

Extension of AK-MCS for the efficient computation of very small failure probabilities

Nassim Razaaly, Pietro Marco Congedo

▶ **To cite this version:**

Nassim Razaaly, Pietro Marco Congedo. Extension of AK-MCS for the efficient computation of very small failure probabilities. [Research Report] RR-9286, INRIA Saclay - Equipe DeFI. 2019. hal-02265311

HAL Id: hal-02265311
https://inria.hal.science/hal-02265311

Submitted on 12 Aug 2019

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.

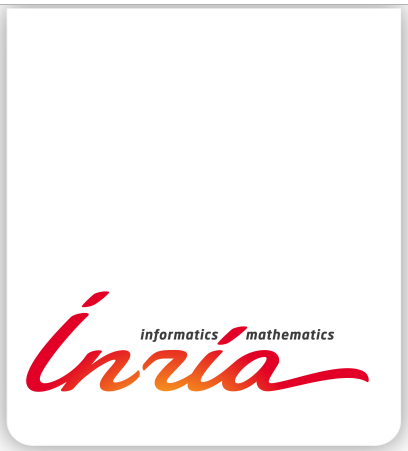


Extension of AK-MCS for the efficient computation of very small failure probabilities

Nassim Razaaly , Pietro Marco Congedo

RESEARCH REPORT
N° 9286
August 11, 2019
Project-Teams DEFI

ISSN 0249-6399 ISRN INRIA/RR--9286--FR+ENG



Extension of AK-MCS for the efficient computation of very small failure probabilities

Nassim Razaaly **, Pietro Marco Congedo *

Project-Teams DEFI

Research Report n° 9286 — August 11, 2019 — 39 pages

* DeFI Team, CMAP - Centre de Mathématiques Appliquées - Ecole Polytechnique, Inria Saclay - Ile de France

RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

1 rue Honoré d'Estienne d'Orves
Bâtiment Alan Turing
Campus de l'École Polytechnique

Abstract: We consider the problem of estimating a probability of failure p_f, defined as the volume of the excursion set of a complex (*e.g.* output of an expensive-to-run finite element model) scalar performance function J below a given threshold, under a probability measure that can be recast as a multivariate standard gaussian law using an isoprobabilistic transformation. We propose a method able to deal with cases characterized by multiple failure regions, possibly very small failure probability p_f (say $\sim 10^{-6} - 10^{-9}$), and when the number of evaluations of J is limited. The present work is an extension of the popular Kriging-based active learning algorithm known as AK-MCS, as presented in [1], permitting to deal with very low failure probabilities. The key idea merely consists in replacing the Monte-Carlo sampling, used in the original formulation to propose candidates and evaluate the failure probability, by a centered isotropic Gaussian sampling in the standard space, which standard deviation is iteratively tuned. This *extreme* AK-MCS (eAK-MCS) inherits its former multi-point enrichment algorithm allowing to add several points at each iteration step, and provide an estimated failure probability based on the Gaussian nature of the Kriging surrogate. Both the efficiency and the accuracy of the proposed method are showcased through its application to two to eight dimensional analytic examples, characterized by very low failure probabilities: $p_f \sim 10^{-6} - 10^{-9}$. Numerical experiments conducted with *unfavorable* initial *Design of Experiment* suggests the ability of the proposed method to detect failure domains.

