



HAL
open science

Estimation of (small) reliable sets using a sequential Bayesian strategy

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

► **To cite this version:**

Romain Ait Abdelmalek-Lomenech, Julien Bect, Vincent Chabridon, Emmanuel Vazquez. Estimation of (small) reliable sets using a sequential Bayesian strategy. MASCOT-NUM 2023, Apr 2023, Le Croisic, France. . hal-04370899

HAL Id: hal-04370899

<https://centralesupelec.hal.science/hal-04370899>

Submitted on 3 Jan 2024

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

ESTIMATION OF (SMALL) RELIABLE SETS USING A SEQUENTIAL BAYESIAN STRATEGY

1. Context & motivations

In engineering, many system design problems involve finding parameters for a system to operate under desired conditions. When time-consuming simulations are used to assess the operating conditions of such systems, it is essential to use search methods that minimize the number of simulations.

Here, we propose a **SUR sequential strategy** [Bec+12] for a problem we refer to as **“reliable set inversion” (RSI)** [Ait+22], in which the system has both **deterministic and uncertain inputs**. In this problem, the objective is to estimate the set of deterministic inputs such that the probability of the outputs exceeding a given level is lower than a threshold.

We specifically focus on the case where the set to retrieve is relatively **small** compared to the space of deterministic inputs.

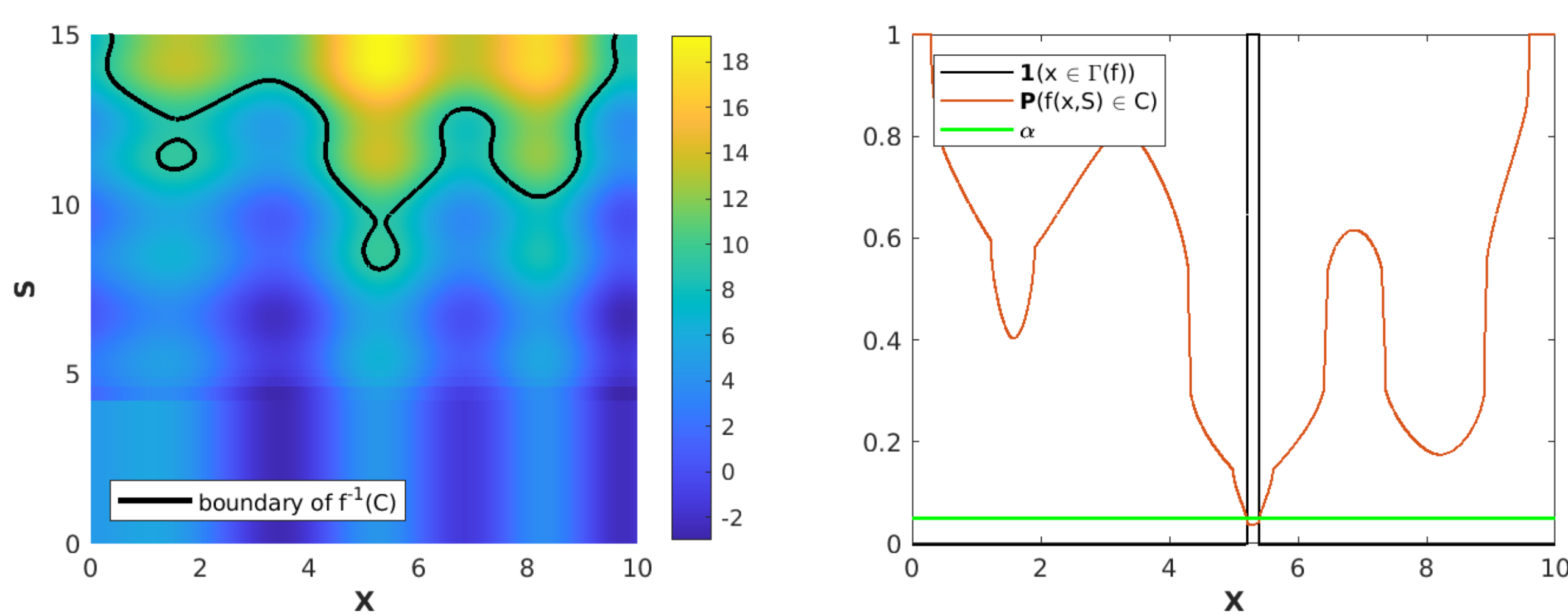
2. The reliable set inversion (RSI) problem

We view our system as a **black-box function** $f : \mathbb{X} \times \mathbb{S} \mapsto \mathbb{R}^q$. In this setting, \mathbb{X} is the space of deterministic inputs and \mathbb{S} the space of **uncertain** inputs.

