{S}$, used to approximate the integrals involved in the QSI-SUR sampling criterion. The approximated sampling criterion is then optimized using an exhaustive search over a subset of the approximation grid $\tilde{D} \subset \tilde{X} \times \tilde{S}$. To construct such subset of candidate points, we evaluate $p_n(x,s)$, and keep a random set of $n_c$ points composed of the point with the highest misclassification probability in $\gamma(\xi)$ and $n_c - 1$ points sampled according to the probability vector proportional to $\min(p_n(x), 1 - p_n(x))$ as $\tilde{D}$.
SM3  Complementary results for numerical experiments

SM3.1  Complementary results for the misclassification-based criterion

SM3.1.1  Artificial case $f_1$

![Figure SM1: Median and several quantiles of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_1$.](image-url)
Figure SM2: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_1$. 
Figure SM3: Different trajectories of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_1$. 
SM3.1.2 Artificial case $f_2$

Figure SM4: Median and several quantiles of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_2$. 
Figure SM5: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_2$. 
Figure SM6: Different trajectories of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_2$. 
SM3.1.3 Artificial case $f_3$

Figure SM7: Median and several quantiles of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_3$. 
Figure SM8: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_3$. 
Figure SM9: Different trajectories of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function $f_3$. 
Figure SM10: Median and several quantiles of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the “volcano” test case.
Figure SM11: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the “volcano” test case.
Figure SM12: Different trajectories of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the “volcano” test case.

SM3.2 Comparison between variants of $\mathcal{H}_n$
Figure SM13: Median of the proportion of misclassified points vs. number of iterations, for 100 repetitions of different possible variants (in terms of $H_n$) of the QSI-SUR method on the function $f_1$, $f_2$, $f_3$ and the “volcano” test case (from left to right, top to bottom).