Bayesian sequential design of computer experiments for quantile set inversion

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

Bayesian sequential design of computer experiments for quantile set inversion

Romain Ait Abdelmalek-Lomenech*[1], Julien Bect[1], Vincent Chabridon[2], and Emmanuel Vazquez[1]

[1]Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 91190 Gif-sur-Yvette, France

[2]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.

*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). The authors report there are no competing interests to declare.

This is an author-generated postprint version.
Accepted for publication in Technometrics, DOI:10.1080/00401706.2024.2394475

Contents

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 various 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 $f : \mathbb{X} \times \mathbb{S} \to \mathbb{R}^q$, where $\mathbb{X}$ and $\mathbb{S}$ are bounded subsets of $\mathbb{R}^{d_\mathbb{X}}$ and $\mathbb{R}^{d_\mathbb{S}}$, 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 $\mathbb{S}$. 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

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

Using the language of machine learning, we can also formulate the QSI problem as that of learning a classifier $\mathbb{X} \to \{0, 1\}$ 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.

An illustrative two-dimensional example of a QSI problem is shown in Figure 1, with one deterministic input variable and one uncertain input variable ($d_\mathbb{X} = d_\mathbb{S} = 1$), critical region $C = (-\infty, 7.5]$ in the output space, and probability threshold $\alpha = 5\%$. The input space $\mathbb{S} = [0, 15]$ for the uncertain variable is equipped with a Beta$(7.5, 1.9)$ distribution, rescaled from $[0, 1]$ to $\mathbb{S}$, which concentrates on large values of S. The set $\Gamma(f)$ to be estimated is the union of two disjoint intervals in $\mathbb{X} = [0, 10]$. It appears clearly, on this example, that an accurate approximation of the boundary of $f^{-1}(C)$, in $\mathbb{X} \times \mathbb{S}$, is only needed in some specific regions of the input space—more specifically, for the points (x, s) of the boundary such that the probability $\mathsf{P}(f(x, S) \in C)$ is close to the threshold α. (See Section 5.3 for

numerical results on this example.)

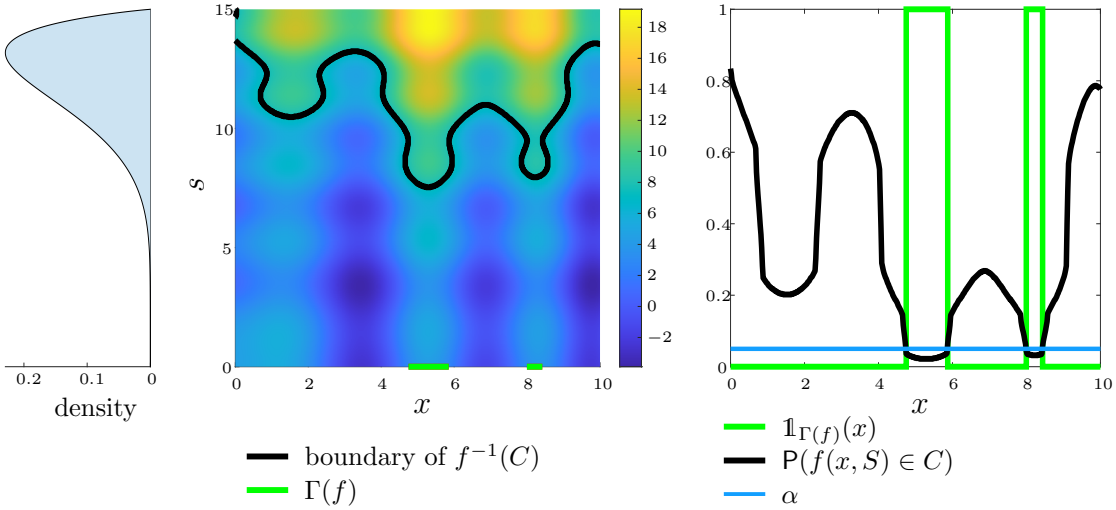

Figure 1: Representation of a two-dimensional QSI problem. Left: probability density function of P_S. Middle: test function $f = f_1$ (see Section 5.3 for details), boundary of $f^{-1}(C)$ and quantile set $\Gamma(f)$ associated to $C = (-\infty, 7.5]$ and $\alpha = 0.05$. Right: indicator function of the quantile set, probability $\mathsf{P}(f(x,S) \in C)$ and probability threshold α.

Remark 1 *When $q = 1$ and C is a semi-infinite interval, there is a direct link between $\Gamma(f)$ and the quantiles of f. For instance, if $C = (-\infty, T]$, the set (1) can be rewritten as*

$$\Gamma(f) = \{x \in \mathbb{X} \,:\, Q_\alpha(f(x,S)) > T\}, \tag{2}$$

where $Q_\alpha(f(x,S))$ denotes the quantile of order α of $f(x,S)$, with $S \sim \mathsf{P}_S$. 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.

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 (see, e.g., 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 minimizing the *expected future uncertainty* on the quantity or object of interest—a set in the present case—given the past observations.

The structure of the article is as follows: Section 2 introduces the framework, while Section 3 gives a brief overview of the literature on Bayesian set inversion strategies, with a particular emphasis on SUR approaches. The core contribution of the 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 various 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 $f : \mathbb{U} \to \mathbb{R}$, 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 $U_1, U_2, \ldots \in \mathbb{U}$ 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 ϵ_n are independent zero-mean Gaussian random variables, with a possibly null variance in the case of a deterministic simulator. As a prior for the unknown function f, we consider a GP model (see, e.g., Rasmussen and Williams, 2006; Santner et al., 2018)—in other words, we assume that f is a sample path of a GP. We denote by ξ this process, and by μ and k its mean and covariance functions.

Denote by $\mathcal{I}_n = \{(U_1, Z_1^{\text{obs}}), \ldots, (U_n, Z_n^{\text{obs}})\}$ the currently available information, and $\mathsf{P}_n = \mathsf{P}(\,\cdot \,|\, \mathcal{I}_n)$ the conditional probability given $\mathcal{I}_n$. Bayesian strategies employ at each step a *sampling criterion*, also referred to as an *acquisition function*, which we will denote by J_n when it is meant to be minimized, or G_n when it is meant to be maximized. This criterion, based on the distribution of ξ under P_n, is used to select the next evaluation point from $\mathbb{U}$. More explicitly: we choose U_{n+1} as an element in $\mathbb{U}$ that minimizes J_n or maximizes G_n:

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

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

Notations. In the rest of the paper, $\mathsf{E}_n = \mathsf{E}(\cdot \mid \mathcal{I}_n)$ 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 $\xi(u)$, and $p_n(u) = \mathsf{P}_n(\xi(u) \in C)$ 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

$$\Lambda(f) = \{x \in \mathbb{X} : f(x) \leq T\}. \tag{3}$$

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 $G_n(x) = \min(p_n(x), 1 - p_n(x))$. This criterion is maximal for any point x such that $\mu_n(x) = T$. Several equivalent criteria lead to the same choice of sampling point, including the entropy of the indicator $\mathbb{1}_C(\xi(x))$ used by Cole et al. (2023), its variance, or the sampling criterion used in the AK-MCS method of Echard et al. (2011).

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

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, 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 $U_{n+1} = u$:

$$J_n(u) \;=\; \mathsf{E}_n\left[\mathcal{H}_{n+1} \mid U_{n+1} = u\right]. \tag{4}$$

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. Equivalently, instead of minimizing the sampling criterion J_n, one can 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, Bect et al. (2012) suggest the integrated probability of misclassification

$$\mathcal{H}_n = \int_{\mathbb{X}} \min(p_n(x), 1 - p_n(x))\,\mathrm{d}x, \tag{5}$$

and $\mathcal{H}_n = \int_{\mathbb{X}} p_n(x)(1-p_n(x))\,\mathrm{d}x$, the integrated variance of $\mathbb{1}_{\Lambda(\xi)}(x)$ as uncertainty measures. Similarly, Marques et al. (2018) propose to use the integrated entropy of the random variable $\mathbb{1}_{(-\infty;T-\epsilon(x)]}(\xi(x)) - \mathbb{1}_{[T+\epsilon(x);+\infty)}(\xi(x))$, with $\epsilon(x) > 0$, to estimate $\Lambda(f)$ in a context where different sources of information can be leveraged.

Picheny et al. (2010) propose a *targeted Integrated Mean Square Error* (tIMSE) reduction strategy, based on the uncertainty measure: $\mathcal{H}_n = \int_{\mathbb{X}} \sigma_n^2(x)\, W_n(x)\,\mathrm{d}x$, where $W_n(x) = \mathsf{E}_n[K(\mu_n(x) - T)]$, with K a kernel (e.g., 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), Azzimonti et al. (2021), and Duhamel et al. (2023).

To conclude this section, let us mention two formulations of the set inversion problem with uncertain input variables, which are related to—but distinct from—the QSI problem. First, Chevalier (2013) considers the task of estimating the set $\{x \in \mathbb{X} : \max_{s\in\mathbb{S}} f(x,s) \le T\}$. In this setting, the proposed uncertainty measure is $\mathcal{H}_n = \int_{\mathbb{X}} p_n^\circ(x)(1 - p_n^\circ(x))\mathrm{d}x$, where $p_n^\circ(x) = \mathsf{P}_n(\max_{s\in\mathbb{S}} \xi(x,s) \le T)$. Second, in the work of El Amri et al. (2023), the objective instead is to estimate the

set $\{x \in \mathbb{X} : \mathsf{E}(f(x,S)) \leq T\}$. To this end, the authors propose a hybrid SUR strategy to choose, sequentially, the deterministic component x and the 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 $Z_n^{obs} = f(X_n, S_n) + \epsilon_n$, i.e., we take $\mathbb{U} = \mathbb{X} \times \mathbb{S}$ and write $U_n = (X_n, S_n)$ from now on.

In the following, we construct a SUR sampling criterion for the QSI problem. For the sake of simplicity, we assume that the distribution P_S admits a density g (with respect to the Lebesgue measure). Consider the random process

$$\tau(x) \;=\; \int_{\mathbb{S}} \mathbb{1}_C\left(\xi(x,s)\right) g(s)\,\mathrm{d}s, \tag{6}$$

which corresponds, for each $x \in \mathbb{X}$, to the stochastic (Bayesian) counterpart of the unknown probability $\mathsf{P}\left(f(x,S) \in C\right)$, and notice that $\Gamma(\xi)$ can be written as

$$\Gamma(\xi) = \{x \in \mathbb{X} : \tau(x) \leq \alpha\}\ . \tag{7}$$

Assume that a sequence $(\widehat{\Gamma}_n)_{n\geq 1}$ of estimators of $\Gamma(\xi)$ has been chosen. We propose to use as uncertainty measure the expected volume of the symmetric difference (see Figure 2) between $\Gamma(\xi)$ and its estimator:

$$\mathcal{H}_n = \mathsf{E}_n\left[\lambda(\Gamma(\xi)\,\Delta\,\widehat{\Gamma}_n)\right], \tag{8}$$

where λ is the usual (Lebesgue) volume measure on $\mathbb{R}^{d_{\mathbb{X}}}$. The SUR strategy derived from (8) consists in minimizing, at each step, the criterion

$$J_n(x,s) = \mathsf{E}_n\left[\mathcal{H}_{n+1} \mid (X_{n+1}, S_{n+1}) = (x,s)\right]\ . \tag{9}$$

SUR strategies using the symmetric difference have been used, in other contexts, by Chevalier (2013) and Azzimonti et al. (2021).

