



**HAL**  
open science

# Bayesian sequential design of computer experiments for quantile set inversion

Romain Ait Abdelmalek-Lomenech, Julien Bect, Vincent Chabridon,  
Emmanuel Vazquez

## ► To cite this version:

Romain Ait Abdelmalek-Lomenech, Julien Bect, Vincent Chabridon, Emmanuel Vazquez. Bayesian sequential design of computer experiments for quantile set inversion. 2023. hal-03835704v2

**HAL Id: hal-03835704**

**<https://centralesupelec.hal.science/hal-03835704v2>**

Preprint submitted on 12 Jul 2023 (v2), last revised 5 Dec 2024 (v7)

**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 Abdelmalek-Lomenech<sup>\*1</sup>, Julien Bect<sup>1</sup>, Vincent Chabridon<sup>2</sup>, and Emmanuel Vazquez<sup>1</sup>

<sup>1</sup>Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 91190 Gif-sur-Yvette, France

<sup>2</sup>EDF 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.

---

<sup>\*</sup>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).

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Framework and notations</b>	<b>4</b>
<b>3</b>	<b>Overview of Bayesian strategies for set inversion</b>	<b>5</b>
3.1	Maximal uncertainty sampling . . . . .	5
3.2	Stepwise uncertainty reduction . . . . .	6
<b>4</b>	<b>Construction of a SUR strategy for QSI</b>	<b>7</b>
4.1	Sampling criterion . . . . .	7
4.2	Approximation of the criterion . . . . .	9
<b>5</b>	<b>Numerical experiments</b>	<b>11</b>
5.1	Artificial examples . . . . .	11
5.2	Application to history matching . . . . .	12
5.3	Implementation details . . . . .	13
5.4	Results . . . . .	15
<b>6</b>	<b>Conclusion</b>	<b>19</b>
	<b>References</b>	<b>20</b>
	<b>SUPPLEMENTARY MATERIAL</b>	<b>24</b>
	<b>SM1 Proof of the expression of <math>\mathcal{H}_n</math></b>	<b>24</b>
	<b>SM2 Implementation details</b>	<b>24</b>
	<b>SM3 Complementary results for numerical experiments</b>	<b>26</b>
	SM3.1 Complementary results for the misclassification-based criterion	26
	SM3.1.1 Artificial case $f_1$ . . . . .	26
	SM3.1.2 Artificial case $f_2$ . . . . .	29
	SM3.1.3 Artificial case $f_3$ . . . . .	32
	SM3.1.4 “Volcano” test case . . . . .	35
	SM3.2 Comparison between variants of $\mathcal{H}_n$ . . . . .	37

# 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  $f : \mathbb{X} \times \mathbb{S} \rightarrow \mathbb{R}^q$ , where  $\mathbb{X}$  and  $\mathbb{S}$  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  $\mathbb{P}_S$  on  $\mathbb{S}$ . 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

$$\Gamma(f) = \{x \in \mathbb{X} : \mathbb{P}(f(x, S) \in C) \leq \alpha\}. \quad (1)$$

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

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*

$$\Gamma(f) = \{x \in \mathbb{X} : q_{1-\alpha}(f(x, S)) \leq T\}, \quad (2)$$

where  $q_{1-\alpha}$  denotes the quantile of order  $1 - \alpha$ . More generally,  $\Gamma(f)$  can be seen as a quantile of the random set  $\{x \in \mathbb{X} : f(x, S) \in C\}$  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

$$Z_n^{\text{obs}} = f(X_n, S_n) + \epsilon_n, \quad (3)$$

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 *Stepwise 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.

**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.*

## 2 Framework and notations

In the following, we consider a function  $f : \mathbb{U} \rightarrow \mathbb{R}$ , where  $\mathbb{U} = \mathbb{X}$  or  $\mathbb{U} = \mathbb{X} \times \mathbb{S}$ , depending on whether there are stochastic input variables or not. We adopt a Bayesian approach to sequentially choose the evaluation points  $U_1, U_2, \dots \in \mathbb{U}$  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. We consider a GP prior denoted by  $\xi \sim \text{GP}(\mu, k)$  about the unknown function  $f$ , where  $\mu$  and  $k$  are mean and covariance functions (see, e.g., Santner et al., 2003; Rasmussen and Williams, 2006).

Denote by  $\mathbb{P}_n$  the conditional probability given  $\mathcal{I}_n = \{(U_1, Z_1^{\text{obs}}), \dots, (U_n, Z_n^{\text{obs}})\}$ , 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  $\mathbb{P}_n$ , and it is used at each step  $n$  to sequentially select the next evaluation point from  $\mathbb{U}$ . Specifically, we choose  $U_{n+1}$  as an element in  $\mathbb{U}$  that minimizes  $J_n$  or maximizes  $G_n$ :

$$U_{n+1} \in \underset{u \in \mathbb{U}}{\text{argmin}} J_n(u) \quad \text{or} \quad U_{n+1} \in \underset{u \in \mathbb{U}}{\text{argmax}} G_n(u).$$

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

**Notations.**  $\mathbb{E}_n = \mathbb{E}(\cdot | \mathcal{I}_n)$  denotes the conditional expectation associated with  $\mathbb{P}_n$ .  $\mu_n(u)$  and  $\sigma_n(u)$  stand for the conditional (posterior) mean and standard deviation of  $\xi(u)$ .  $p_n(u) = \mathbb{P}_n(\xi(u) \in C)$  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  $x \in \mathbb{X}$  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  $\mathbb{U} = \mathbb{X}$ , 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

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

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  $\mathbb{1}_C(\xi(x))$  (see Bryan et al. (2005)), 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  $G_n(x) = \mathbb{E}_n \left[ \max \left( 0, (\kappa \sigma_n(x))^\delta - |\xi(x) - T|^\delta \right) \right]$ , with  $\kappa > 0$  and  $\delta \in \{1, 2\}$ . Those sampling criteria, inspired by the expected improvement criterion for optimization (Mockus et al., 1978; Jones et al., 1998), were 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  $\mathcal{H}_n$ , at each step  $n$ , that depends on the currently available information  $\mathcal{I}_n$ . Then, a sampling criterion  $J_n$  is built by considering the expectation of  $\mathcal{H}_{n+1}$  conditional on  $\mathcal{I}_n$ , for a given choice of  $U_{n+1} = u$ :

$$J_n(u) = \mathbb{E}_n [\mathcal{H}_{n+1} \mid U_{n+1} = u]. \quad (5)$$

Notice that  $\mathcal{H}_{n+1}$  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  $G_n(u) = \mathcal{H}_n - J_n(u)$ .

We now give more details and first focus on the case of deterministic inversion ( $\mathbb{U} = \mathbb{X}$ ). Several approaches have been developed in the past years. For instance, Vazquez and Piera-Martinez (2006); Vazquez and Bect (2009) propose to estimate the volume of  $\gamma(f)$  using the uncertainty measure

$$\mathcal{H}_n = \left( \int_{\mathbb{X}} \sqrt{\min(p_n(x), 1 - p_n(x))} dx \right)^2, \quad (6)$$

where  $\min(p_n(x), 1 - p_n(x))$  is the probability of misclassification at  $x$  using the classifier  $\mathbb{1}_{\{p_n(x) \leq \frac{1}{2}\}}(x)$ . (The problem of estimating the volume of a set is directly linked to that of set inversion and they both boil down to the problem of estimating

the function  $x \mapsto \mathbb{1}_{\gamma(\xi)}(x)$ .) [Bect et al. \(2012\)](#) suggest other uncertainty measures, such as the integrated probability of misclassification:

$$\mathcal{H}_n = \int_{\mathbb{X}} \min(p_n(x), 1 - p_n(x)) dx, \quad (7)$$

and the integrated posterior variance of  $\mathbb{1}_{\gamma(\xi)}(x)$ :  $\mathcal{H}_n = \int_{\mathbb{X}} p_n(x)(1 - p_n(x)) dx$ . [Picheny et al. \(2010\)](#) propose a *targeted Integrated Mean Square Error* (tIMSE) based on the uncertainty measure:  $\mathcal{H}_n = \int_{\mathbb{X}} \sigma_n^2(x) W_n(x) dx$ , where  $W_n(x) = \mathbb{E}_n[g(\mu_n(x) - T)]$ , 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  $\{x \in \mathbb{X} : \max_{s \in \mathbb{S}} f(x, s) \leq T\}$ . In this setting, the proposed uncertainty measure is

$$\mathcal{H}_n = \int_{\mathbb{X}} p_n^\vee(x)(1 - p_n^\vee(x)) dx, \quad (8)$$

where  $p_n^\vee(x) = \mathbb{P}_n(\max_{s \in \mathbb{S}} \xi(x, s) \leq T)$ . Second, in the work of [El Amri et al. \(2023\)](#), the objective instead is to estimate the set  $\{x \in \mathbb{X} : \mathbb{E}(f(x, S)) \leq T\}$ . 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

$$\tau(x) = \int_{\mathbb{S}} \mathbb{1}_C(\xi(x, s)) d\mathbb{P}_S(s) = \mathbb{P}_S(\{s \in \mathbb{S} : \xi(x, s) \in C\}), \quad (9)$$



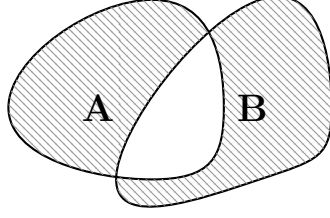


Figure 1: Symmetric difference  $A\Delta B = (A \setminus B) \cup (B \setminus A)$  between  $A$  and  $B$  (shaded area).

which corresponds, for each  $x \in \mathbb{X}$ , 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

$$\Gamma(\xi) = \{x \in \mathbb{X} : \tau(x) \leq \alpha\} . \quad (10)$$

Assume that a sequence  $(\widehat{\Gamma}_n)_{n \geq 1}$  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:

$$\mathcal{H}_n = \mathbb{E}_n \left[ \lambda(\Gamma(\xi) \Delta \widehat{\Gamma}_n) \right], \quad n \geq 1, \quad (11)$$

where  $\lambda$  is the usual (Lebesgue) volume measure on  $\mathbb{R}^{d_x}$ .