Key-words: Low Failure Probability AK-MCS Importance Sampling Rare Events

Extension de AK-MCS pour le calcul efficace de très faibles probabilités d'échec

Résumé : Nous considérons le problème de l'estimation d'une probabilité de défaillance p_f, définie comme le volume de l'ensemble d'excursion d'une fonction scalaire (sortie d'un modèle coûteux à exécuter) J au-dessus d'un seuil donné, sous une mesure de probabilité qui peut être reformulée comme une loi gaussienne standard multivariée en utilisant une transformation iso-probabiliste. Nous proposons une méthode capable de traiter les cas caractérisés par plusieurs régions de défaillance, éventuellement une très petite probabilité de défaillance p_f (par exemple $sim10^{-6} - 10^{-9}$), et lorsque le nombre d'évaluations de J est limité. Le présent travail est une extension du célèbre algorithme dapprentissage actif basé sur Kriging, appelé AK-MCS, présenté dans [1], permettant de traiter des faibles probabilités déchec. L'idée principale consiste simplement à remplacer l'échantillonnage de Monte-Carlo, utilisé dans la formulation d'origine pour proposer des candidats et évaluer la probabilité de défaillance, par un échantillonnage gaussien centré isotrope dans l'espace standard, dont l'écart-type itérativement modifié. Cet *extrême* AK-MCS (eAK-MCS) hérite de son ancien algorithme d'enrichissement multi-points, permettant d'ajouter plusieurs points à chaque étape d'itération et de fournir une probabilité de défaillance estimée basée sur la nature gaussienne du métamodèle (Kriging). Lefficacité et la précision de la méthode proposée sont illustrées par son application à des exemples analytiques de deux à huit dimensions, caractérisées par de très faibles probabilités de défaillance: $p_f \sim 10^{-6} - 10^{-9}$. Des expériences numériques menées avec un plan d'expérience initial défavorable suggèrent la capacité de la méthode proposée à détecter les domaines de défaillance.

Mots-clés : Faible Probabilité de défaillanceAK-MCS Echantillonage d'Importance Evènements Rares

1. Introduction

Various methods have been proposed in the literature to estimate a probability of failure and the reader may refer to [2] for a critical review.

A first popular class of methods relies on the notion of so-called *Most Probable Failure Point* (MPFP), defined in the standard space as the closest point to the origin, belonging to the failure domain. Local approximation of the expensive-to-evaluate Limit-State Function (LSF) through a linear or quadratic Taylor expansion around this MPFP lead, respectively, to FORM and SORM, including in cases involving multiple MPFPs [3]. The FORM step require the evaluation of the gradient of the performance function, while SORM is based on the computation of the Hessian at the MPFP. Both are generally obtained by means of a finite difference scheme, known to be less efficient in high dimensional space, and subject to numerical instabilities. Morevover, FORM/SORM do not assess the error resulting from the Taylor series assumption.

Sampling methods include the well-known Monte-Carlo (MC) characterized by a low convergence rate, Importance Sampling (IS) [4, 5] relying on a prudent choice of the importance sampling density and Subset Simulation (SS) [6].

Surrogate-based methods rely on the substitution of the LSF by a metamodel, orders of magnitude faster to evaluate. The approximate model can be used in conjunction with sampling methods, to improve the latter or to correct the potential bias due to the surrogate model, such as AK-MCS [7], AK-MCSi [8], AK-IS [9], Meta-IS [10], MetaAK-IS^2 [11], KAIS [12], AK-SS [13], ^{2}SMART [14] and ASVR [15].

In particular, the AK-MCS algorithm as described in [1] presents attractive feature w.r.t. the original AK-MCS [7], such as a multipoint refinement strategy based on a clustering technique among samples generated with MC, and a stopping criterion based on the Kriging surrogate accuracy.

Amonsgt aforementioned algorithms, only a few permit to deal with very small failure probabilities ($10^{-5} - 10^{-9}$) and multiple failure regions: Meta-IS, BSS, ASVR, ^{2}SMART and AK-MCSi. Some other methods such as SORM or AK-IS are suitable for very small failure probabilities, but rely on the existence of an assumed unique MPFP.

In this paper, we propose an *extreme* version of AK-MCS [1], called eAK-MCS, in order to evaluate very small failure probabilities. It consists mainly in reformulating the MCS-based samples generation and the failure proba-

bility estimation. Indeed a centered isotropic Gaussian distribution, whose standard deviation is iteratively tuned, is used to generate the candidate samples for the refinement step; the resulting surrogate adaptive strategy is suitable for very small failure probabilities, while AK-MCS would require an untractable number of candidate samples. The surrogate exploitation stp aiming at estimating the fialure probability could potentially be performed with any suitbale sampling method (IS, SS...); here, IS is employed based on the centered isotropic Gaussian distribution. Unlike AK-MCS [1], the proposed method requires to work in the standard space, resorting if necessary to an isoprobabilistic transformation.