When considering the Bayes-optimal estimator

$$\widehat{\Gamma}_n = \left\{x \in \mathbb{X} : \pi_n(x) > \frac{1}{2}\right\}, \tag{10}$$

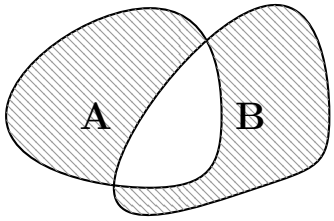

Figure 2: Symmetric difference $A \Delta B = (A \setminus B) \cup (B \setminus A)$ between A and B (shaded area).

where $\pi_n(x) = \mathsf{P}_n(\tau(x) \leq \alpha)$, the uncertainty measure $\mathcal{H}_n$ can be expressed as the integrated probability of misclassification associated to the classifier $\mathbb{1}_{\widehat{\Gamma}_n}$:

$$\mathcal{H}_n = \int_{\mathbb{X}} \min(\pi_n(x),\, 1 - \pi_n(x))\, \mathrm{d}x\,. \tag{11}$$

The proof of this simple result can be found in (Appendix A). As a consequence, the SUR strategy derived from (8) consists in minimizing at each step the criterion

$$\begin{aligned} J_n(\breve{x}, \breve{s}) &= \mathsf{E}_n\left[\mathcal{H}_{n+1} \mid (X_{n+1}, S_{n+1}) = (\breve{x}, \breve{s})\right] \\ &= \int_{\mathbb{X}} \mathsf{E}_n\left[\min(\pi_{n+1}(x),\, 1 - \pi_{n+1}(x)) \mid (X_{n+1}, S_{n+1}) = (\breve{x}, \breve{s})\right] \mathrm{d}x\,. \end{aligned} \tag{12}$$

Here, notice that we use the new notations $\breve{x}$ and $\breve{s}$ for the components of the candidate point, since x is now used as an integration variable.

Remark 2 *Other choices for the uncertainty measure are possible. Notably, we could use any increasing transformation of* $\min(\pi_n(x),\, 1 - \pi_n(x))$ *in the integral* (11) *to define the measure. In particular, we can construct a* variance*-based measure* $\mathcal{H}_n^v = \int_{\mathbb{X}} \pi_n(x)(1 - \pi_n(x))\, \mathrm{d}x$*, and an* entropy*-based one* $\mathcal{H}_n^e = -\int_{\mathbb{X}} \pi_n(x) \log_2(\pi_n(x))\, \mathrm{d}x - \int_{\mathbb{X}} (1 - \pi_n(x)) \log_2(1 - \pi_n(x))\, \mathrm{d}x$*. We focus in the following on the misclassification-based QSI-SUR strategy* (12)*. A comparative benchmark provided as Supplementary Material shows that the other variants yield almost identical results on the four examples of Section 5.*

Remark 3 *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* π_n *by* p_n *and* $\Gamma(\xi)$ *by* $\Lambda(\xi)$*.*

4.2 Approximation of the criterion

It appears from the definition of $\pi_n(x)$ that the proposed criterion does not admit an explicit form, and thus must be approximated. Indeed, it is based on the conditional distribution of $\tau(x)$, which is intractable to the best of our knowledge.

In particular, two major issues arise in the numerical evaluation of $J_n(\breve{x},\, \breve{s})$ at a given point $(\breve{x}, \breve{s}) \in \mathbb{X} \times \mathbb{S}$—namely, the evaluation of the integral over $\mathbb{X}$ and the lack of closed-form formula for the integrand

$$\mathsf{E}_n \left[\min(\pi_{n+1}(x), 1 - \pi_{n+1}(x)) \mid (X_{n+1}, S_{n+1}) = (\breve{x}, \breve{s})\right]. \tag{13}$$

To tackle these issues, we propose an approximation of the criterion based on two ingredients. First, using a suitable auxiliary sampling density, the integral over $\mathbb{X}$ is estimated using importance sampling. Second, at a given point x, the integrand (13) is approached using Monte Carlo simulations of conditional sample paths of the process ξ.

The interested reader can refer to Appendix B for details on this approximation scheme.

5 Numerical experiments

5.1 Implementation of the QSI-SUR strategy

Bayesian model: The underlying function is modeled using a GP prior with a constant mean function and an anisotropic Matérn covariance function. The parameters of the GP prior are estimated, at each step, using the restricted maximum likelihood (ReML) method (see, e.g., Stein, 1999), with the constraint that the regularity parameter υ of the kernel should belong to $\left\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, +\infty\right\}$. Note that the limit case $\upsilon \to +\infty$ corresponds to the Gaussian kernel. For numerical purposes, in order to limit the occurrence of ill-conditioned covariance matrices, a nugget of value 10^{-6} is added. Regarding the initial training points, we use a pseudo[1]-maximin LHS, following the rule of thumb which consists in taking an initial design of size $n_0 = 10d$ (see, e.g., Loeppky et al., 2009), where $d = d_{\mathbb{X}} + d_{\mathbb{S}}$ is the dimension of the input space.

Construction of approximation grid: At each step, we first sample 100 points according to P_S as a discretization grid $\widetilde{\mathbb{S}}_n$ of $\mathbb{S}$. Then, a set of $n_{\mathbb{X}} = 500\, d_{\mathbb{X}}$

[1]We call "pseudo-maximin" the best LHS, in the sense of the minimal distance between points, in a collection of 1000 independent LHSs.

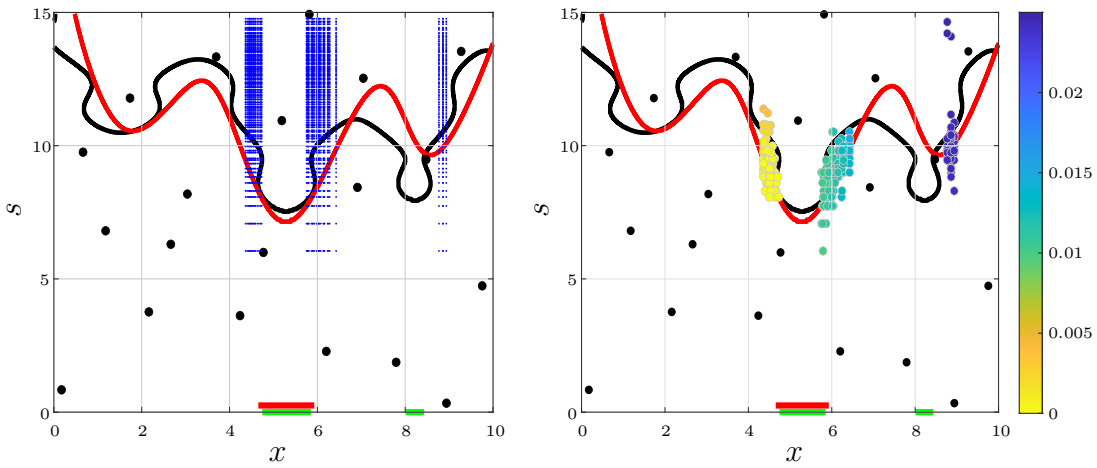

Figure 3: Examples of approximation grid $\widetilde{\mathbb{X}}_n \times \widetilde{\mathbb{S}}_n$ (left) and QSI-SUR criterion values on $\widetilde{\mathbb{U}}_n \subset \widetilde{\mathbb{X}}_n \times \widetilde{\mathbb{S}}_n$ (right), for the acquisition of the first evaluation point on f_1. Black dots represent the initial design, the red and black curves respectively the estimated and true boundary of $\Lambda(f)$. On the x-axis, the segments represent the currently estimated quantile set (red), and the true $\Gamma(f)$ (green).

points is uniformly sampled in $\mathbb{X}$. 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 $\widetilde{\mathbb{S}}$ (with respect to P_n). Finally, a subset $\widetilde{\mathbb{X}}_n$ of $n_{\widetilde{\mathbb{X}}} = 40$ points is constructed, composed of the point with the highest misclassification probability, and $n_{\widetilde{\mathbb{X}}} - 1 = 39$ points drawn (without replacement) according to the discrete probability distribution $p_{\mathbb{X}}(x) \propto \min(\pi_n(x), 1-\pi_n(x))$. Note that this distribution is chosen to ensure that the elements of $\widetilde{\mathbb{X}}_n$ are concentrated in the areas of $\mathbb{X}$ where the probability of misclassification is high. This procedure gives us a product set $\widetilde{\mathbb{X}}_n \times \widetilde{\mathbb{S}}_n$, used to approximate the integrals involved in the QSI-SUR sampling criterion. The integrand (13) arising in the criterion is approximated using a Gauss-Hermite quadrature of 10 points coupled with 100 conditioned sample paths. (See Appendix B for more details on the procedure used to approximate the criterion).

Optimization of the criterion: The approximated sampling criterion is then optimized using an exhaustive search over a subset $\widetilde{\mathbb{U}}_n \subset \widetilde{\mathbb{X}}_n \times \widetilde{\mathbb{S}}_n$ of 250 candidate points. To construct this subset of candidate points, we follow the same idea that for the construction of $\widetilde{\mathbb{X}}$, using this time sampling probabilities proportional to the probability of misclassification $\min(p_n(x,s), 1-p_n(x,s))$ of $\xi(x,s) \in C$.

The main ingredients of this implementation are illustrated in Figure 3.

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* $\widetilde{\mathbb{X}}_n$ *and* $\widetilde{\mathbb{S}}_n$ *must be constructed such that* $\mathrm{card}(\widetilde{\mathbb{X}}_n \times \widetilde{\mathbb{S}}_n)$ *is sufficiently low to accommodate a given computation time.*

5.2 Comparative methods and performance metric

Due to the absence in the literature of strategies tailored specifically to the QSI problem, we propose to compare the performance of the QSI-SUR criterion against methods that aim at approximating the set $\Lambda(f)$ in the joint space $\mathbb{X} \times \mathbb{S}$.

Using the same Bayesian modeling as for the QSI-SUR criterion, the results are first compared to the SUR criterion of Bect et al. (2012), hereafter denoted as "Joint-SUR", using the closed-form expression of Chevalier et al. (2014). We draw $5d_{\mathbb{X}} \times 10^4$ points according to the uniform distribution on $\mathbb{X} \times \mathbb{S}$. The criterion is approximated using a subset of 4000 points constructed from the points with the highest misclassification probability $\min(p_n(x,s), 1 - p_n(x,s))$ and a sample (without replacement) according to this probability. It is then optimized using an exhaustive search on 250 of those points.