The SUR strategy derived from (11) consists in minimizing, at each step, the criterion

$$J_n(x, s) = \mathbb{E}_n [\mathcal{H}_{n+1} \mid (X_{n+1}, S_{n+1}) = (x, s)] . \quad (12)$$

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

$$\widehat{\Gamma}_n = \left\{ x \in \mathbb{X} : \pi_n(x) > \frac{1}{2} \right\}, \quad (13)$$

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  $\mathbb{1}_{\widehat{\Gamma}_n}$ , leading to

$$\mathcal{H}_n = \int_{\mathbb{X}} \min(\pi_n(x), 1 - \pi_n(x)) dx . \quad (14)$$

The proof of this simple result can be found in the supplementary material. As a consequence, the SUR strategy derived from (11) consists in minimizing at each step the criterion

$$\begin{aligned} J_n(\check{x}, \check{s}) &= \mathbb{E}_n [\mathcal{H}_{n+1} \mid (X_{n+1}, S_{n+1}) = (\check{x}, \check{s})] \\ &= \int_{\mathbb{X}} \mathbb{E}_n [\min(\pi_{n+1}(x), 1 - \pi_{n+1}(x)) \mid (X_{n+1}, S_{n+1}) = (\check{x}, \check{s})] dx. \end{aligned} \quad (15)$$

Here, notice that we use the new notations  $\check{x}$  and  $\check{s}$ , 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 (14) to define the measure. In particular, we can construct a *variance*-based measure  $\mathcal{H}_n = \int_{\mathbb{X}} \pi_n(x)(1 - \pi_n(x)) dx$ , and an *entropy*-based one  $\mathcal{H}_n = - \int_{\mathbb{X}} \pi_n(x) \log_2(\pi_n(x)) dx - \int_{\mathbb{X}} (1 - \pi_n(x)) \log_2(1 - \pi_n(x)) dx$ .

For the sake of clarity, in the following, we focus on the misclassification-based QSI-SUR strategy described by (15). 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  $J_n(\check{x}, \check{s})$ , defined by (15), at a given point  $(\check{x}, \check{s}) \in \mathbb{X} \times \mathbb{S}$ . 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  $J_n(\check{x}, \check{s})$ . The first and obvious one lies in the approximation of the integral over  $\mathbb{X}$ . A possible solution, which is applied in this work, is to estimate the integral (with respect to the uniform distribution on  $\mathbb{X}$ ) using importance sampling. This allows to non-uniformly sample an approximation grid  $\tilde{\mathbb{X}}$  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  $\tilde{\mathbb{X}}$  of elements of  $\mathbb{X}$  sampled from a density  $p_{\mathbb{X}}$ , we use an importance

sampling approximation:

$$J_n(\check{x}, \check{s}) \approx \sum_{x \in \tilde{\mathbb{X}}} \frac{1}{p_{\mathbb{X}}(x)} \mathbb{E}_n [\min(\pi_{n+1}(x), 1 - \pi_{n+1}(x)) \mid (X_{n+1}, S_{n+1}) = (\check{x}, \check{s})]. \quad (16)$$

The second difficulty comes from the fact that the integrand

$$\mathbb{E}_n [\min(\pi_{n+1}(x), 1 - \pi_{n+1}(x)) \mid (X_{n+1}, S_{n+1}) = (\check{x}, \check{s})] \quad (17)$$

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 (17) using quantization of the distribution  $\mathbb{P}_S$  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  $\tilde{\mathbb{S}}$  of  $\mathbb{S}$ , and a family  $(w_{\mathbb{S}}(s))_{s \in \tilde{\mathbb{S}}}$  of positive real numbers such that  $\mathbb{P}_{\tilde{\mathbb{S}}} = \sum_{s \in \tilde{\mathbb{S}}} w_{\mathbb{S}}(s) \delta_s$  is a “good” approximation of  $\mathbb{P}_S$ , where  $\delta_s$  denotes the Dirac measure at  $s$ . This can be achieved, for instance, by defining  $\tilde{\mathbb{S}}$  as a collection of  $n_{\mathbb{S}}$  i.i.d. samples from  $\mathbb{P}_S$  and fixing  $w_{\mathbb{S}}(s) = \frac{1}{n_{\mathbb{S}}}$  for all  $s \in \tilde{\mathbb{S}}$ . For other methods, the reader can refer to [Graf and Luschgy \(2000\)](#) for more information about quantization.

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

$$\tilde{\pi}_{n+1}^i(x) = \frac{1}{M} \sum_{m=1}^M \left( \sum_{s \in \tilde{\mathbb{S}}} w_{\mathbb{S}}(s) \mathbf{1}_C(\xi_{i,m}(x, s)) \right). \quad (18)$$

It follows from the law of large numbers, that

$$\tilde{\pi}_{n+1}^i(x) \xrightarrow[M \rightarrow \infty]{\text{a.s.}} \mathbb{P} \left( \xi(x, \tilde{S}) \in C \mid \mathcal{I}_n, \xi(\check{x}, \check{s}) = z_i \right), \quad (19)$$

with  $\tilde{S} \sim \mathbb{P}_{\tilde{\mathbb{S}}}$ . As a consequence, for a sufficiently large  $M$ , it is possible to use

$$j_n^x(\check{x}, \check{s}) = \sum_{i=1}^N w_{\xi}(z_i) \min(\tilde{\pi}_{n+1}^i(x), 1 - \tilde{\pi}_{n+1}^i(x)) \quad (20)$$

as an approximation of (17).

Finally, combining (16) and (20), the criterion  $J_n(\check{x}, \check{s})$  is then approximated by

$$\tilde{J}_n(\check{x}, \check{s}) = \sum_{x \in \tilde{\mathbb{X}}} \frac{1}{p_{\mathbb{X}}(x)} j_n^x(\check{x}, \check{s}). \quad (21)$$

Algorithm 1 summarizes the numerical procedure.

---

**Algorithm 1** Procedure to approximate  $J_n$

---

**For each candidate point**  $(\check{x}, \check{s})$ :

1. Generate  $\{z_1, \dots, z_N\}$  and  $(w_\xi(z_i))_{i \in \{1, \dots, N\}}$ , a quantization of the distribution of  $\xi(\check{x}, \check{s})$  with respect to  $\mathbb{P}_n$ .
  2. For  $i = 1, \dots, N$ :
    - 2.1 Generate  $M$  sample paths of  $\xi$  on  $\tilde{\mathbb{X}} \times \tilde{\mathbb{S}}$  according to  $\mathbb{P}_n(\cdot | \xi(\check{x}, \check{s}) = z_i)$ .
    - 2.2 Compute  $\tilde{\pi}_{n+1}^i(x)$  using (18), for all  $x \in \tilde{\mathbb{X}}$ .
  3. Compute the approximation  $\tilde{J}_n(\check{x}, \check{s})$  of the criterion according to (21)
- 

**Remark 3** *For a better numerical efficiency, the simulations of the sample paths of  $\xi$  under  $\mathbb{P}_n(\cdot | \xi(\check{x}, \check{s}) = z_i)$  are preferably carried out using reconditioning of sample paths. A description of this procedure is given by [Villemonteix et al. \(2009\)](#), Section 5.1.*

## 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  $\mathbb{X} = [0; 10]$  and  $\mathbb{S} = [0; 15]$ , is a modified Branin-Hoo function:

$$f_1(x, s) = \frac{1}{12}b(x, s) + 3 \sin\left(x^{\frac{5}{4}}\right) + 3 \sin\left(s^{\frac{5}{4}}\right), \quad (22)$$

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  $T = 7.5$ ,

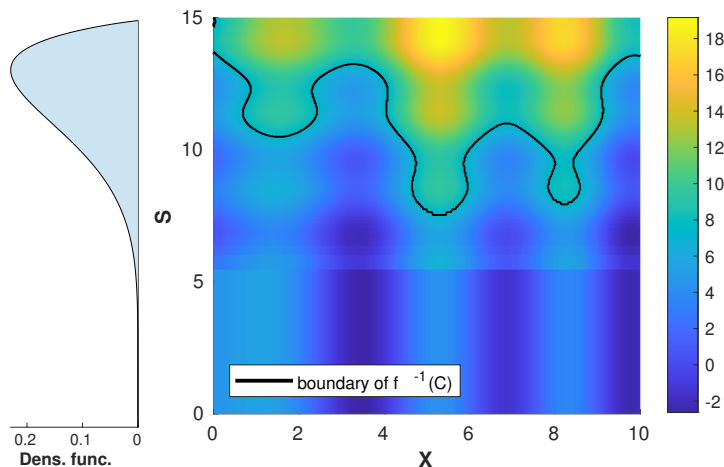


Figure 2: Representation of  $f_1$  (right) and probability density function associated to  $\mathbb{P}_S$  (left). The black curve represents the contour of  $\gamma(f)$ .

$\alpha = 0.05$  and  $\mathbb{P}_S$  the rescaled Beta distribution  $\mathcal{B}(7.5, 1.9)$  on  $\mathbb{S}$ . The associated set  $\Gamma(f_1)$  is represented in Figure 3.

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

$$f_2((x_1, x_2), (s_1, s_2)) = c(x_1, s_1) + c(x_2, s_2), \quad (23)$$

with  $c$  the “six-hump camel” function  $c(x, s) = \left(4 - 2.1x^2 + \frac{x^4}{3}\right)x^2 + xs + (4s^2 - 4)s^2$  (Dixon and Szegő, 1978). We take  $C = (-\infty; T]$  with  $T = 1.2$ ,  $\alpha = 0.15$  and  $\mathbb{P}_S$  the uniform distribution on  $\mathbb{S}$ . 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  $\mathbb{X} = [0; 1]^2$  and  $\mathbb{S} = [0; 1]^2$  equipped with the uniform distribution. We define  $\alpha = 0.6$  and  $C = (-\infty, T]$ , with  $T = 1.1$ .

## 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

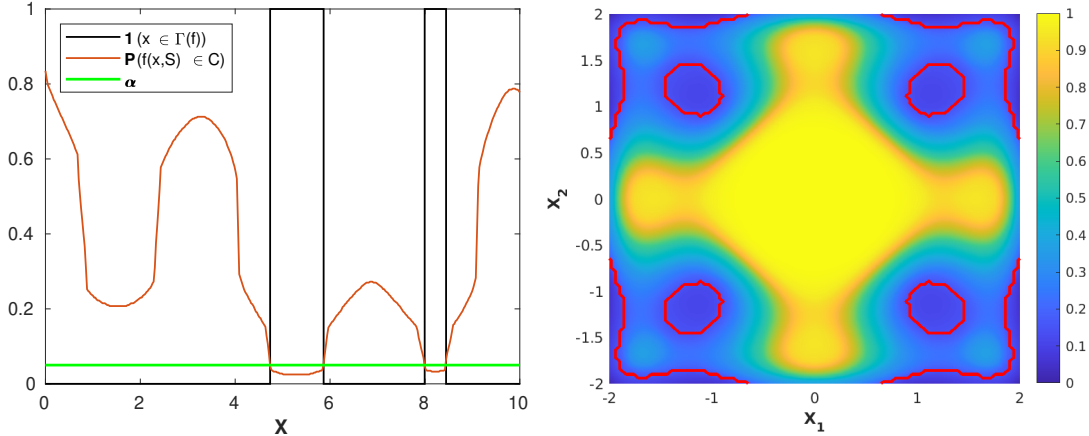


Figure 3: Representation of  $x \mapsto \mathbb{P}(\{s \in \mathbb{S} : f(x, s) \in C\})$  and  $\Gamma(f)$  for the artificial test functions  $f_1$  (left) and  $f_2$  (right).

the soil (Durrande and Le Riche, 2017).