Additionally, the approach followed by eAK-MCS is significantly different from the one adopted in AK-MCSi: the latter proposes to split the large MC population (required by standard AK-MCS) into several populations of smaller sizes, and thus, to perform sequential MC simulations. Although adressing the RAM issue, AK-MCSi still requires to perform a considerable number of metamodels evaluations (similarly to AK-MCS), impacting significantly the computational burden.

The main advantage of eAK-MCS relies on its strong similarities w.r.t. AK-MCS, which refinement strategy has been derived to other fields such as quantile estimation [1] and (quantile-based) optimization under uncertainty [16]. As a consequence, eAK-MCS could enable the adaptation of those methodologies for very small probabilities. Another interesting feature is that eAK-MCS aims at refining *directly* the LSS, while for instance, BSS or ^{2}SMART adaptively improve LSSs corresponding to intermediate thresholds since those two methods are combined with SS. Finally, its simplicity presents a particularly favorable compromise between efficiency/accuracy and ease of implementation.

The proposed paper is structured as follows. Section 2 presents the problem description and general concepts such as Gaussian Processes and the theory of Importance Sampling. In Section 3, the proposed algorithm, *extended* AK-MCS (eAK-MCS), is described. Academic application examples are treated in Section 4. Conclusions are finally drawn in Section 5.

2. General Definitions

This section is devoted to introduce some concepts and to provide an introduction to some of the methods used in the proposed algorithm. The

problem of interest is formulated in Subsection 2.1. Gaussian Processes are concisely described in Subsection 2.2. Model accuracy and IS theory are commented respectively in Subsections 2.3 and 2.4.

2.1. Problem Description

Let us introduce a probabilistic model, described by a physical d-dimensional random vector $\mathbf{Y}$ fully defined by its joint probability density function (PDF) $f_{\mathbf{Y}}$, and a performance function $J : \mathbb{Y} \subseteq \mathbb{R}^d \to \mathbb{R}$, representing the system response. In reliability analysis, the failure probability p_f is defined as the probability that the model response J is smaller than a threshold value u:

$$p_f = \mathbb{P}_{\mathbf{Y}}(J(\mathbf{Y}) < u), \tag{1}$$

$\mathbb{P}_{\mathbf{Y}}$ denoting the probability measure induced by the random vector $\mathbf{Y}$. It is assumed possible to recast the problem (Eq. 1) using an isoprobabilistic transformation T (e.g. Rosenblatt/Nataf transform [17, 18, 19]), used to define the standard random variables $\mathbf{X} = T(\mathbf{Y})$ and the performance function $G = J \circ T^{-1}$ in the standard space. This latter point is crucial for the viability of the present approach, and might constitute a limitation for some industrial cases. We recall that $\mathbf{X} \sim \mathcal{N}(0, I_d)$ is the d-dimensional standard Gaussian vector, described by its PDF denoted as $f_{\mathbf{X}} = f_{\mathcal{N}(0,I_d)}$. The failure probability reads then

$$p_f = \mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) < u) = \mathbb{E}_{\mathbf{X}}[\mathbb{1}_{G<u}(\mathbf{X})] = \int_{\mathbb{R}^d}, \mathbb{1}_{G<u}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \tag{2}$$

$\mathbb{1}_{G<u}$ being the indicator function such that $\mathbb{1}_{G<u} = 1$ for $G < u$, $\mathbb{1}_{G<u} = 0$ otherwise; $\mathbb{P}_{\mathbf{X}}$ and $\mathbb{E}_{\mathbf{X}}$ denote respectively the probability measure and the expectation operator induced by the random vector $\mathbf{X}$.

2.2. Gaussian Processes

Computing a metamodel requires the construction of a *Design of Experiments* (DoE), denoted here by $\mathcal{X} = \{\mathbf{x}_1, ..., \mathbf{x}_m\}$. We choose to build a Gaussian Process (GP) of the performance function G; its *best linear unbiased estimation* (BLUE, see [20, 21]) at an unknown location $\mathbf{x}$ is shown to be a Gaussian random variable $\hat{G}(\mathbf{x})$:

$$\hat{G}(\mathbf{x}) \sim \mathcal{N}(\mu_{\hat{G}}(\mathbf{x}), \sigma^2_{\hat{G}}(\mathbf{x})),$$