We also compare against the criterion of Ranjan et al. (2008) with $\kappa = 1.96$ and the misclassification probability criterion of Bryan et al. (2005), both evaluated on $5d_{\mathbb{X}} \times 10^4$ points sampled uniformly on $\mathbb{X} \times \mathbb{S}$.

The Entropy Contour Locator (ECL) criterion of Cole et al. (2023) is also considered, with 4000 candidate points, using the Python implementation[2] provided by the authors, with minor modifications to fit our Bayesian modeling. As explained in Section 3.1, ECL and the misclassification probability criterion are equivalent in principle: the main difference lies in the use, in ECL, of a local optimizer from the best candidate point (hence the use of a smaller number of candidate points).

Finally, as a baseline, we include the results obtained by uniform random sampling on the space $\mathbb{X} \times \mathbb{S}$.

Except for the ECL criterion, all the experiments are carried out using Matlab R2022a and the STK toolbox v2.8.1 (Bect et al., 2023). The implementation of the methods[3] (except ECL) and the scripts used to run the numerical experiments[4] are available online.

[2] https://bitbucket.org/gramacylab/nasa/src/master/

[3] https://github.com/stk-kriging/contrib-qsi

[4] https://github.com/stk-kriging/qsi-paper-experiments

To assess the performances of the methods, we compare at each step the proportions of misclassified points obtained on a prediction grid composed of the product of a Sobol' sequence of 2^{12} points in $\mathbb{X}$ and the inversion (with respect to the cumulative distribution function of P_S) of a Sobol sequence of 2^{10} points in $[0,1]^{d_\mathbb{S}}$. Considering that the Bayes-optimal estimator (10) is expensive to approximate on such a large grid, we use instead the estimator $\widehat{\Gamma}_n = \{x \in \mathbb{X} : \mathsf{E}_n(\tilde{\tau}(x)) \leq \alpha\}$, where $\tilde{\tau}(x)$ is the approximation of $\tau(x)$ defined by averaging over the 2^{10} selected points of $\mathbb{S}$.

Each method is run 100 times on each test case, using different initial designs, to study performances variability.

5.3 Synthetic examples

We propose first three synthetic examples, with scalar output values ($q = 1$) and noise-free observations.

The first test function (see Figure 1), defined on $\mathbb{X} = [0; 10]$ and $\mathbb{S} = [0; 15]$, is a modified Branin-Hoo function $f_1(x, s) = \frac{1}{12}\, b(x, s) + 3 \sin\left(x^{\frac{5}{4}}\right) + \sin\left(s^{\frac{5}{4}}\right)$, 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 $\mathbb{S}$. The associated set $\Gamma(f_1)$ is represented in Figure 1.

The test function f_2 is defined by $f_2\,(x, s) = \frac{1}{2}\, c(x_1, s_1) + \frac{1}{2}\, c(x_2, s_2)$, on $\mathbb{X} = [-2; 2]^2$ and $S = [-1; 1]^2$, 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 Szegö, 1978). We take $C = (-\infty; T]$ with $T = 1.2$, $\alpha = 0.15$ and P_S the uniform distribution on $\mathbb{S}$.

As a third test function f_3, we consider the Hartman4 function (Picheny et al., 2013), defined on $\mathbb{X} = [0; 1]^2$ and $\mathbb{S} = [0; 1]^2$ equipped with the uniform distribution. For this example we set $\alpha = 0.6$ and $C = [T, +\infty)$, with $T = -1.1$.

From the median performance results in Figure 4, it can be observed that the new sampling criteria tends to perform better on the QSI problem than the state-of-the-art methods focusing on the set $\Lambda(f)$ in the joint space $\mathbb{X} \times \mathbb{S}$. More specifically, it performs much better on the first two cases, and has similar performance on the third one. This remains true when looking at the 75th and 95th percentiles, as illustrated in Figure 5 for f_1 (and in the Supplementary Material for the other 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

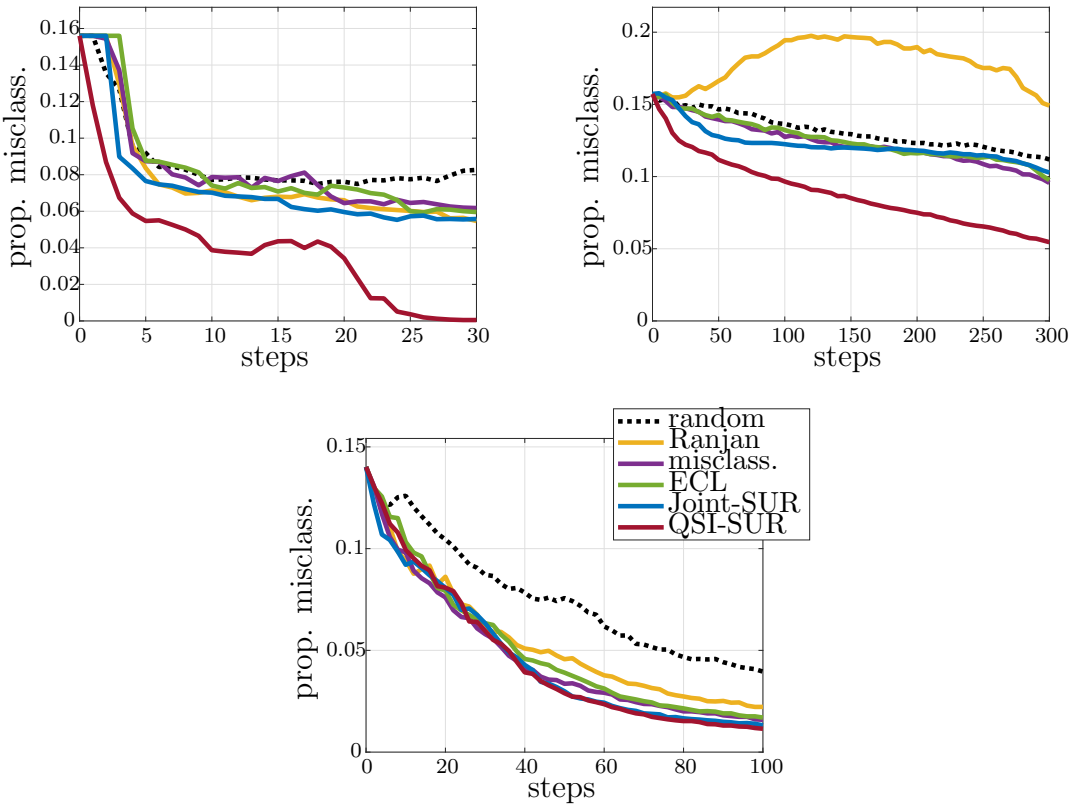

Figure 4: Median of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test functions f_1 (top left), f_2 (top right) and f_3 (bottom).

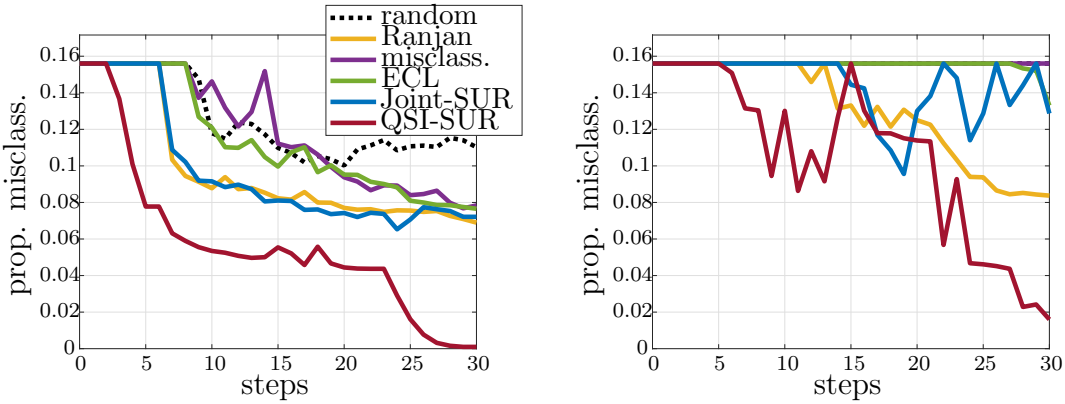

Figure 5: Quantiles of order 0.75 (left) and 0.95 (right) of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_1.

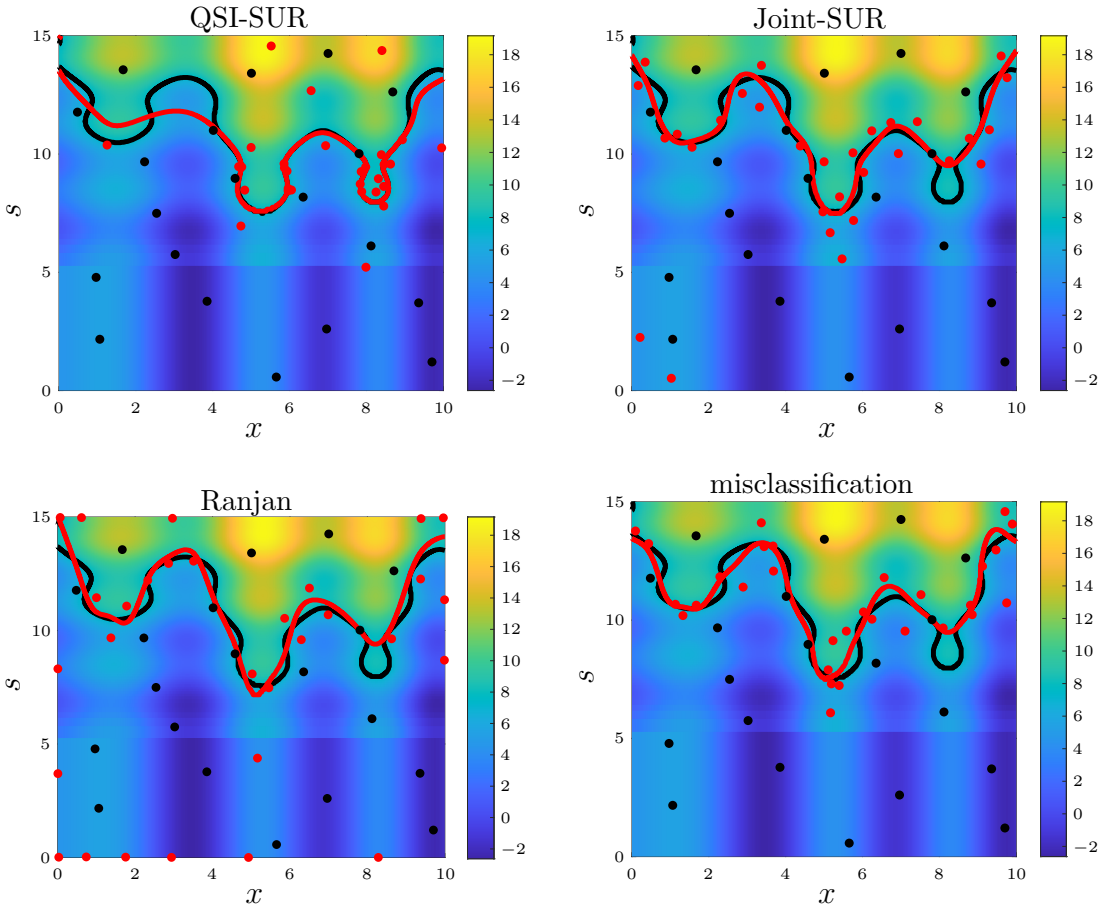

Figure 6: Examples of sequential designs (red dots) obtained after 30 steps on the function f_1, with the QSI-SUR and three others sampling criteria. Black dots represent the initial design points, black curve the boundary of $\Lambda(f)$, and red curve the boundary of $\Lambda(\mu_n)$.

evaluations around specific zones of $\Lambda(f)$ which are particularly relevant for the approximation of $\Gamma(f)$. This phenomenon, related to both the geometry of $\Lambda(f)$ and the distribution P_S, can be visualized in Figure 6 for f_1. In relation to the sequential designs displayed in Figure 6, it can be observed in Figure 7 that the competitor method (here, maximum misclassification probability criterion) tends to select points that are close to the estimated boundary of $\Lambda(f)$, or/and have high variance. In comparison, the QSI-SUR strategy focuses mainly on areas of the joint space $\mathbb{X} \times \mathbb{S}$ that are susceptible—given the current data—to provide information about $\Gamma(f)$. 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.