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

Given the real measurements  $(y_i)_{i \in \{1, \dots, 220\}}$  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  $x \in \mathbb{X}$  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  $v$  of the kernel belongs to  $\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, +\infty\}$ , where the limit case  $v \rightarrow +\infty$  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  $\mathbb{P}_S$  as a discretization grid  $\tilde{\mathbb{S}}$  of  $\mathbb{S}$ . Then, a set of  $n_X$  points is uniformly sampled in  $\mathbb{X}$  (with  $n_X = 500d_X$  in our experiments). For each of these points, we approximate the misclassification probability  $\min(\pi_n(x), 1 - \pi_n(x))$  using Monte Carlo simulations of  $\xi(x, \cdot)$  on  $\tilde{\mathbb{S}}$  (with respect to  $\mathbb{P}_n$ ). Finally, a subset  $\tilde{\mathbb{X}}$  is constructed from the point with the highest misclassification probability, and a sample of  $n_\pi - 1$  points drawn, according to the discrete probability distribution  $p_X(x) \propto \min(\pi_n(x), 1 - \pi_n(x))$ . This particular distribution is chosen to ensure that the elements of  $\tilde{\mathbb{X}}$  are concentrated in the areas of  $\mathbb{X}$  where the probability of misclassification is high. This procedure gives us a product set  $\tilde{\mathbb{X}} \times \tilde{\mathbb{S}}$ , used to approximate the integrals involved in the QSI-SUR sampling criterion, as explained in [Section 4.2](#).

The approximated sampling criterion is then optimized using an exhaustive search over a subset  $\tilde{\mathbb{D}} \subset \tilde{\mathbb{X}} \times \tilde{\mathbb{S}}$  of candidate points, over which a Gauss-Hermite quadrature is used to approach the integrand (20). To construct this subset of candidate points, we evaluate the probability of misclassification  $\min(p_n(x, s), 1 - p_n(x, s))$  of  $\xi(x, s) \in C$ , and, following the same idea as previously, keep the point with the highest misclassification probability and a random set of  $n_c - 1$  points drawn, according to the probability vector proportional to  $\min(p_n(x, s), 1 - p_n(x, s))$  as  $\tilde{\mathbb{D}}$ .

We follow the rule of thumb which consists of taking an initial design of size  $n_0$  equals to ten times the input dimension of the problem ([Loeppky et al., 2009](#)). The values of the simulation parameters used for the QSI-SUR implementation (and the comparative methods implementations, see supplementary material) are summarized in [Table 1](#).

All the numerical experiments are carried out using Matlab R2022a and the STK toolbox ([Bect et al., 2022](#)).

**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  $\text{card}(\tilde{\mathbb{X}} \times \tilde{\mathbb{S}}) = n_\pi n_S$  is sufficiently low to accommodate a given computation time.*

Test case	$\alpha$	$d_{\mathbb{X}} + d_{\mathbb{S}}$	$n_0$	$n_{\mathbb{X}}$	$n_{\mathbb{S}}$	$n_{\pi}$	$n_c$	$N$	$M$
$f_1$	0.05	1 + 1	20	500	100	15	200	15	250
$f_2$	0.15	2 + 2	40	1000	25	60	200	15	250
$f_3$	0.60	2 + 2	40	1000	25	60	200	15	250
“Volcano”	0.90	5 + 2	70	2500	75	20	200	15	250

Table 1: Summary of the simulation parameters used in the numerical experiments. Note that  $\text{card}(\tilde{\mathbb{X}} \times \tilde{\mathbb{S}}) = n_{\pi}n_{\mathbb{S}} = 1500$  for all the examples (cf. [Remark 4](#)).

## 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  $\gamma(f) = \{(x, s) \in \mathbb{X} \times \mathbb{S} : f(x, s) \in C\}$ . 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 (7) 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<sup>1</sup>-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  $\mathbb{X}$  and the inversion (with respect to the cumulative distribution function of  $\mathbb{P}_{\mathbb{S}}$ ) of a Sobol sequence of  $2^8$  points in  $[0, 1]^{d_{\mathbb{S}}}$ . Considering that the Bayes-optimal estimator (13) 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  $\hat{\Gamma}_n = \{x \in \mathbb{X} : \mathbb{E}_n(\tilde{\tau}(x)) \leq \alpha\}$ , where  $\tilde{\tau}(x)$  is the approximation of  $\tau(x)$  defined by average over the  $2^8$  selected points of  $\mathbb{S}$ .

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 fo-

---

<sup>1</sup>We call “pseudo-maximin” the best LHS, in the sense of the minimal distance between points, in a collection of 1000 independent LHSs.



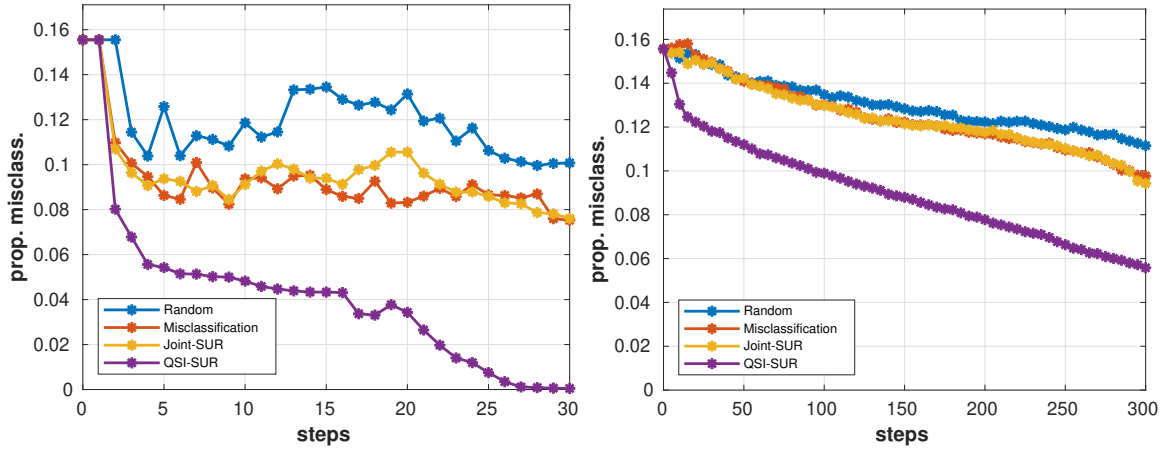


Figure 4: Median of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test functions  $f_1$  (left) and  $f_2$  (right).

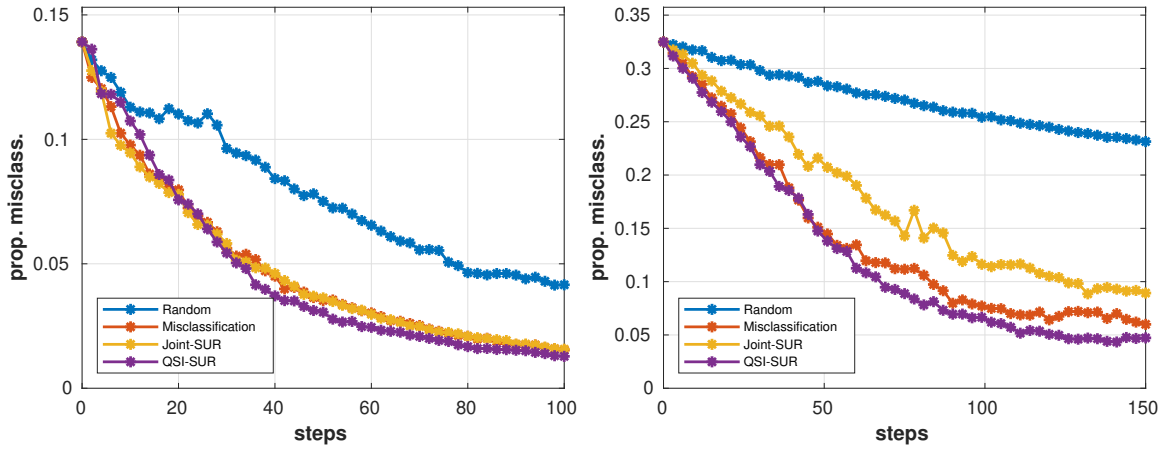


Figure 5: Median of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the function  $f_3$  (left) and the “volcano” test case (right).

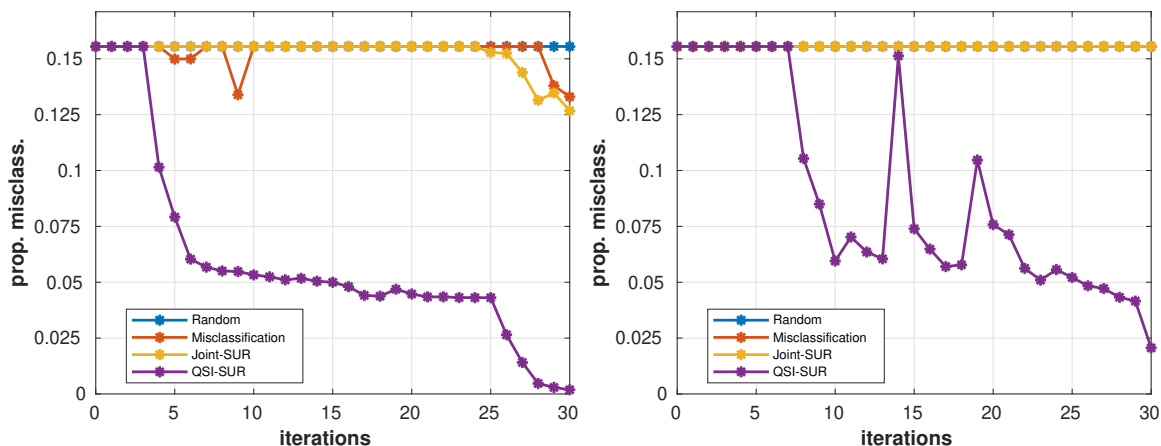


Figure 6: Quantiles of order 0.75 (left) and 0.95 (right) of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_1$ .

cusing on the set  $\gamma(f)$  in the joint space  $\mathbb{X} \times \mathbb{S}$ . 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  $\mathbb{X} \times \mathbb{S}$ , the new criterion tend to concentrate 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”.