- Given a critical region $C \subset \mathbb{R}^q$ and a threshold $\alpha \in]0, 1[$, **our objective is to estimate the “reliable” set**

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

where S is a random variable with **known distribution** \mathbb{P}_S on \mathbb{S} .



3. A SUR criterion for the RSI problem

- Assuming $\xi \sim \text{GP}(\mu, k)$ is a **Gaussian Process prior** on f , with μ and k respectively the mean and covariance functions of the process, define the **integrated process** τ and the associated **random excursion set** $\Gamma(\xi)$ as:

$$\tau(\xi, x) = \int_{\mathbb{S}} \mathbb{1}(\xi(x, s) \in C) d\mathbb{P}_S(s) \quad \text{and} \quad \Gamma(\xi) = \{x \in \mathbb{X}, \tau(\xi, x) \leq \alpha\}.$$

- Define an uncertainty measure \mathcal{H}_n :

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

with $\pi_n(x) = \mathbb{P}_n(\tau(\xi, x) \leq \alpha)$, where \mathbb{P}_n is the distribution of ξ given the current information

$$\mathcal{I}_n = \{(X_1, S_1, \xi(X_1, S_1)), \dots, (X_n, S_n, \xi(X_n, S_n))\}.$$

- The proposed SUR strategy chooses the point **minimizing the expected future uncertainty**:

$$(X_{n+1}, S_{n+1}) \in \arg \min_{(x, s) \in \mathbb{X} \times \mathbb{S}} \mathbb{E}_n(\mathcal{H}_{n+1} | X_{n+1} = x, S_{n+1} = s),$$

where \mathbb{E}_n is the conditional expectation with respect to \mathcal{I}_n .

4. An approximated sampling criterion

- Given a discrete approximation grid $\tilde{\mathbb{X}} \times \tilde{\mathbb{S}}$ and a subset $\Theta \subset \tilde{\mathbb{X}} \times \tilde{\mathbb{S}}$, define the discretized approximated process [Azz+16]

$$\tilde{\xi}(x, s) = \begin{cases} \xi(x, s) & \text{if } (x, s) \in \Theta \\ \mathbb{E}_n[\xi(x, s) | \xi(\Theta)] & \text{otherwise.} \end{cases}$$

- Monte Carlo simulations:** Given a candidate point $(\check{x}, \check{s}) \in \Theta$ and a quantization $(w_i, z_i)_{i=1, \dots, I}$ of the distribution of $\xi(\check{x}, \check{s})$ given \mathcal{I}_n , we can generate sample paths ξ_1^i, \dots, ξ_M^i of $\tilde{\xi}$ on $\tilde{\mathbb{X}} \times \tilde{\mathbb{S}}$ conditional to \mathcal{I}_n and $\{\xi(\check{x}, \check{s}) = z_i\}$.

Then we can write an approximated criterion \tilde{J}_n on Θ

$$\tilde{J}_n(\check{x}, \check{s}) = \sum_{i=1}^I \frac{w_i}{M} \sum_{m=1}^M \sum_{x \in \tilde{\mathbb{X}}} \mathbb{1}(\tau(\xi_m^i, x) \leq \alpha) \approx J_n(\check{x}, \check{s})$$

5. An SMC-based optimization algorithm

Considering the small volume of $\Gamma(f)$ relatively to \mathbb{X} , the idea is to build a sequence of sets of decreasing volume

$$\Gamma_1(f) \supset \Gamma_2(f) \supset \dots \supset \Gamma_T(f) = \Gamma(f)$$

such that $\Gamma_{t+1}(f)$ is “easy” to estimate knowing $\Gamma_t(f)$.