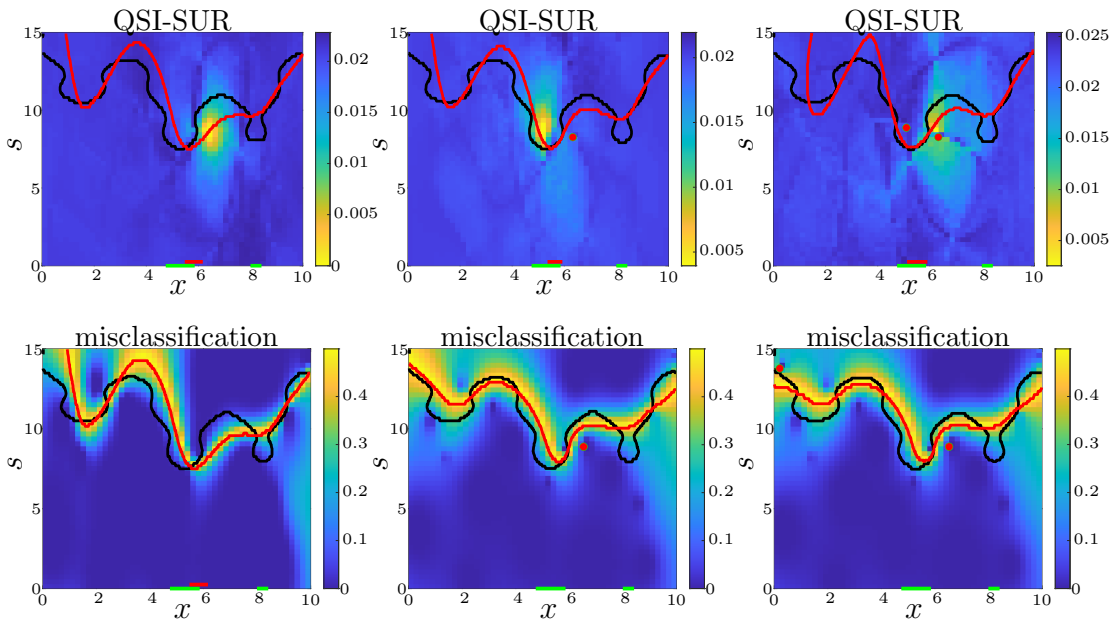

Figure 7: Examples of values of the QSI-SUR (top) and maximum misclassification probability (bottom) criteria, for step 1 to 3 (left to right), for the function f_1. The black curve represents the boundary of $\Lambda(f)$, and the red curve its estimation. Red dots are the points evaluated at the previous steps. The segments on the x-axis represent $\Gamma(f)$ (green) and its estimation (red). Evaluated on a grid of 50×50 points.

This gain in performance has to be put in perspective with the higher computation time of the proposed method, which is, depending on the test case, between 5 and 25 times slower than the simplest methods (e.g., the method of Ranjan et al. (2008), see Appendix C). It is only slightly more expensive, however, than the "Joint-SUR" method (approximately twice slower in our benchmark). This higher computation time remains quite acceptable for expensive computer models, but simpler methods remain relevant for moderately expensive ones (taking, e.g., a few minutes per run).

5.4 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 a 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), which simulates 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 model can be seen as a function $v : \mathbb{X} \times \mathbb{S} \to \mathbb{R}^{220}$, with inputs $(x_1, \dots, x_5) \in \mathbb{X} = [0,1]^5$ 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 G and Poisson ratio ν of the material, which are written as

$$G(s_1) = 2000 + 100(2s_1 - 1) \quad \text{and} \quad \nu(s_2) = 0.25 + 0.3(2s_2 - 1).$$

The uncertain variables are assumed independent and identically distributed, following a Beta distribution with parameters $(2, 2)$.

Given the real measurements $(y_i)_{i\in\{1,\dots,220\}}$ of the displacement at the surface of the volcano (illustrated in Figure 8) and considering the mean absolute error of the simulated displacement against the real measures $V(x,s) = \frac{1}{220}\sum_{i=1}^{220} |v_i(x,s) - y_i|$, our objective is to retrieve the set of plausible parameters for the Mogi model. More specifically, a vector of parameters $x \in \mathbb{X}$ is considered to be "plausible" if it yields an error $V(x,S)$ strictly less than 0.015 with a probability larger than 10%. 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 $\alpha = 90\%$.

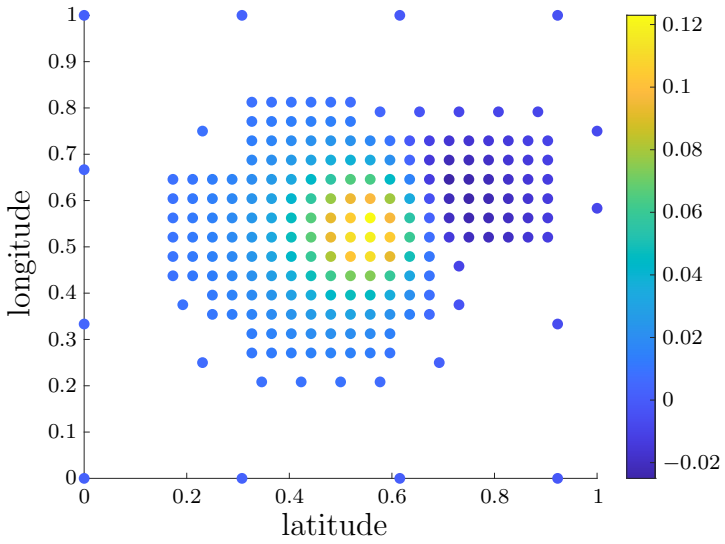

Figure 8: True measures of the displacement at the volcano surface with respect to the normalized latitude and longitude.

We observe in Figure 9 that the median proportion of misclassified points decreases similarly to the top competitors (namely, ECL, probability of misclassification and "Joint-SUR"), with approximately equal median performances after 150 steps. However, our strategy proposes the best "worst case" results, as indicated by the quantile of order 0.95 of the proportion of misclassified points.

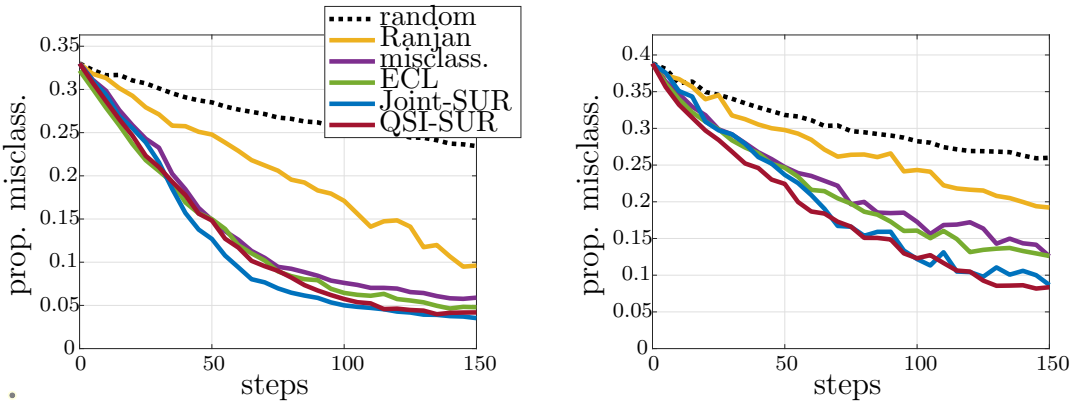

Figure 9: Median (left) and quantile of order 0.95 (right) of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on "volcano" test case.

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 to the heavy use of conditioned Gaussian trajectory simulations. Future work will concentrate on reducing this numerical complexity and making the method applicable to harder test problems, notably in the case of high-dimensional inputs and small quantile sets $\Gamma(f)$. In an other direction, the proposed method could benefit from some adaptations to make it more applicable to real-life problem—in particular, its adaptation to batch design in order to tackle cases where several instances of the simulator can be run in parallel.

Acknowledgments. The authors are grateful to Rodolphe Le Riche and Valérie Cayol 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. (2023). STK: a Small (Matlab/Octave) Toolbox for Kriging. Release 2.8.1.

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.

Cole, D. A., R. B. Gramacy, J. E. Warner, G. F. Bomarito, P. E. Leser, and W. P. Leser (2023). Entropy-based adaptive design for contour finding and estimating reliability. *Journal of Quality Technology 55*(1), 43–60.

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. *Physica D: Nonlinear Phenomena 455*, 133893.

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.

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.

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 (2018). *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.

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

A Proof of the expression of $\mathcal{H}_n$

Let us remark that

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

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

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

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

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

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

B Approximation of the criterion

We give here the details about the approximation of the criterion $J_n(\breve{x}, \breve{s})$. The same procedure can be adapted for the variations of the criterion based on the

variance or the entropy.

The integral on $\mathbb{X}$ (with respect to the uniform distribution) is estimated using an importance sampling scheme. This allows to non-uniformly sample an approximation grid $\widetilde{\mathbb{X}}$ for our integral in order, for instance, to concentrate the sampled points in uncertain areas of $\Gamma(\xi)$. Given a random finite collection $\widetilde{\mathbb{X}}$ of elements of $\mathbb{X}$ sampled from a density $p_{\mathbb{X}}$, we use the following importance sampling approximation:

$$J_n(\breve{x},\, \breve{s}) \;\approx\; \sum_{x\in\widetilde{\mathbb{X}}} \frac{1}{p_{\mathbb{X}}(x)} \mathsf{E}_n \left[\min(\pi_{n+1}(x), 1-\pi_{n+1}(x)) \mid (X_{n+1}, S_{n+1}) = (\breve{x}, \breve{s})\right]. \tag{SM1}$$

We propose to estimate the integrand (13) using quantization of the distribution P_S together with Monte Carlo simulations of the process ξ, in the spirit of Villemonteix et al. (2009).