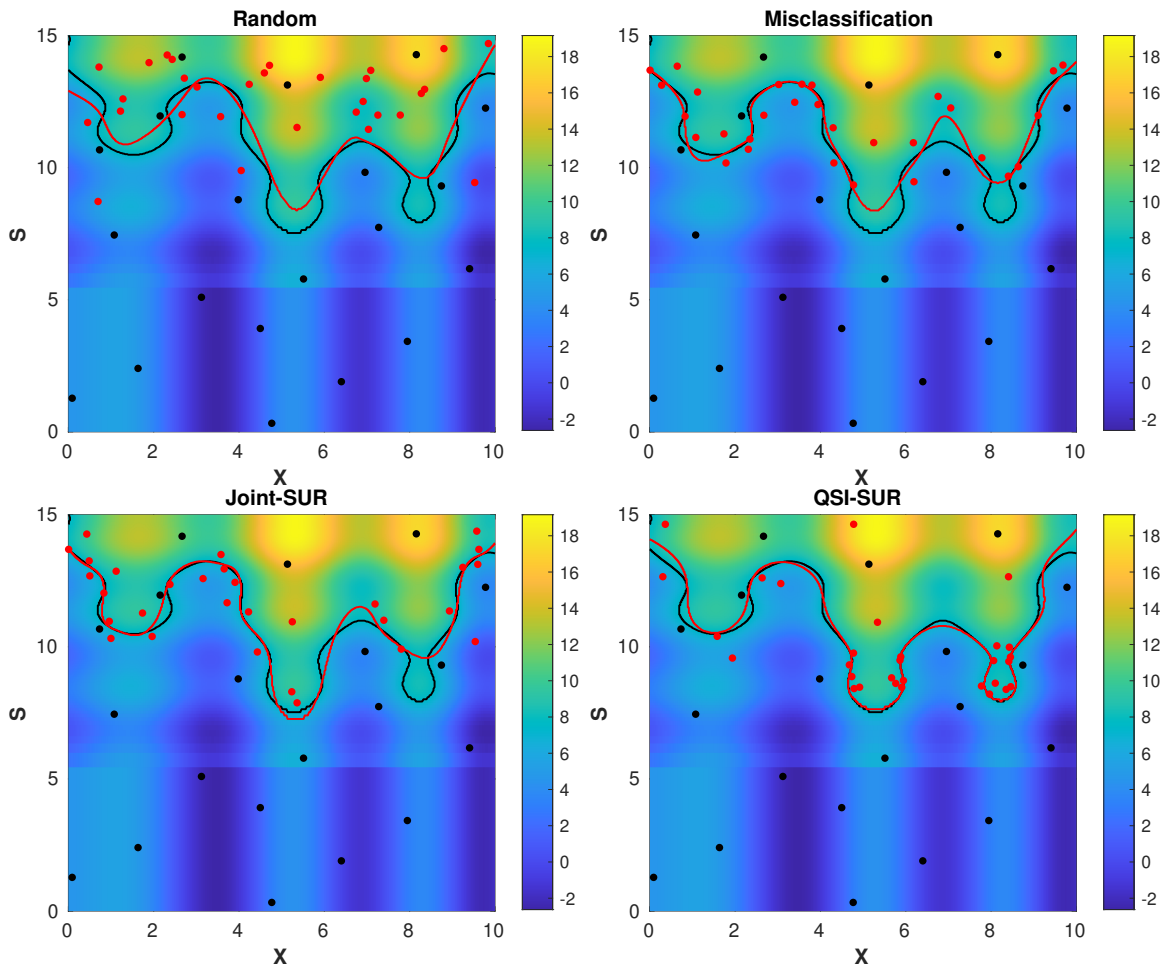


Figure 7: Examples of sequential designs (red dots) obtained after  $n = 30$  iterations on the function  $f_1$ , with the four sampling criteria. Black dots represent the initial design points, black curve the boundary of  $\gamma(f)$ , and red curve the boundary of  $\gamma(\mu_n)$ .

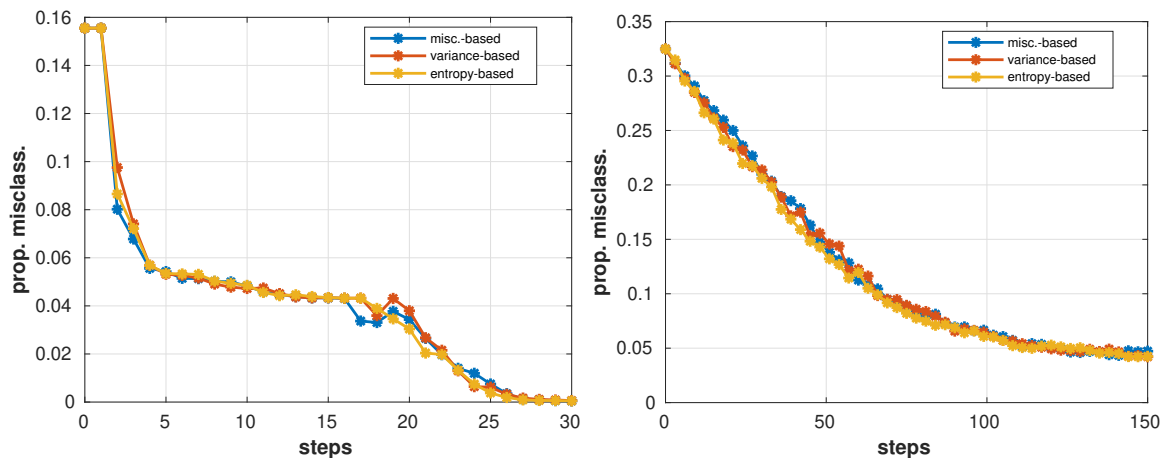


Figure 8: Median of the proportion of misclassified points vs. number of iterations, for 100 repetitions of different possible variants (in terms of  $\mathcal{H}_n$ ) of the QSI-SUR method on the function  $f_1$  (left) and the “volcano” test case (right).

## 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 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

- Azzimonti, D., D. Ginsbourger, C. Chevalier, J. Bect, and Y. Richet (2021). Adaptive design of experiments for conservative estimation of excursion sets. *Technometrics* 63(1), 13–26.
- Bect, J., F. Bachoc, and D. Ginsbourger (2019). A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli* 25(4A), 2883–2919.
- Bect, J., D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing* 22, 773–793.
- Bect, J., E. Vazquez, et al. (2022). STK: a Small (Matlab/Octave) Toolbox for Kriging. Release 2.7.0.
- Bichon, B. J., M. S. Eldred, L. P. Swiler, S. Mahadevan, and J. M. McFarland (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal* 46(10), 2459–2468.
- Branin, F. H. and S. K. Hoo (1972). A method for finding multiple extrema of a function of n variables. In F. A. Lootsma (Ed.), *Numerical methods of Nonlinear Optimization*, pp. 231–237. Academic Press.
- Bryan, B., R. C. Nichol, C. R. Genovese, J. Schneider, C. J. Miller, and L. Wasserman (2005). Active learning for identifying function threshold boundaries. In Y. Weiss, B. Schölkopf, and J. Platt (Eds.), *Advances in Neural Information Processing Systems*, Volume 18. MIT Press.
- Chevalier, C. (2013). *Fast uncertainty reduction strategies relying on Gaussian process models*. Ph. D. thesis, University of Bern.
- Chevalier, C., J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, and Y. Richet (2014). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics* 56(4), 455–465.
- Chevalier, C., D. Ginsbourger, J. Bect, and I. Molchanov (2013). Estimating and quantifying uncertainties on level sets using the Vorob’ev expectation and deviation

- with Gaussian process models. In D. Uciński, A. C. Atkinson, and M. Patan (Eds.), *mODa 10 – Advances in Model-Oriented Design and Analysis*, pp. 35–43. Springer.
- Dixon, L. and G. P. Szegö (1978). The global optimization problem: an introduction. In L. C. W. Dixon and G. P. Szegö (Eds.), *Towards Global Optimization 2*. North Holland.
- Duhamel, C., C. Helbert, M. Munoz Zuniga, C. Prieur, and D. Sinoquet (2023). A SUR version of the Bichon criterion for excursion set estimation. *Statistics and Computing* 33, Article number: 41.
- Durrande, N. and R. Le Riche (2017). Introduction to Gaussian Process surrogate models. HAL cel-01618068. Lecture notes, 4th MDIS-Form@ter workshop, October 16–20, 2017, Besse en Chandesse, France.
- Echard, B., N. Gayton, and M. Lemaire (2011). AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Structural Safety* 33(2), 145–154.
- El Amri, R., C. Helbert, M. Munoz Zuniga, C. Prieur, and D. Sinoquet (2023). Feasible set estimation under functional uncertainty by Gaussian Process modelling. HAL preprint hal-02986558v3.
- Graf, S. and H. Luschgy (2000). *Foundations of Quantization for Probability Distributions*. Lecture Notes in Mathematics. Springer.
- Jaulin, L. and E. Walter (1993). Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica* 29(4), 1053–1064.
- Jones, D. R., M. Schonlau, and W. J. Welch (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* 13, 455–492.
- Loeppky, J. L., J. Sacks, and W. J. Welch (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics* 51(4), 366–376.
- Marques, A., R. Lam, and K. Willcox (2018). Contour location via entropy reduction leveraging multiple information sources. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, pp. 1–11.

- Marrel, A., B. Iooss, and V. Chabridon (2022). The ICSCREAM Methodology: Identification of Penalizing Configurations in Computer Experiments Using Screening and Metamodel—Applications in Thermal Hydraulics. *Nuclear Science and Engineering* 196(3), 301–321.
- Mockus, J. B., V. Tiesis, and A. Žilinskas (1978). The application of Bayesian methods for seeking the extremum. In L. C. W. Dixon and G. P. Szegö (Eds.), *Towards Global Optimization 2*, pp. 117–129. North Holland.
- Mogi, K. (1958). Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them. *Bulletin of the Earthquake Research Institute* 36, 99–134.
- Molchanov, I. S. (1991). Empirical estimation of distribution quantiles of random closed sets. *Theory of Probability & Its Applications* 35(3), 594–600.
- Picheny, V., D. Ginsbourger, O. Roustant, R. T. Haftka, and N.-H. Kim (2010). Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design* 132(7), 071008 (9 pages).
- Picheny, V., T. Wagner, and D. Ginsbourger (2013). A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization* 48(3), 607–626.
- Ranjan, P., D. Bingham, and G. Michailidis (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics* 50(4), 527–541.
- Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*. MIT Press.
- Richet, Y. and V. Bacchi (2019). Inversion algorithm for civil flood defense optimization: Application to two-dimensional numerical model of the Garonne river in france. *Frontiers in Environmental Science* 7(160), 1–16.
- Santner, T. J., B. J. Williams, and W. I. Notz (2003). *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer.

- Sire, C. (2022). Robust inversion under uncertainty for flooding risk analysis. Talk given at the SIAM Conference on Uncertainty Quantification (UQ22), MS10, April 12, Atlanta.
- Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging*. Springer.
- Vazquez, E. and J. Bect (2009). A sequential Bayesian algorithm to estimate a probability of failure. *IFAC Proceedings Volumes 42*(10), 546–550.
- Vazquez, E. and M. Piera-Martinez (2006). Estimation of the volume of an excursion set of a Gaussian process using intrinsic kriging. arXiv preprint math/0611273.
- Villemonteix, J., E. Vazquez, and E. Walter (2009). An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization 44*, 509–534.
- Williamson, D., M. Goldstein, L. Allison, A. Blaker, P. Challenor, L. Jackson, and K. Yamazaki (2013). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics 41*, 1703–1729.