where $\mathcal{N}$ refers to the univariate gaussian law, which PDF and Cumulative Density Function (CDF) are denoted respectively by ϕ and Φ. $\mu_{\hat{G}}(\mathbf{x})$ and $\sigma^2_{\hat{G}}(\mathbf{x})$ refer respectively to the so-called predictive mean and variance of the GP at $\mathbf{x}$. The metamodel of the original performance function is denoted by:

$$\tilde{G}(\mathbf{x}) = \mu_{\hat{G}}(\mathbf{x}).$$

The stationary anisotropic Matérn covariance function with regularity $\frac{5}{2}$ is used, the hyperparameters being tuned by maximizing the likelihood [22]. It is equivalent to the Simple Kriging formulation, assuming a known null mean. Note that the algorithm presented in this paper is obviously also compatible with any GP-based metamodel, for instance with more sophisticated trends (e.g. PC-Kriging [1]) or kernels. In cases where an isoprobabilistic transform T is used to recast the *physical* problem into the *standard* one, it might be recommended to build a surrogate in the *physical* space, namely for $J(\mathbf{y})$, obtaining: $\tilde{G}(\mathbf{x}) = \mu_{\hat{G}}(\mathbf{x}) = \mu_{\hat{J}}(\mathbf{y})$ and $\sigma_{\hat{G}}(\mathbf{x}) = \sigma_{\hat{J}}(\mathbf{y})$, where $\mathbf{y} = T^{-1}(\mathbf{x})$. Considering for instance bounded input distributions in the physical space, the isoprobabilistic transform T might be highly non-linear. Consequently, a linear Limit-State Surface (LSS) easy to approximate in the physical space might result in a LSS characterized by a intricate geometry.

2.3. Model Accuracy

Accounting for the prediction uncertainty in the GP model, we define [1, 23] the lower $\{\mathbf{x} \in \mathbb{R}^d : \ \mu_{\hat{G}}(\mathbf{x}) - k\sigma_{\hat{G}}(\mathbf{x}) = u\}$ and upper $\{\mathbf{x} \in \mathbb{R}^d : \ \mu_{\hat{G}}(\mathbf{x}) + k\sigma_{\hat{G}}(\mathbf{x}) = u\}$ boundaries of the *predicted* LSS $\{\mathbf{x} \in \mathbb{R}^d : \ \mu_{\hat{G}}(\mathbf{x}) = u\}$, where k sets the confidence level, (typically $1.96 = \Phi^{-1}(97.5\%)$). Analogously, the lower and upper bounds, and *predicted* failure domains are respectively defined as:

$$\begin{aligned}
\mathcal{D}_f^{(k)-} &= \{\mathbf{x} \in \mathbb{R}^d : \ \mu_{\hat{G}}(\mathbf{x}) + k\sigma_{\hat{G}}(\mathbf{x}) < u\} \\
\mathcal{D}_f^{(k)+} &= \{\mathbf{x} \in \mathbb{R}^d : \ \mu_{\hat{G}}(\mathbf{x}) - k\sigma_{\hat{G}}(\mathbf{x}) < u\} \\
\mathcal{D}_f^{0} &= \{\mathbf{x} \in \mathbb{R}^d : \ \mu_{\hat{G}}(\mathbf{x}) < u\}
\end{aligned} \tag{3}$$

with $\mathcal{D}_f^{(k)-} \subset \mathcal{D}_f^0 \subset \mathcal{D}_f^{(k)+}$. The *predicted* failure probability $p_{\tilde{f}}$ and its lower and upper bounds, resp. $p_{\tilde{f}}^{(k)-}$ and $p_{\tilde{f}}^{(k)+}$ are defined as:

$$\begin{aligned} p_{\tilde{f}}^{(k)-} &= \mathbb{P}_{\mathbf{X}}(\mu_{\hat{G}}(\mathbf{X}) + k\sigma_{\hat{G}}(\mathbf{X}) < u) &&= \mathbb{E}_{\mathbf{X}}[\mathbb{1}_{\mu_{\hat{G}}+k\sigma_{\hat{G}}<u}(\mathbf{X})] \\ p_{\tilde{f}}^{(k)+} &= \mathbb{P}_{\mathbf{X}}(\mu_{\hat{G}}(\mathbf{X}) - k\sigma_{\hat{G}}(\mathbf{X}) < u) &&= \mathbb{E}_{\mathbf{X}}[\mathbb{1}_{\mu_{\hat{G}}-k\sigma_{\hat{G}}<u}(\mathbf{X})] \\ p_{\tilde{f}} &= \mathbb{P}_{\mathbf{X}}(\mu_{\hat{G}}(\mathbf{X}) < u) &&= \mathbb{E}_{\mathbf{X}}[\mathbb{1}_{\mu_{\hat{G}}<u}(\mathbf{X})] \end{aligned} \tag{4}$$