Consider a finite subset $\widetilde{\mathbb{S}}$ of $\mathbb{S}$, and a family $(w_{\mathbb{S}}(s))_{s\in\widetilde{\mathbb{S}}}$ of positive real numbers such that $\mathsf{P}_{\tilde{S}} = \sum_{s\in\widetilde{\mathbb{S}}} w_{\mathbb{S}}(s)\delta_s$ is a "good" approximation of P_S, where δ_s denotes the Dirac measure at s. This can be achieved, for instance (as done in Section 5), by defining $\widetilde{\mathbb{S}}$ as a collection of $n_{\mathbb{S}}$ i.i.d. samples from P_S and fixing $w_{\mathbb{S}}(s) = \frac{1}{n_{\mathbb{S}}}$ for all $s \in \widetilde{\mathbb{S}}$. For more information about quantization, the reader can refer to Graf and Luschgy (2000).

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

$$\tilde{\pi}^i_{n+1}(x) = \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}_{[0,\alpha]}\left(\sum_{s\in\widetilde{\mathbb{S}}} w_{\mathbb{S}}(s)\, \mathbb{1}_C(\xi_{i,m}(x,s))\right). \tag{SM2}$$

For a sufficiently large M and a "good" quantization $\mathsf{P}_{\tilde{S}}$, we have

$$\tilde{\pi}^i_{n+1}(x) \;\approx\; \mathsf{P}\left(\tau(x) \le \alpha \mid \mathcal{I}_n\,,\, \xi(\breve{x}, \breve{s}) = z_i\right). \tag{SM3}$$

As a consequence, it is possible to use

$$j^x_n(\breve{x}, \breve{s}) = \sum_{i=1}^{N} w_\xi(z_i)\, \min(\tilde{\pi}^i_{n+1}(x), 1-\tilde{\pi}^i_{n+1}(x)) \tag{SM4}$$

as an approximation of (13).

Combining (SM1) and (SM4), the criterion $J_n(\breve{x}, \breve{s})$ is then approximated by

$$\tilde{J}_n(\breve{x}, \breve{s}) \;=\; \sum_{x \in \widetilde{\mathbb{X}}} \frac{1}{p_{\mathbb{X}}(x)}\, j_n^x(\breve{x}, \breve{s}). \tag{SM5}$$

Remark SM1 *For a better numerical efficiency, the simulations of the sample paths of ξ under $\mathsf{P}_n(\cdot \mid \xi(\breve{x}, \breve{s}) = Z_i)$ are preferably carried out using reconditioning of sample paths. A description of this procedure is given by Villemonteix et al. (2009), Section 5.1.*

C Details on computational cost

Due to the major implementation differences between the Entropy Contour Locator (ECL) method of Cole et al. (2023) and the others competitors, we exclude it of this benchmark. We focus here on the strategies implemented in Matlab using the STK toolbox v2.8.1 (Bect et al., 2023). These experiments are conducted using Matlab R2022a and the same parameters as described in Section 5, on a computer equipped with a CPU AMD Ryzen 7 3700x with 32GB of RAM.

	Ranjan	misclass.	Joint-SUR	QSI-SUR
f_1	0.15	0.14	3.74	3.77
f_2	0.29	0.23	7.01	5.91
f_3	0.30	0.23	6.57	5.35
Volcano	1.24	0.76	12.77	11.30

Table SM1: Runtime (in seconds) to complete the first step. Average over 10 runs.

	Ranjan	misclass.	Joint-SUR	QSI-SUR
f_1	1	0.81	19.26	21.94
f_2	1	0.84	4.96	9.45
f_3	1	0.83	10.02	11.05
Volcano	1	0.53	4.26	7.19

Table SM2: Normalize total runtime (in seconds). Average over 10 runs.

D Comparison between variants of the QSI-SUR criterion

Following Remark 2, we display here a brief comparison of several variants of the QSI-SUR criterion—namely, the misclassification probability-based, variance-based, and entropy-based sampling criteria.

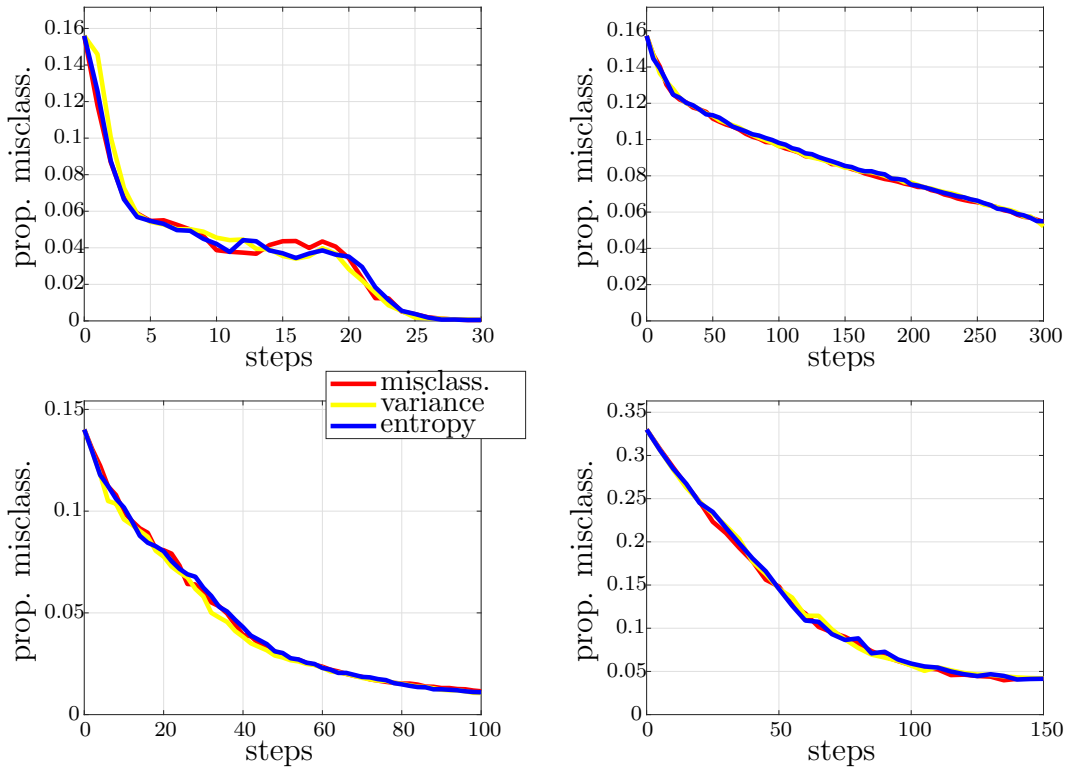

Figure SM1: Median of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test functions f_1 (top left), f_2 (top right), f_3 (bottom left) and the volcano case (bottom right).

E Complementary results for the examples in the article

In this section, some complementary details on the numerical experiments of Section 5 are given. This include, for all the competitors, the 100 sample paths and the quantiles of order 75% and 95% of the error (proportion of misclassified points) as a function of the number of steps.

E.1 Synthetic example f_1

See Figures SM2–SM4.

E.2 Synthetic example f_2

See Figures SM5–SM7.

E.3 Synthetic example f_3

See Figures SM8–SM10.

E.4 Volcano test case

See Figures SM11–SM13.

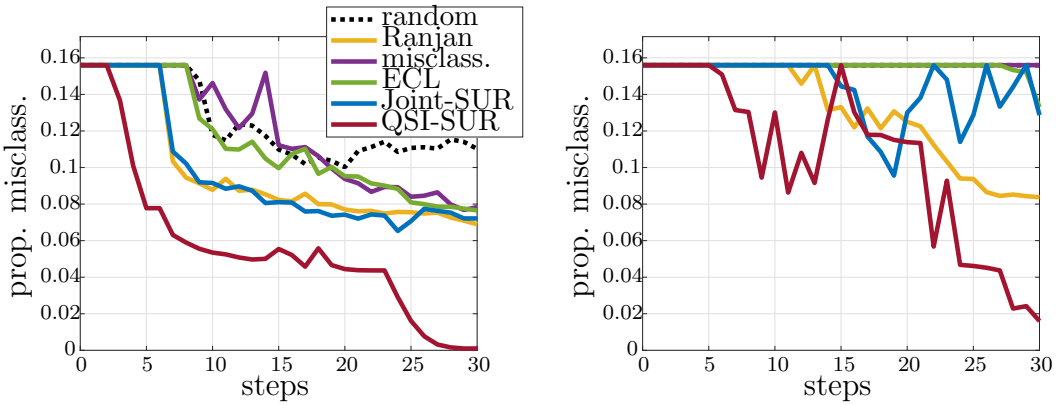

Figure SM2: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_1.