---

## SUPPLEMENTARY MATERIAL

---

### SM1 Proof of the expression of $\mathcal{H}_n$

Let us remark that

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

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

$$\begin{aligned} \mathbb{E}_n \left[ \lambda \left( \Gamma(\xi)\Delta\widehat{\Gamma}_n \right) \right] &= \int_{\mathbb{X}} \mathbb{E}_n \left[ \mathbb{1}_{\Gamma(\xi)\Delta\widehat{\Gamma}_n}(x) \right] dx \\ &= \int_{\mathbb{X}} \mathbb{E}_n \left[ \mathbb{1}_{\{c_n=0\}}(x) \mathbb{1}_{\Gamma(\xi)}(x) \right] \\ &\quad + \int_{\mathbb{X}} \mathbb{E}_n \left[ \mathbb{1}_{\{c_n=1\}}(x) (1 - \mathbb{1}_{\Gamma(\xi)}(x)) \right] dx \\ &= \int_{\mathbb{X}} \mathbb{1}_{\{c_n=0\}}(x) \pi_n(x) dx + \int_{\mathbb{X}} \mathbb{1}_{\{c_n=1\}}(x) (1 - \pi_n(x)) dx. \end{aligned}$$

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

$$\begin{aligned} &\mathbb{1}_{\{c_n=0\}}(x)\pi_n(x) + \mathbb{1}_{\{c_n=1\}}(x)(1 - \pi_n(x)) \\ &= \mathbb{1}_{\{\pi_n \leq \frac{1}{2}\}}(x)\pi_n(x) + \mathbb{1}_{\{\pi_n > \frac{1}{2}\}}(x)(1 - \pi_n(x)) \\ &= \min(\pi_n(x), 1 - \pi_n(x)) \end{aligned}$$

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  $\mathcal{U}(\mathbb{X}) \otimes \mathbb{P}_S$ .
- **Misclassification:** at each step, we sample  $n_\pi \times n_S$  points in  $\mathbb{X} \times \mathbb{S}$  according to  $\mathcal{U}(\mathbb{X}) \otimes \mathbb{P}_S$ . 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  $\mathbb{X} \times \mathbb{S}$  with respect to  $\mathcal{U}(\mathbb{X}) \otimes \mathbb{P}_S$ . 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  $p_{\mathbb{X} \times \mathbb{S}} \propto \min(p_n(x, s), 1 - p_n(x, s))$ , giving us the discretization grid  $\tilde{\mathbb{D}}$ . 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  $\tilde{\mathbb{D}}$ .

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

- **QSI-SUR:** at each step, we first sample  $n_S$  points according to  $\mathbb{P}_S$  as discretization grid  $\tilde{\mathbb{S}}$ . We then uniformly sample a set of  $n_X$  points in  $\mathbb{X}$ . For each of those points, we approximate  $\pi_n(x)$  using Monte Carlo simulation of  $\xi(x, \cdot)$  on  $\tilde{\mathbb{S}}$  (with respect to  $\mathbb{P}_n$ ). 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 give us a product set  $\tilde{\mathbb{X}} \times \tilde{\mathbb{S}}$ , 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  $\tilde{\mathbb{D}} \subset \tilde{\mathbb{X}} \times \tilde{\mathbb{S}}$ . 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  $\tilde{\mathbb{D}}$ .

## SM3 Complementary results for numerical experiments

### SM3.1 Complementary results for the misclassification-based criterion

#### SM3.1.1 Artificial case $f_1$

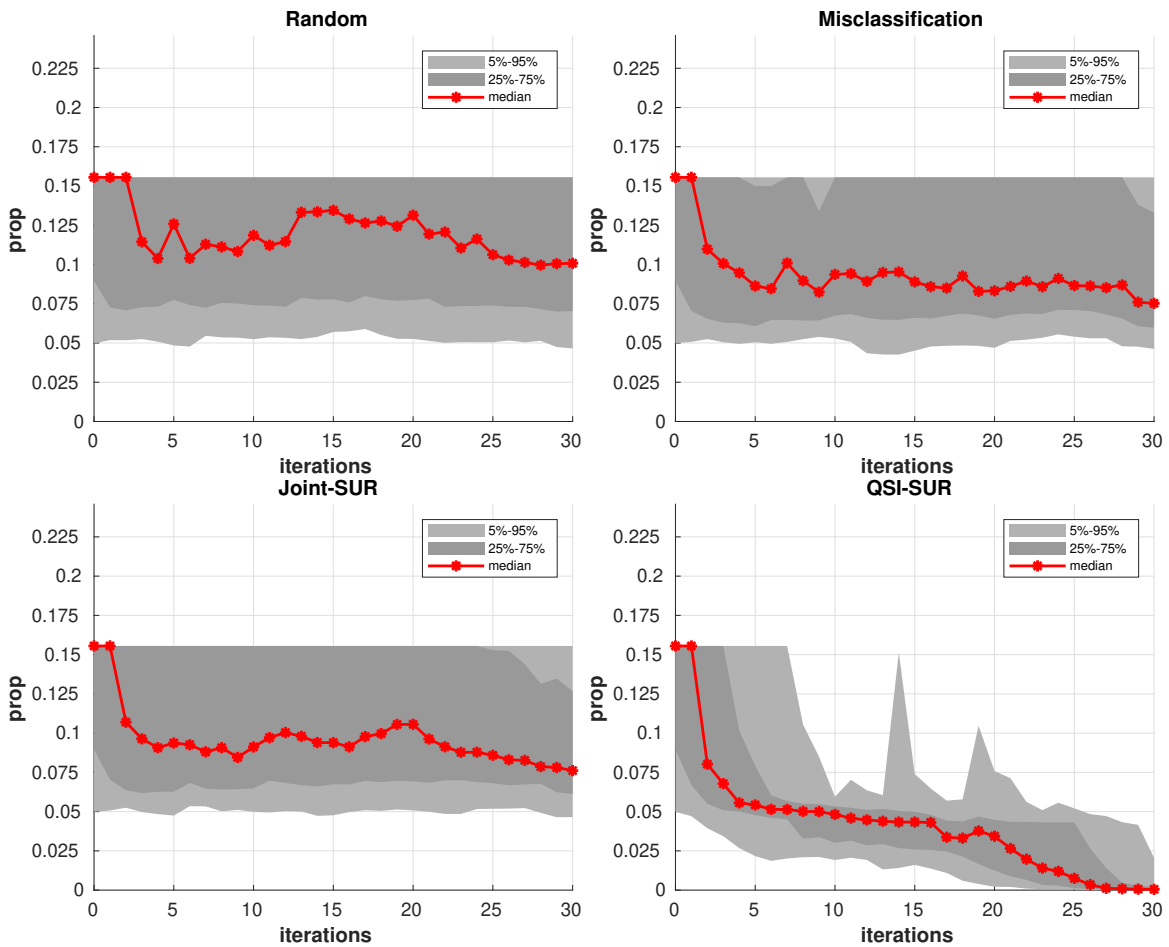


Figure SM1: Median and several quantiles of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_1$ .

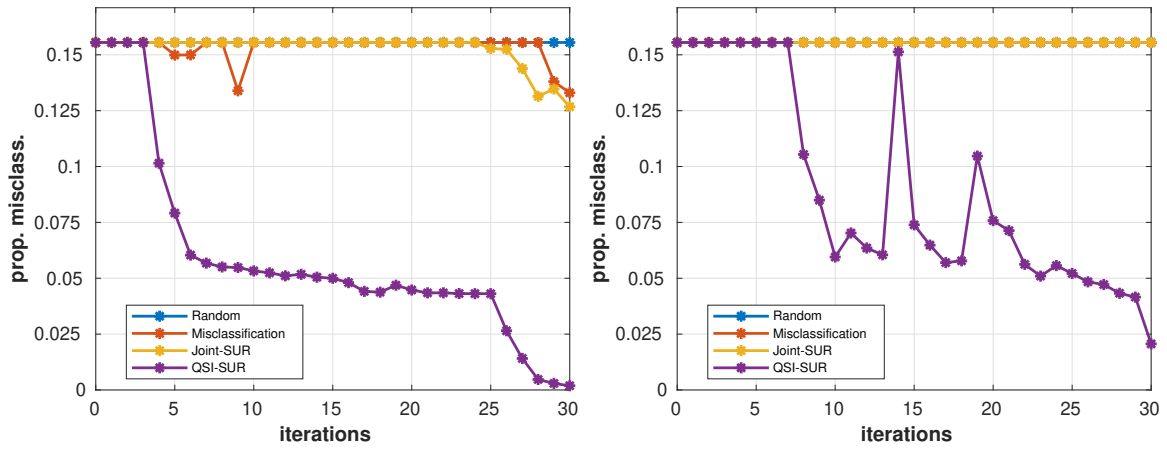


Figure SM2: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_1$ .