The so-called *Limit State Margin* (LSM) [1] $\mathbb{M}_f^{(k)} = \mathcal{D}_f^{(k)+} \backslash \mathcal{D}_f^{(k)-}$ is a natural region where to focus the design enrichment. Given a set of samples $S = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, $S^{(k)} = S \cap \mathbb{M}_f^{(k)}$ refers to samples in S belonging to the LSM $\mathbb{M}_f^{(k)}$, containing points of interest for enrichment among S, since they should lie close to the true LSS. Note that $S^{(k)}$ can be empty for a given set S.

2.4. Importance Sampling Theory

We recall the IS method, by considering the following generic estimation:

$$p_g = \mathbb{E}_{\mathbf{X}}[g(\mathbf{X})], \tag{5}$$

where g refers, for instance, to $\mathbb{1}_{\mu_{\hat{G}}<u}(\mathbf{x})$, $\mathbb{1}_{\mu_{\hat{G}}+k\sigma_{\hat{G}}<u}(\mathbf{x})$ or $\mathbb{1}_{\mu_{\hat{G}}-k\sigma_{\hat{G}}<u}(\mathbf{x})$. Let h be a proposal PDF, a.k.a. biasing/instrumental PDF or Importance Sampling Density (ISD) assumed to dominate $gf_{\mathbf{X}}$ in the absolutely continuous sense:

$$h(\mathbf{x}) = 0 \implies g(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) = 0, \ \forall \mathbf{x} \in \mathbb{R}^d. \tag{6}$$

Then, p_g may be rewritten as follows:

$$p_g = \int_{\mathbb{R}^d} g(\mathbf{x}) \frac{f_{\mathbf{X}}(\mathbf{x})}{h(\mathbf{x})} h(\mathbf{x}) d\mathbf{x} = \mathbb{E}_h \left[g(\mathbf{X}) \frac{f_{\mathbf{X}}(\mathbf{X})}{h(\mathbf{X})} \right]. \tag{7}$$

It easily leads to the *importance sampling estimator*:

$$\hat{p}_g = \frac{1}{N} \sum_{i=1}^{N} g(\mathbf{x}_i) \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{h(\mathbf{x}_i)}, \tag{8}$$

where $\mathbf{x}_1, ..., \mathbf{x}_N \overset{\text{iid}}{\sim} h$. This estimator is unbiased and its quality may be measured by means of its variance estimator:

$$\hat{\sigma}_g^2 = \frac{1}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N} g(\mathbf{x}_i)^2 \frac{f_{\mathbf{X}}(\mathbf{x}_i)^2}{h(\mathbf{x}_i)^2} - \hat{p}_g^2\right). \tag{9}$$

The corresponding Coefficient of Variation (CoV) $\hat{\delta}_g$, quantifying the estimation accuracy, is defined by:

$$\hat{\delta}_g = \frac{\hat{\sigma}_g}{\hat{p}_g}, \tag{10}$$

provided $\hat{p}_g \neq 0$.

The accuracy of the approximation given by IS critically depends on the choice of the ISD h. In this work, we mainly focus on ISDs in the form $\mathcal{N}(0, \gamma^2 I_d)$, with $\gamma > 0$. A discussion on how to tune the parameter γ can be found in Subsection 3.3. Note that a gaussian mixture ISD with suitably tuned parameters might be used [24].

3. The eAK-MCS Algorithm

In this section, we describe the method proposed in this paper, i.e. the eAK-MCS algorithm, aiming at building a GP-based surrogate, refining it iteratively in the LSS $\{G(\mathbf{x}) = u\}$, and estimating the *predicted* failure probability $\hat{p}_{\tilde{f}}$ as an approximation of p_f.