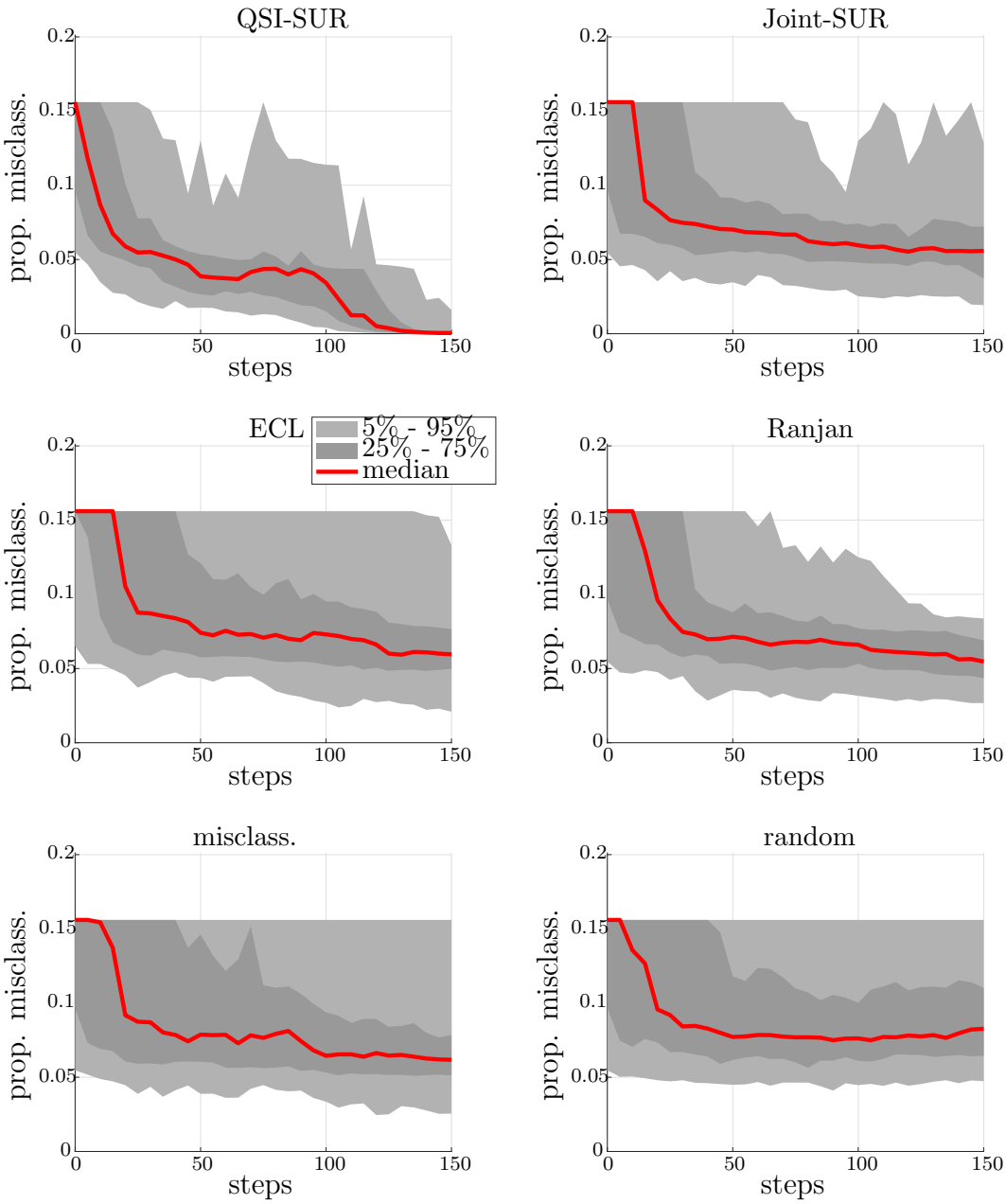

Figure SM3: Median and several quantiles of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_1.

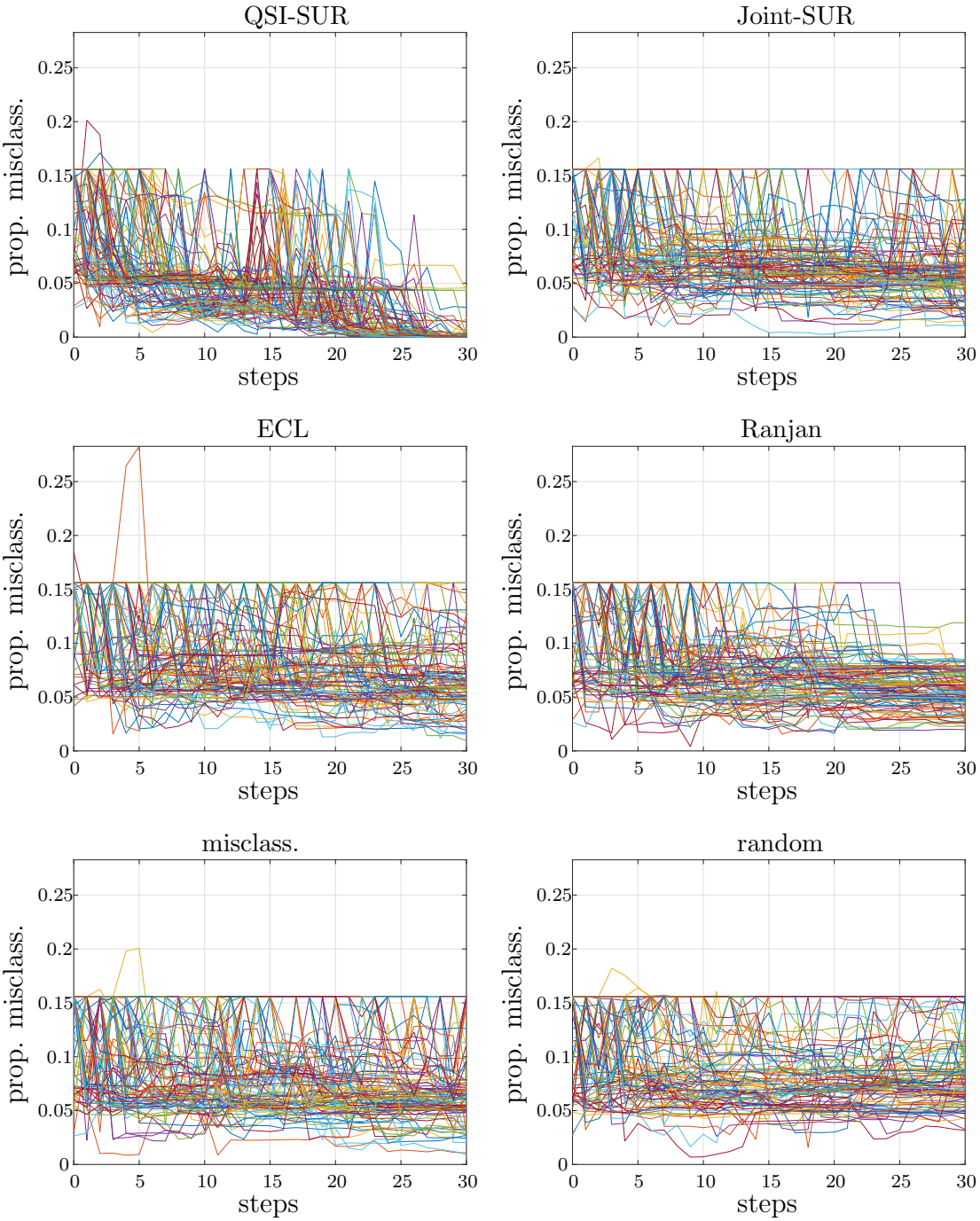

Figure SM4: Different sample paths of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_1.

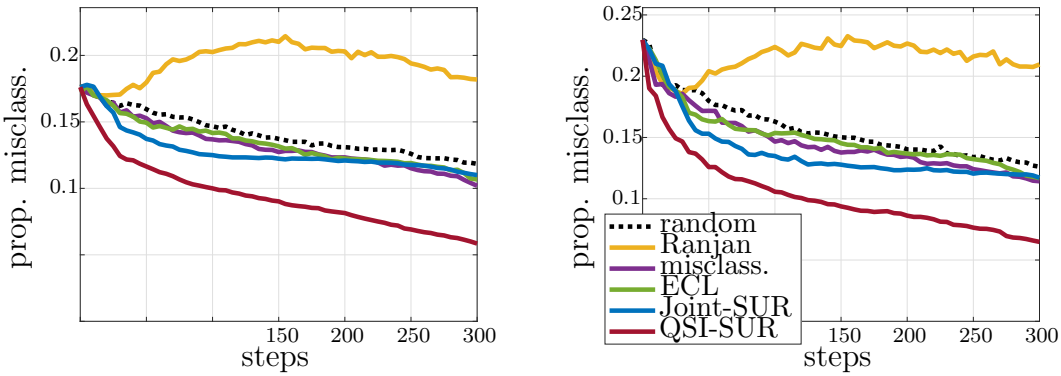

Figure SM5: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_2.

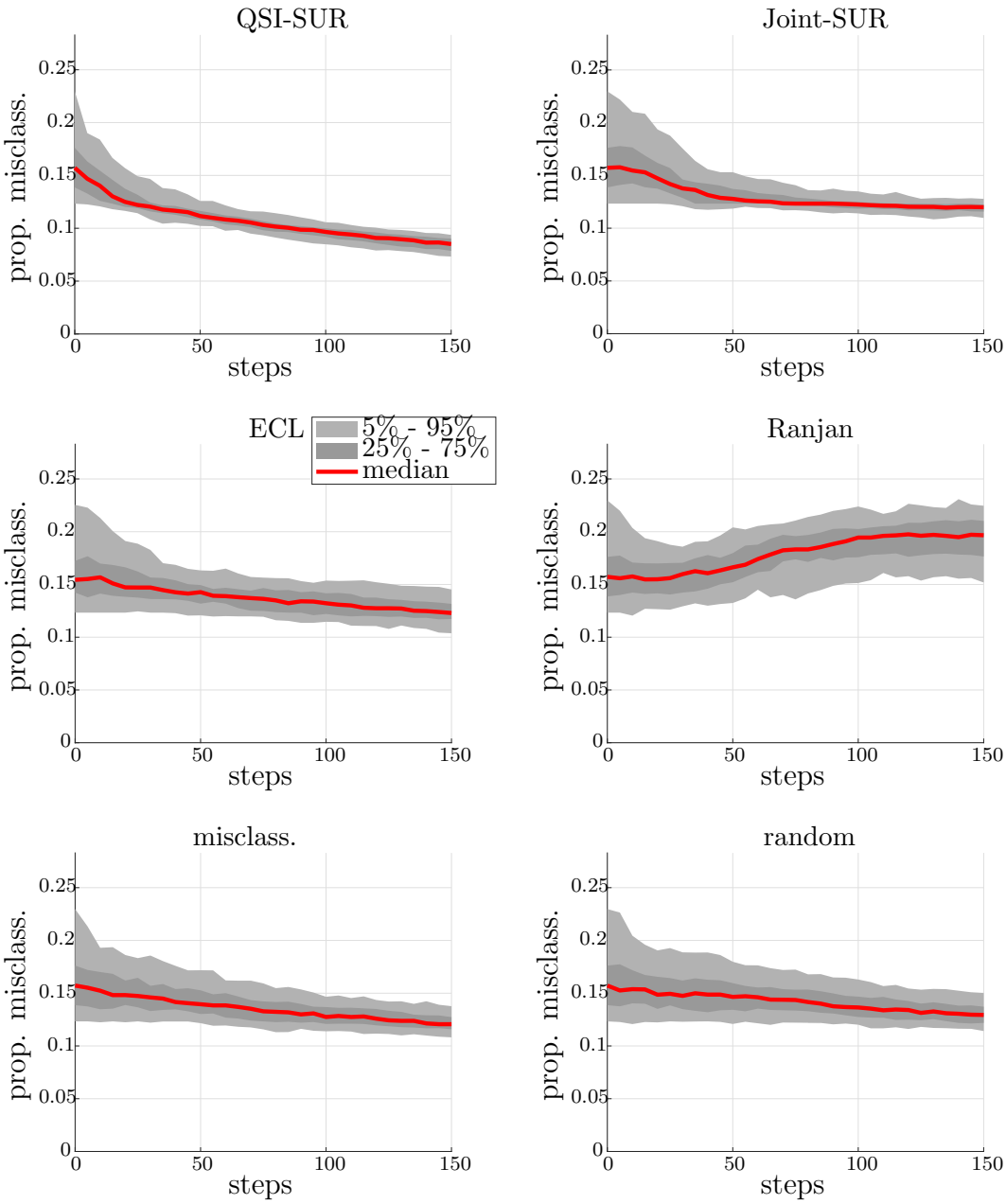

Figure SM6: Median and several quantiles of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_2.