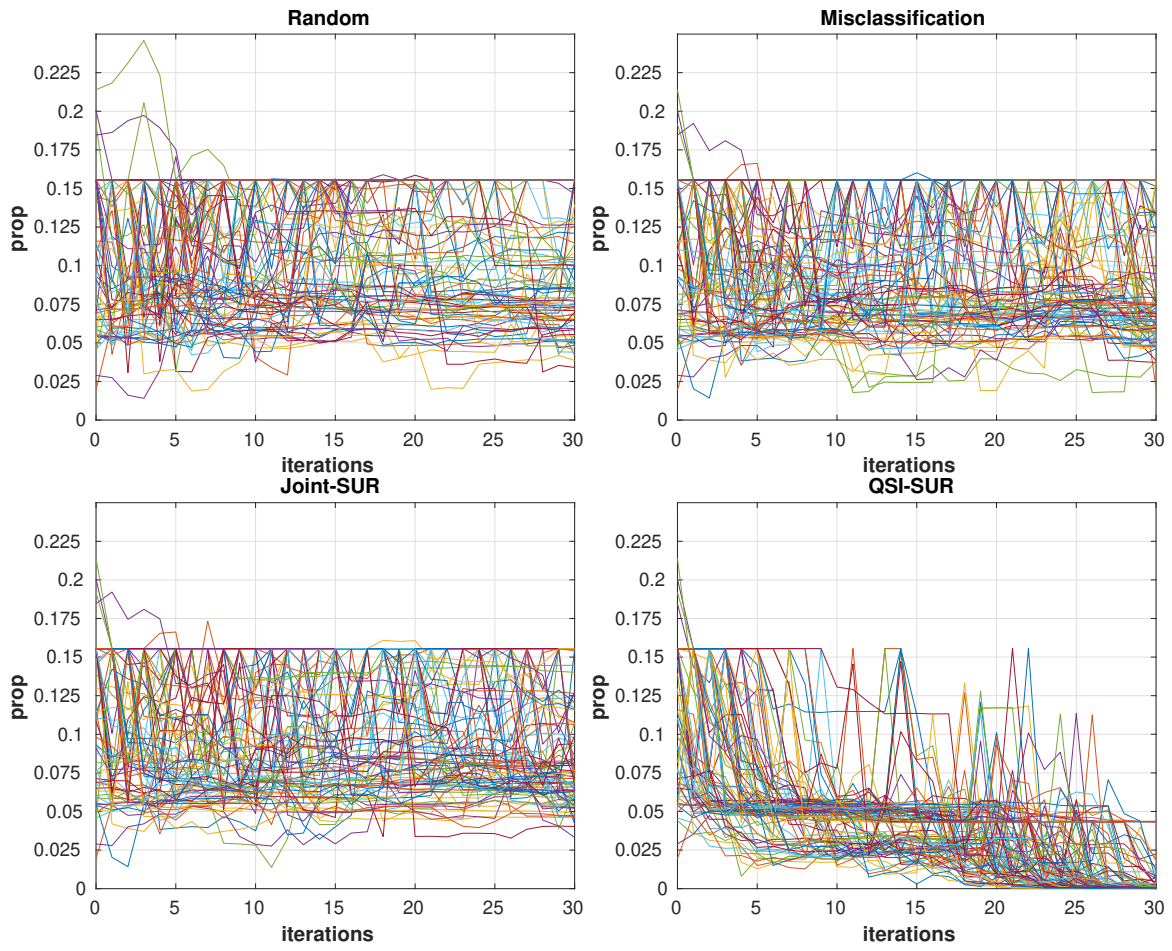


Figure SM3: Different trajectories of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_1$ .

### SM3.1.2 Artificial case $f_2$

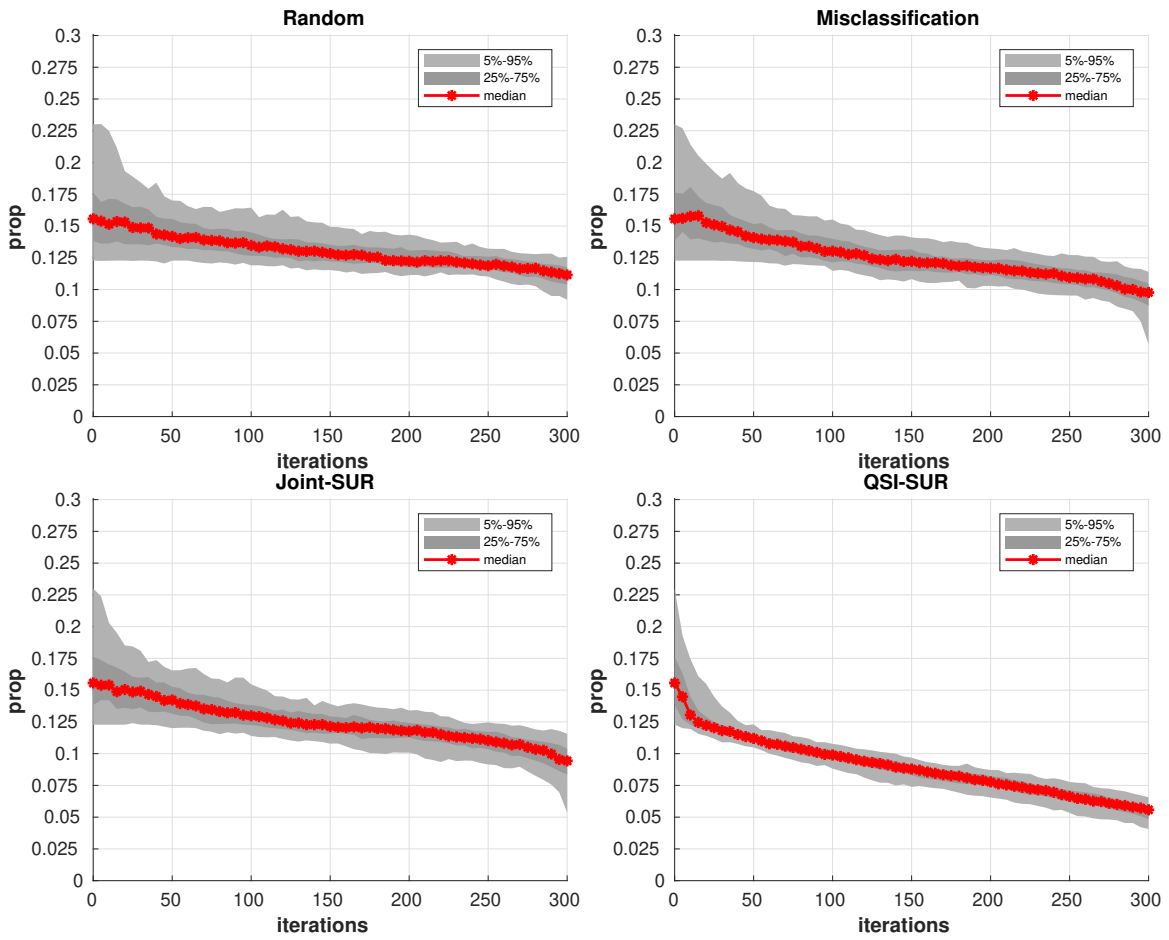


Figure SM4: Median and several quantiles of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_2$ .

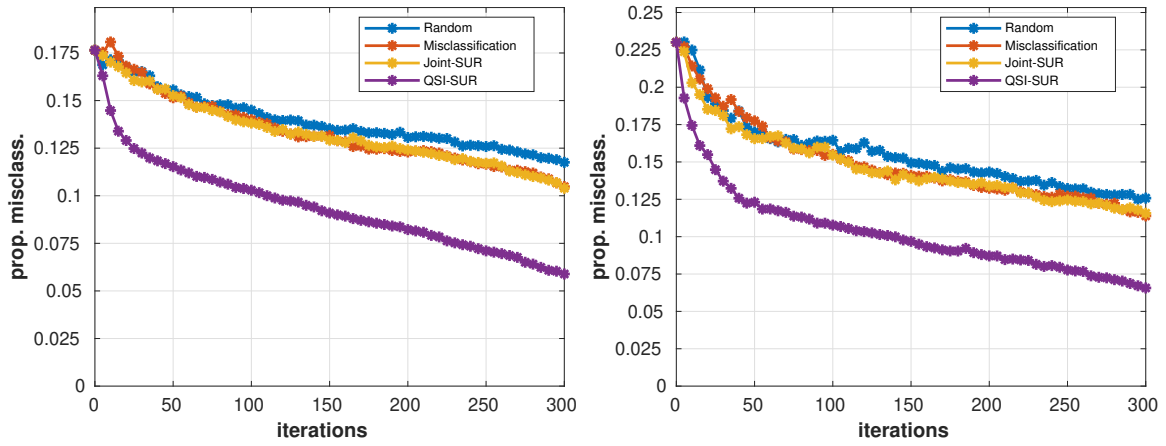


Figure SM5: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_2$ .

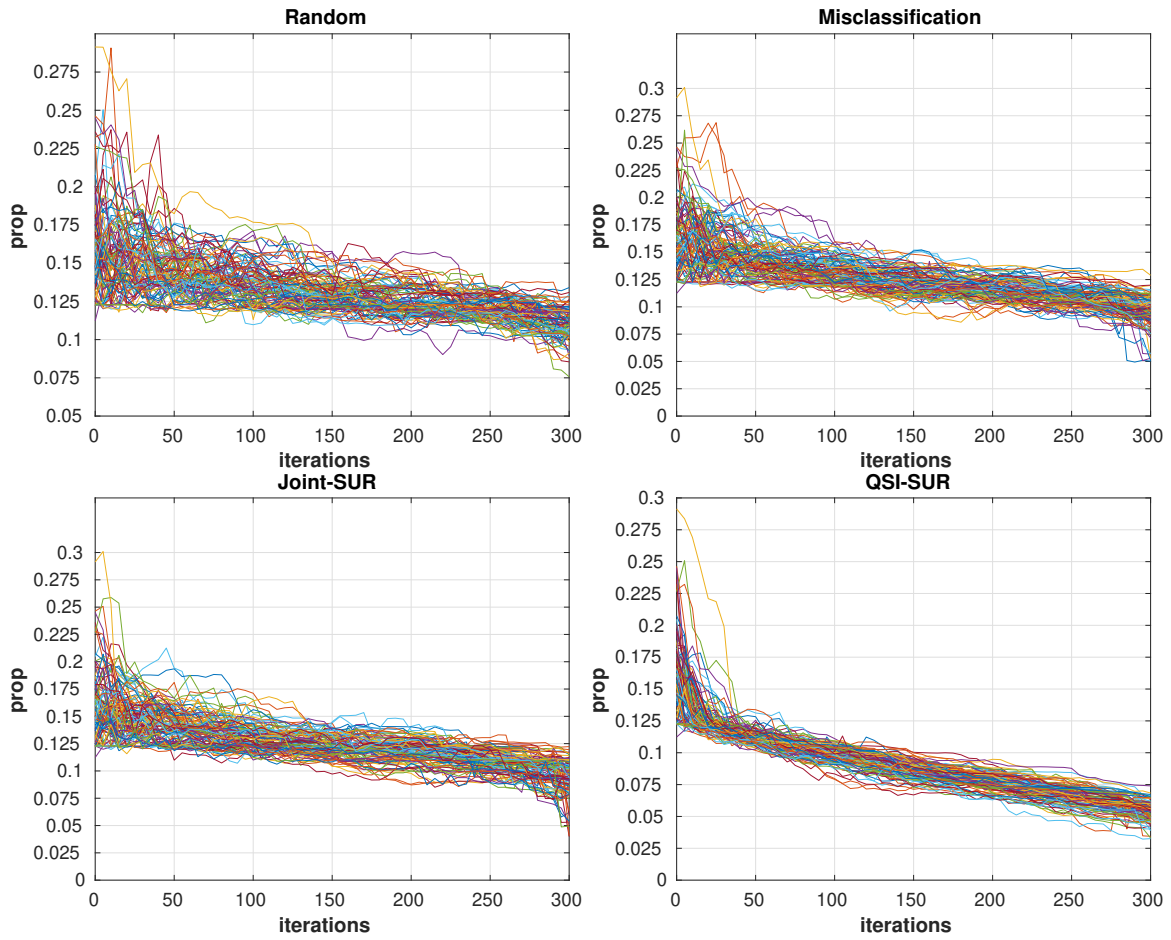


Figure SM6: Different trajectories of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_2$ .