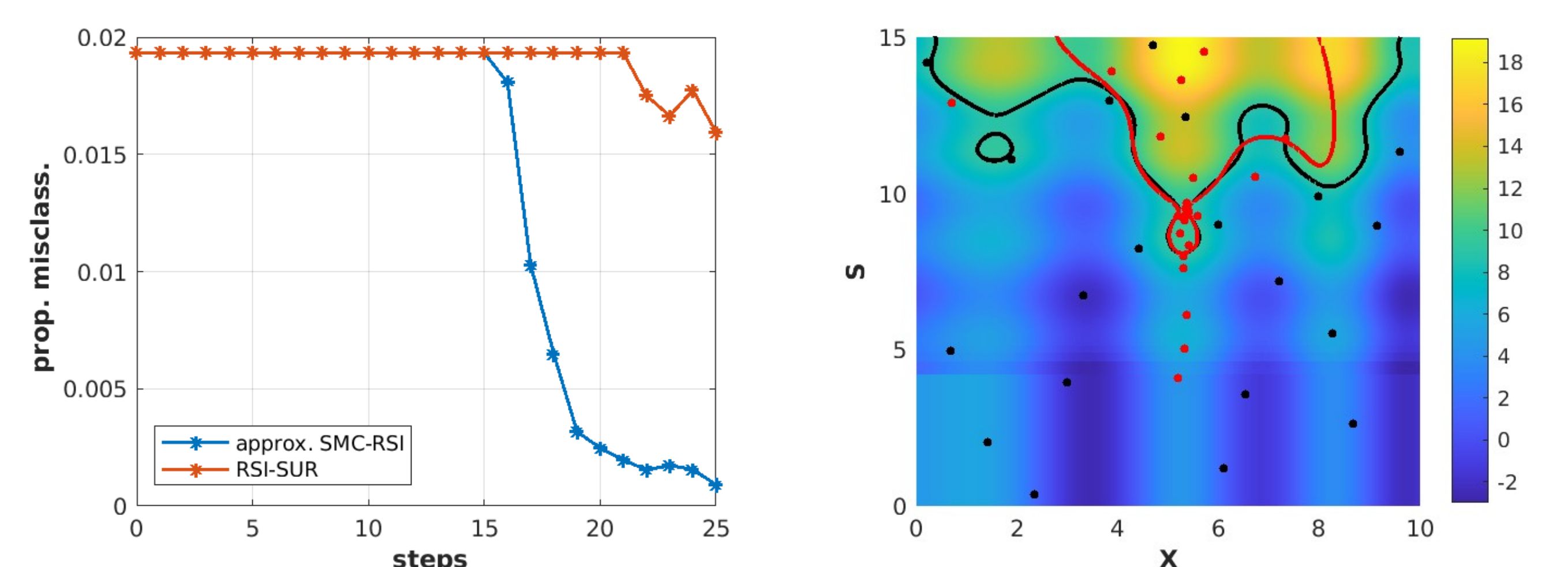
At each step n :

- Build a set $\tilde{\mathbb{S}}_n \subset \mathbb{S}$ whose points are i.i.d. according to \mathbb{P}_S .
- Generate a set of particles $\tilde{\mathbb{X}}_n \subset \mathbb{X}$ using sequential Monte Carlo with target distribution $\pi_n^t(x) = \mathbb{P}_n(x \in \Gamma_t(\xi))$
 - Approximate $q_n(x) \propto \frac{\pi_n^t(x)}{\pi_{n-1}^t(x)}$ for all $x \in \tilde{\mathbb{X}}_n$.
 - Sample n_X particles according to $\sum_{x \in \tilde{\mathbb{X}}_n} q_n(x) \delta_{\{x\}}$ (residual resampling).
 - Move the particles according to a random walk Metropolis-Hastings algorithm.
- Draw a subset $\Theta \subset \tilde{\mathbb{X}}_n \times \tilde{\mathbb{S}}_n$ according to the distribution proportional to $\min(p_n(x, s), 1 - p_n(x, s))$, with $p_n(x, s) = \mathbb{P}_n(\xi(x, s) \in C)$.
- Minimize \tilde{J}_n on Θ .
- If $\min_{(x, s) \in \Theta} \tilde{J}_n(x, s)$ is less than a threshold, set

$$\Gamma_{t+1}(f) = \left\{ x \in \mathbb{X} : \mathbb{P}(f(x, S) \in C) \leq q_\beta \left(\mathbb{P}_n(\xi(x, S) \in C), x \in \tilde{\mathbb{X}}_n \right) \right\}.$$

6. Numerical results

We compare, on an artificial example, the median of the proportion of misclassified points obtained with our improved method to the one obtained with the previous RSI-SUR sampling criterion.



References

- [Ait+22] R. Ait Abdelmalek-Lomenech et al. *Bayesian sequential design of computer experiments to estimate reliable sets*. 2022. arXiv: 2211.01008 [stat.ML].
[Azz+16] D. Azzimonti et al. “Quantifying Uncertainties on Excursion Sets Under a Gaussian Random Field Prior”. In: *SIAM/ASA Journal on Uncertainty Quantification* 4.1 (2016), pp. 850–874.
[Bec+12] J. Bect et al. “Sequential design of computer experiments for the estimation of a probability of failure”. In: *Statistics and Computing* 22.3 (2012), pp. 773–793.