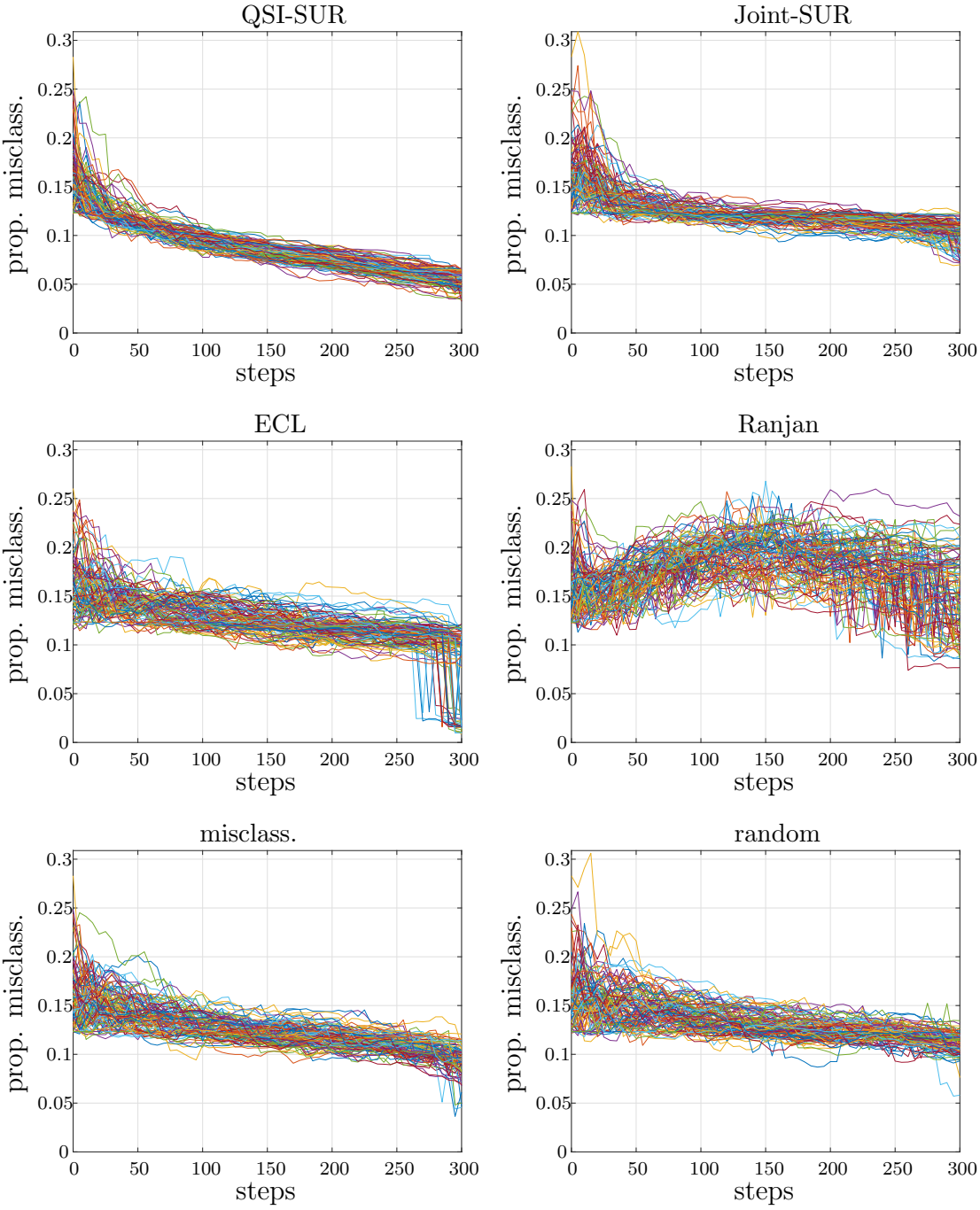

Figure SM7: Different sample paths of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_2.

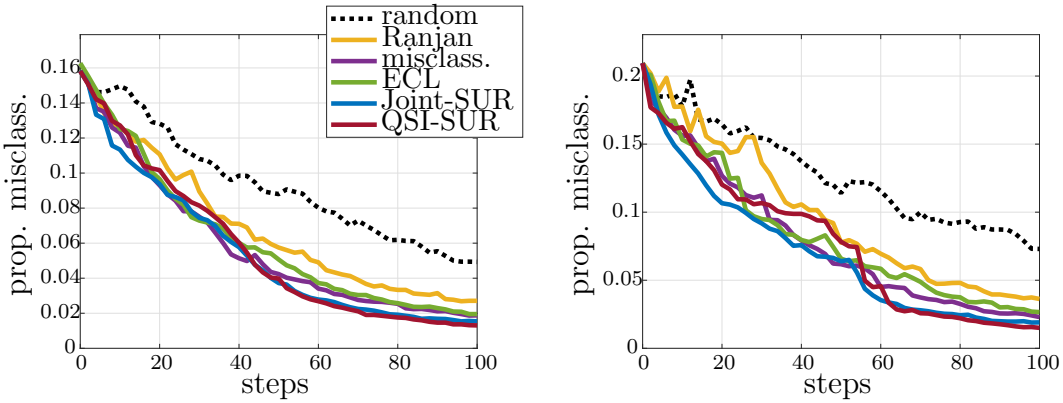

Figure SM8: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_3.

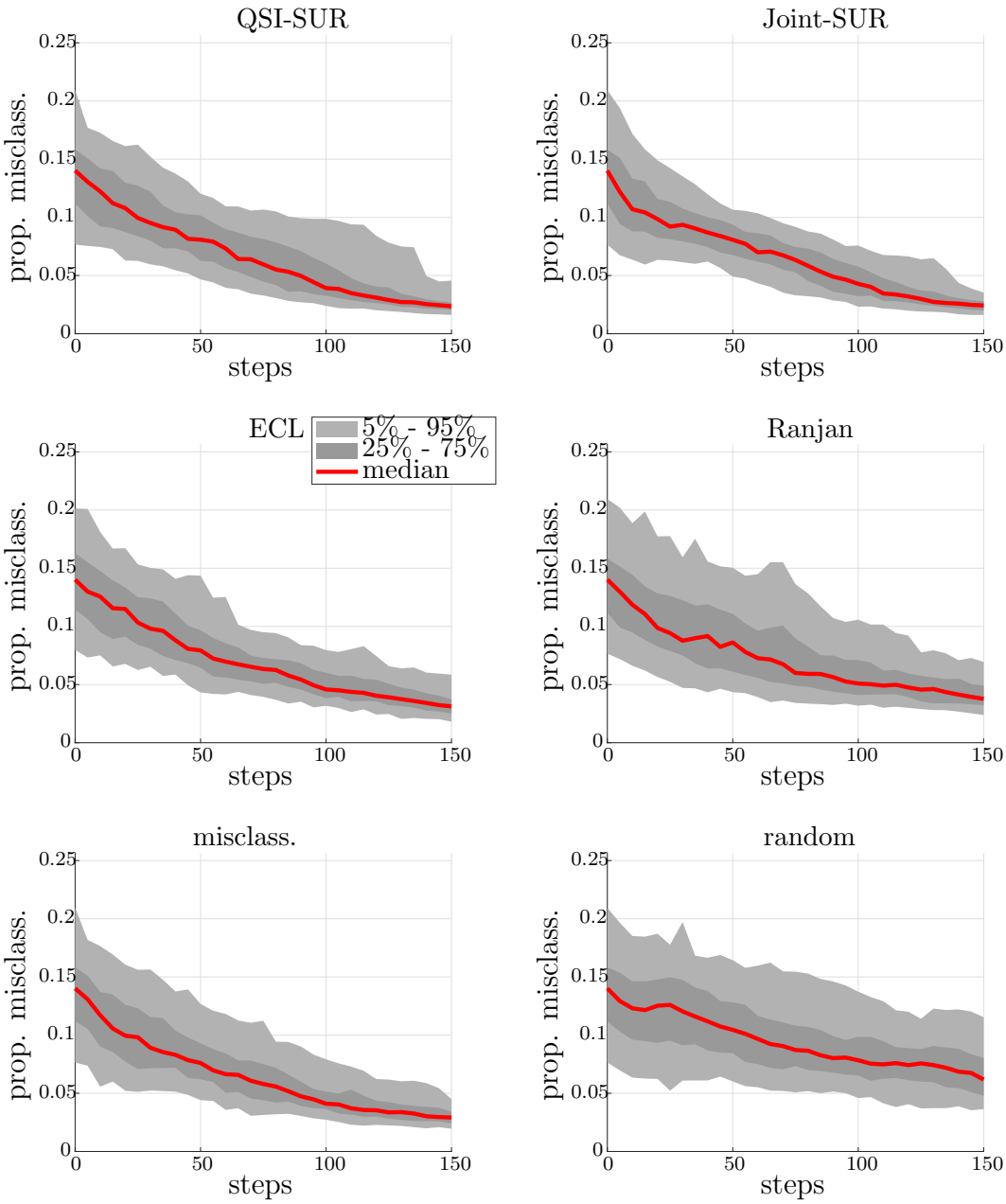

Figure SM9: Median and several quantiles of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_3.