### SM3.1.3 Artificial case $f_3$

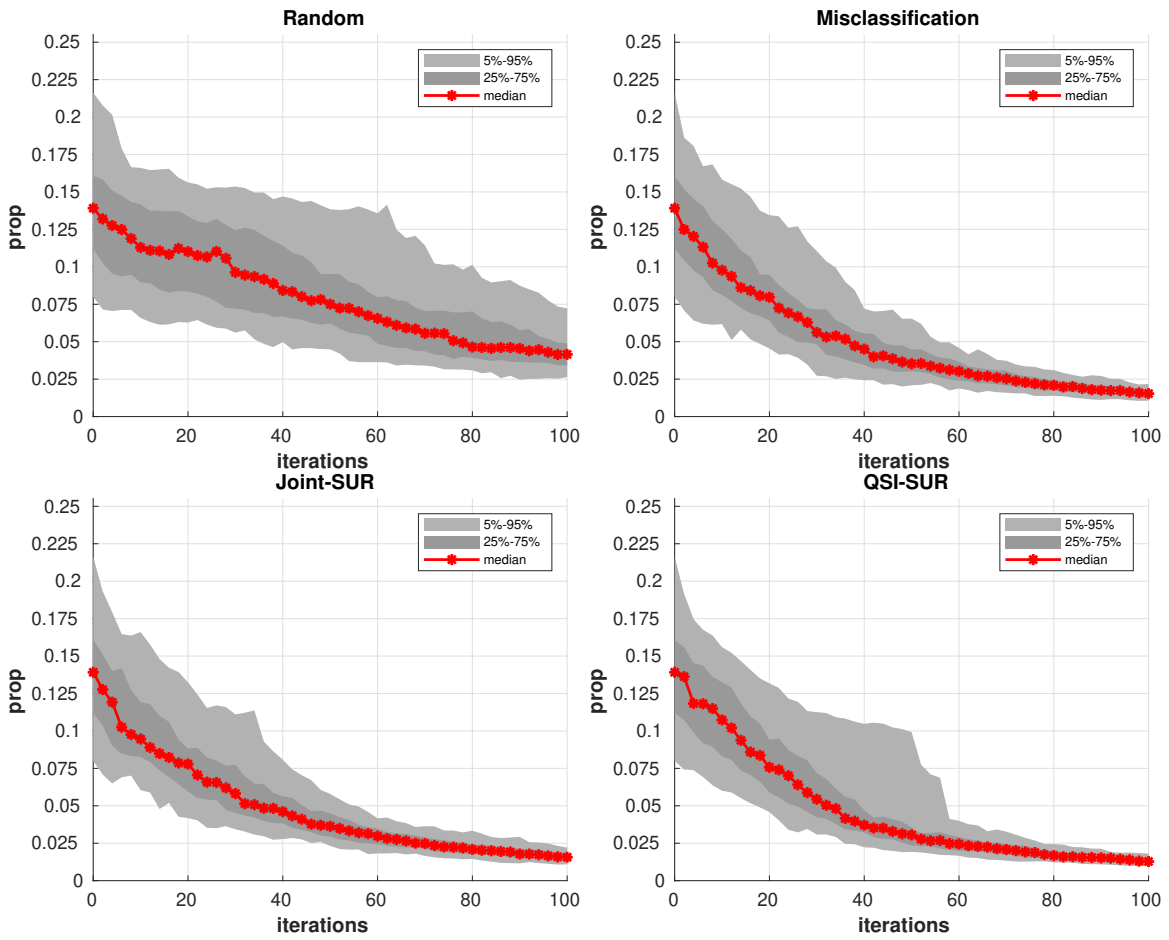


Figure SM7: Median and several quantiles of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_3$ .

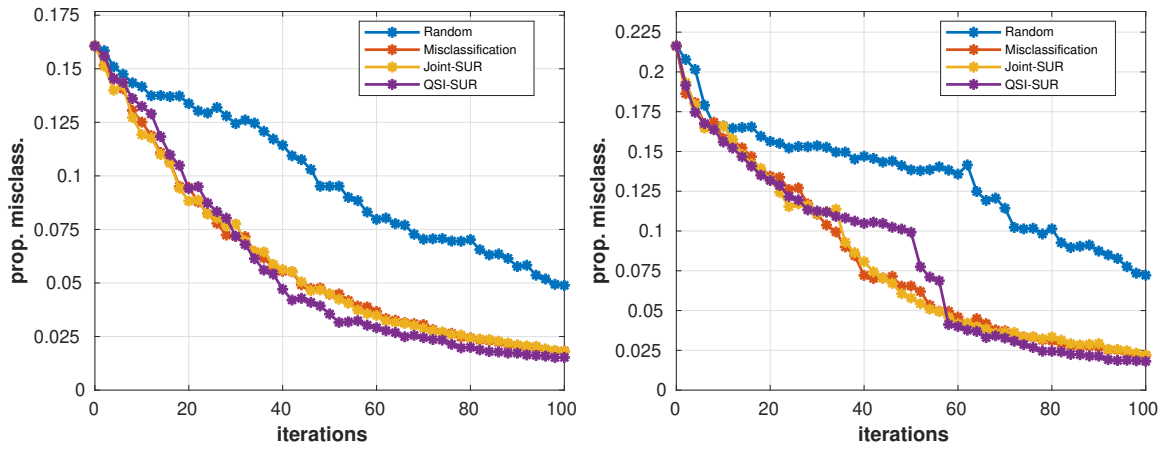


Figure SM8: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_3$ .

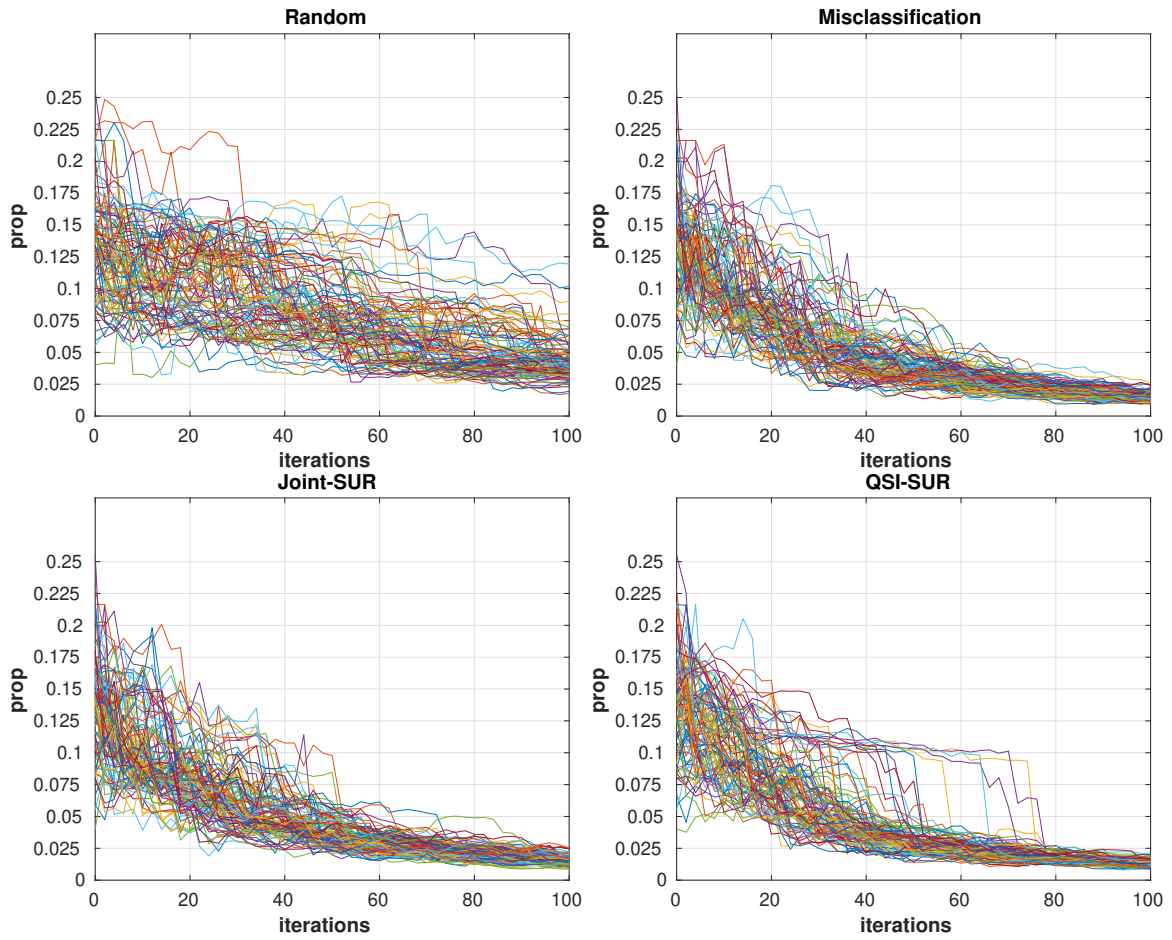


Figure SM9: Different trajectories of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the test function  $f_3$ .

### SM3.1.4 “Volcano” test case

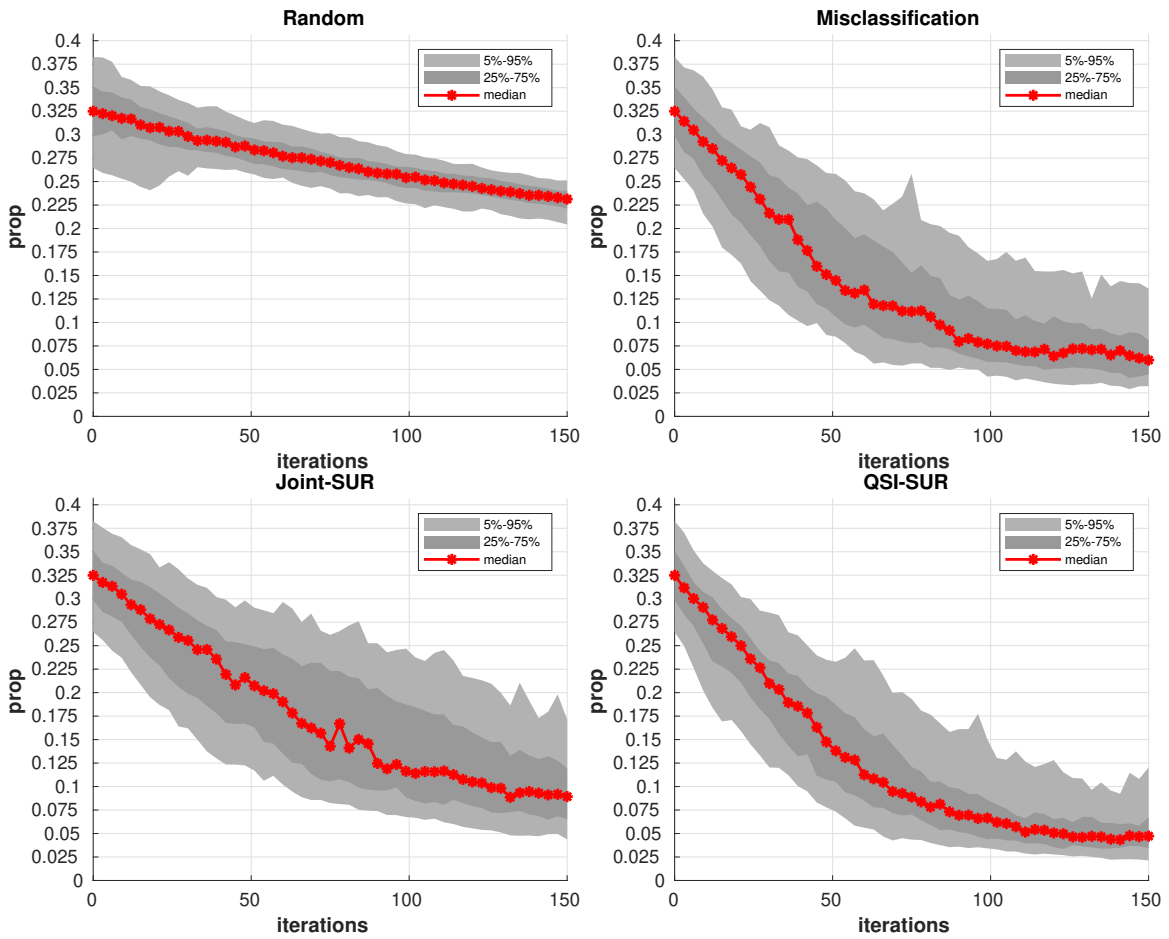


Figure SM10: Median and several quantiles of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the “volcano” test case.

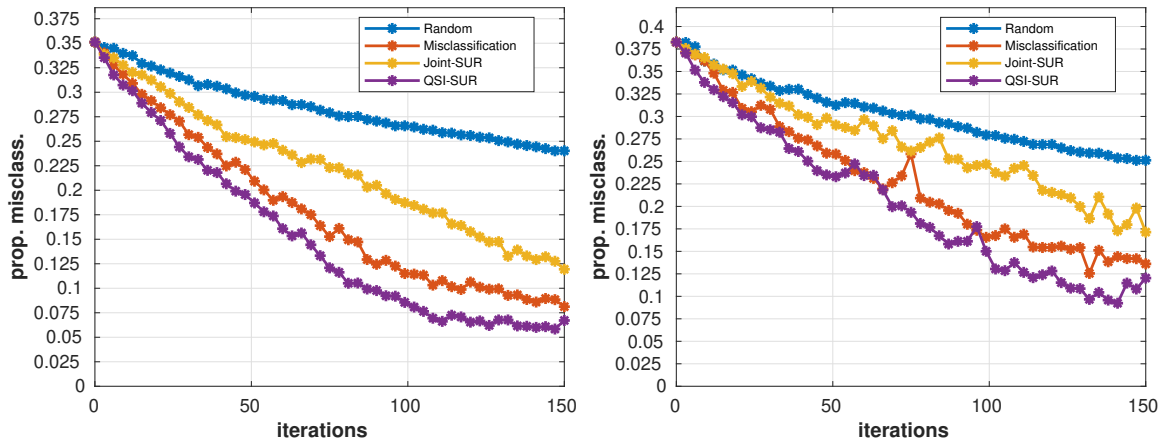


Figure SM11: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the “volcano” test case.

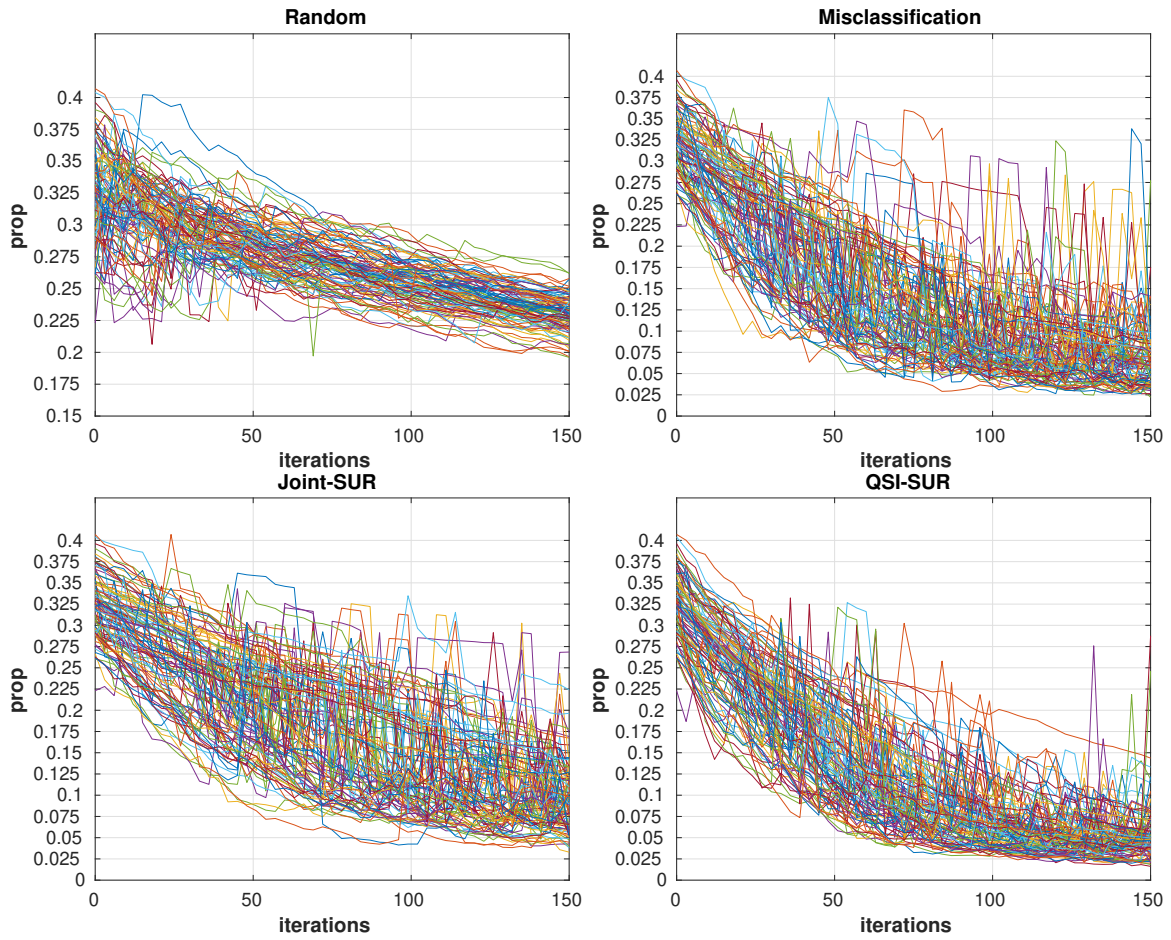


Figure SM12: Different trajectories of the proportion of misclassified points vs. number of iterations, for 100 repetitions of the algorithms on the “volcano” test case.

### SM3.2 Comparison between variants of $\mathcal{H}_n$

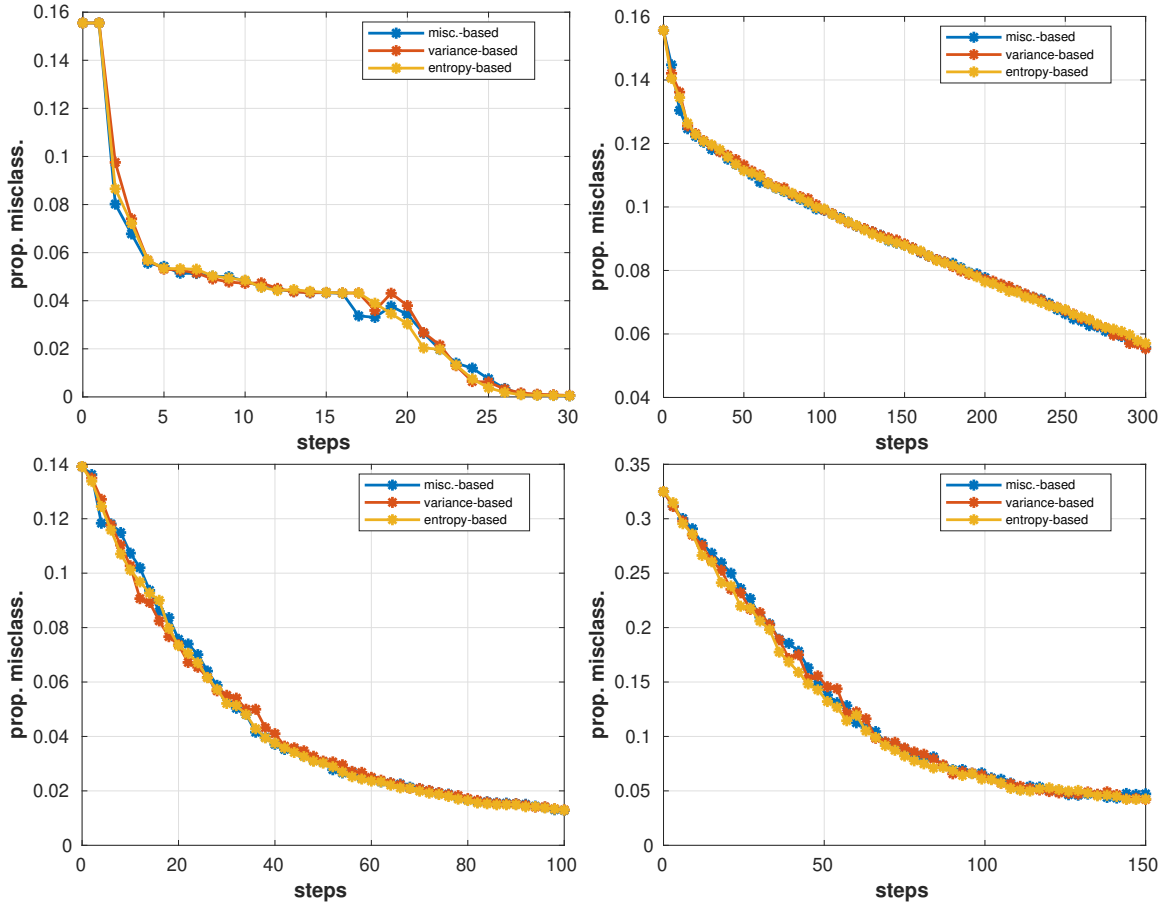


Figure SM13: Median of the proportion of misclassified points vs. number of iterations, for 100 repetitions of different possible variants (in terms of  $\mathcal{H}_n$ ) of the QSI-SUR method on the function  $f_1$ ,  $f_2$ ,  $f_3$  and the “volcano” test case (from left to right, top to bottom).