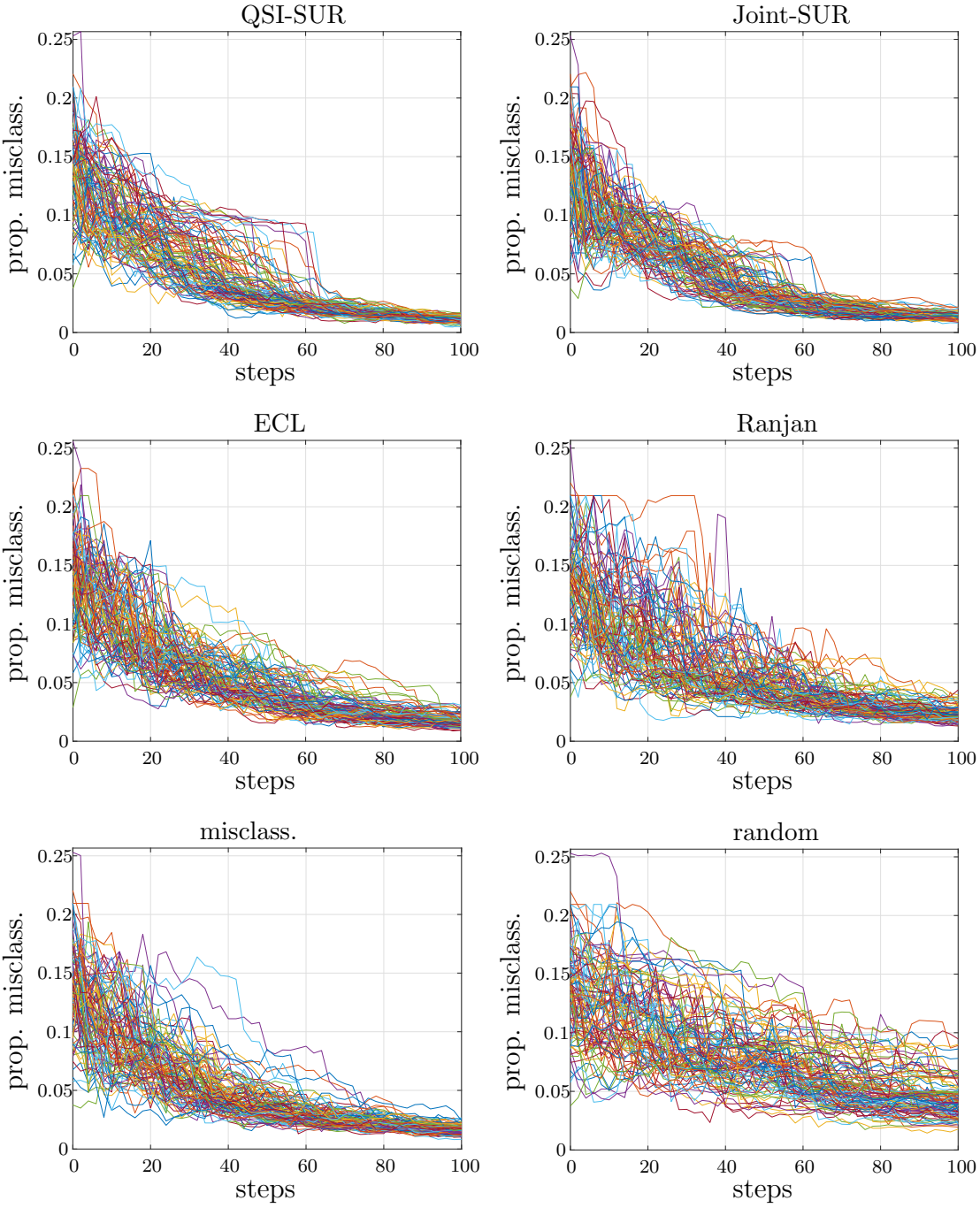

Figure SM10: Different sample paths of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test function f_3.

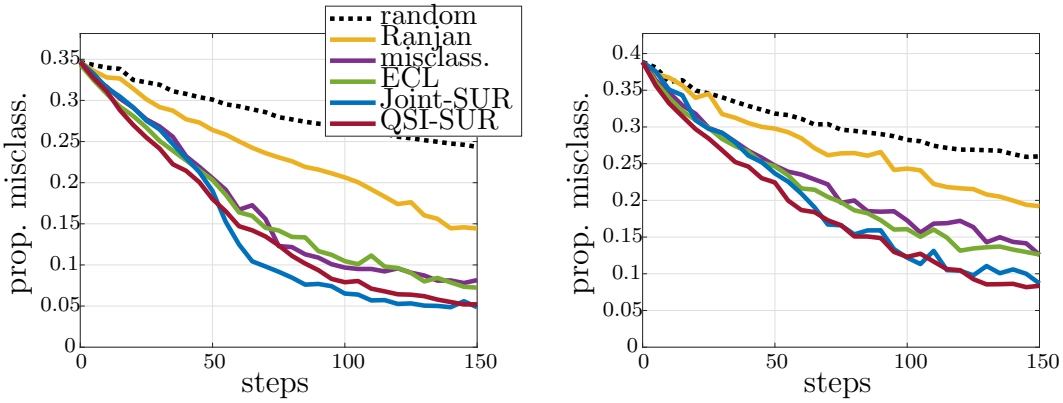

Figure SM11: Quantiles of level 0.75 and 0.95 for the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test case volcano.

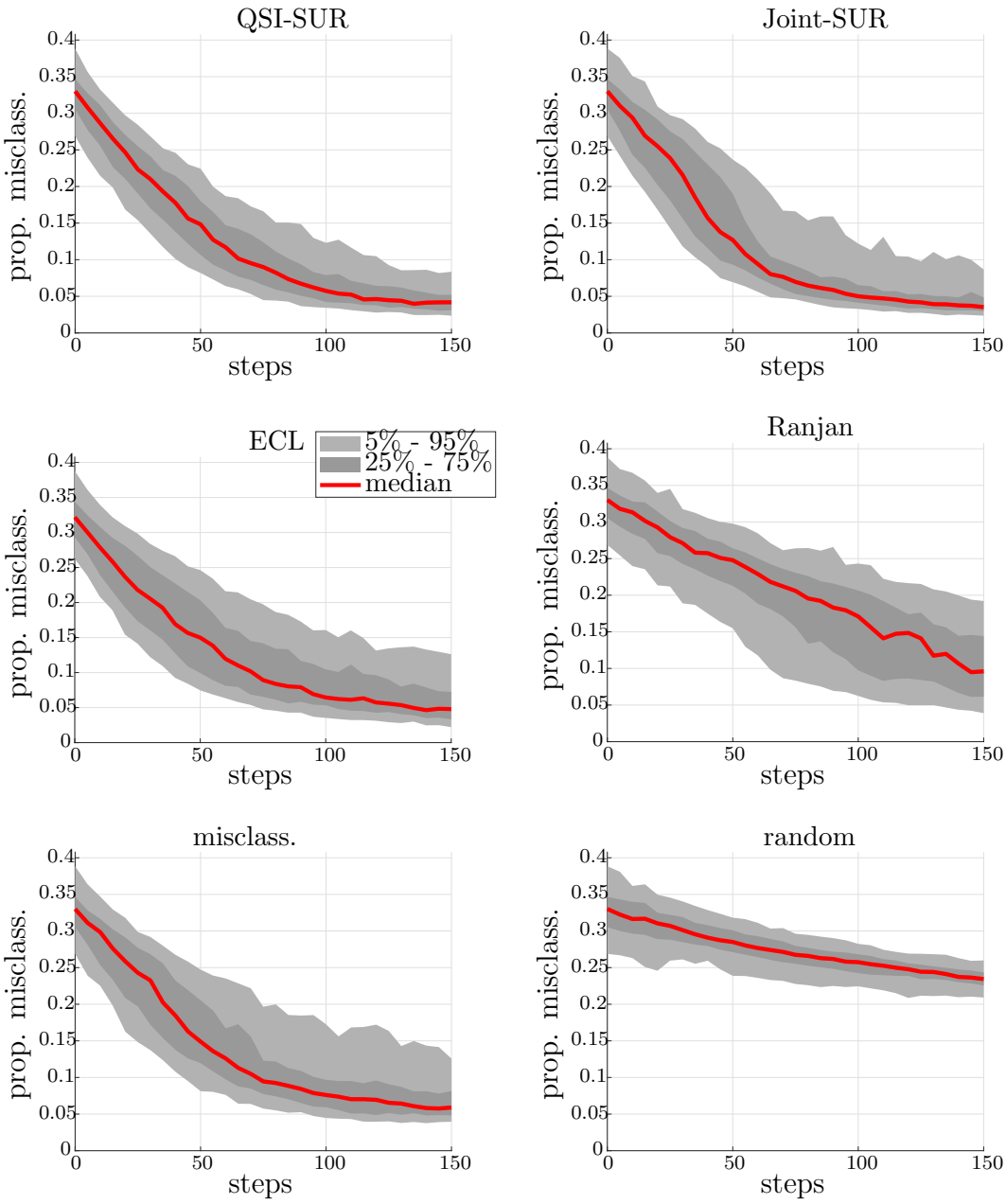

Figure SM12: Median and several quantiles of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test case volcano.

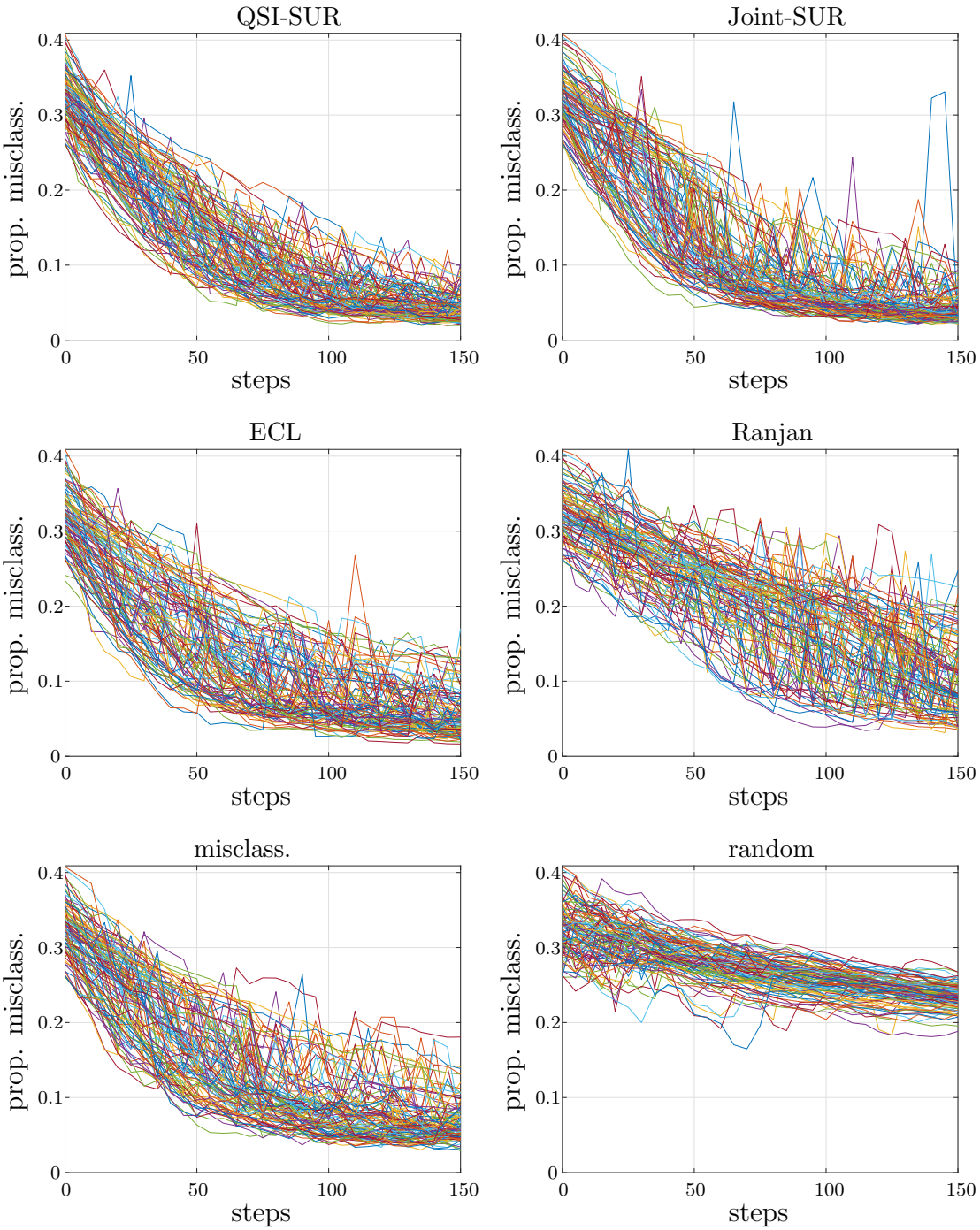

Figure SM13: Different sample paths of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the test case volcano.