The main steps can be summarized as follows (similarly to AK-MCS [1]):

1. *Initial DoE*: An experimental design $\mathcal{X}$ is generated by Latin-Hypercube Sampling (LHS) (Subsection 3.1).
2. *Metamodel Update*: The exact response $\mathcal{Y}$ of the exact performance function G is carried out on $\mathcal{X}$. The metamodel is calibrated based on $\{\mathcal{X}, \mathcal{Y}\}$ (Subsection 2.2).
3. *Candidates*: A set of N_C candidate points $\mathcal{S}$ is generated (Subsection 3.2).
4. *Selection Step*: The *selection step* determines the sample(s) $\mathbf{x}^*$ to be added to the experimental design: $\mathcal{X} \leftarrow \{\mathcal{X}, \mathbf{x}^*\}$ (Subsection 3.4).
5. *Stopping Criterion*: If a stopping criterion is satisfied (Subsection 3.5), the enrichment stops. The failure probability is estimated using IS on the metamodel. Otherwise the algorithm goes back to step 2.

W.r.t. the AK-MCS algorithm as presented in [1], the fundamental difference lies in the generation of candidate points (Step 3) relying on an isotropic centered Gaussian distribution and is the main contribution of the paper. The exploitation step aiming at estimating the failure probability from the approximate model and a sampling method is presented here with IS associated to the aforementioned distribution as the ISD, while any other sampling method such as SS or IS with another ISD could be used.

The candidate selection procedure and the stopping criterion, described respectively in Step 4 and 5 are inspired mainly from [1] but slightly modified and suggested here to improve the efficiency of the proposed algorithm. Finally, in Section 3.6, we indicate some typical values for the tuning parameters.

3.1. Initial DoE

An initial design of size $n_0 = 5d$ [25, 23, 10] is generated in the standard space, as described in [25]. A compact subset $\mathbb{X}_0 = \prod_{i=1}^{d}[q_\epsilon^i, q_{1-\epsilon}^i]$ is constructed, where q_ϵ^i and $q_{1-\epsilon}^i$ are respectively the quantiles of order ϵ and $1-\epsilon$ of the ith input variable. Working in the standard space, $\mathbb{X}_0$ simply reads $[\Phi(\epsilon), \Phi(-\epsilon)]^d$. A LHS design on $[0,1]^d$ of size n_0 (criterion *maximin* here) is then scaled to $\mathbb{X}_0$ using an affine mapping.

3.2. Candidates Generation

Using MC samples as candidates points as in the original AK-MCS algorithm [7, 1] is not suitable when p_f is very small: possibly only a few of them would lie in the LSM $\mathbb{M}_f^{(k)}$. To tackle this issue, samples are generated [24] from the centered uncorrelated multivariate gaussian $\mathcal{N}(0, \gamma^2 I_d)$, where the choice of $\gamma \geqslant 1$ is discussed in Section 3.3. This process is summarized below (Algorithm 1).

Note that too *distant* samples from the origin are discarded, the choice of

the user-defined parameter r_{max} (Step 2) being discussed in Subsection 3.6.

Algorithm 1: Candidates Generation

Input: $N_C,\ r_{max},\ p,\ \gamma$
Output: Set $\mathcal{S}$ of N_C samples

1 Generation of pN_C samples $\{\mathbf{x}_1, ..., \mathbf{x}_{pN_C}\} \overset{\text{iid}}{\sim} \mathcal{N}(0, \gamma^2 I_d)$.
2 Discard Distant samples:
$I = \{i \in [1, pN_C] \text{ s.t. } \|\mathbf{x}_i\|_2 < r_{max}\} = \{i_1, ..., i_{|I|}\}$ with $i_k < i_{k+1}$ and $|I| \geqslant N_C$. (If not, increase p).
3 Candidate: $\mathcal{S} = \{\mathbf{x}_{i_1}, ..., \mathbf{x}_{i_{N_C}}\}$.

3.3. Choice of γ

We describe hereafter a procedure to tune the parameter γ, associated to the ISD $f_{\mathcal{N}(0,\gamma^2 I_d)}$ used both for the candidate samples generation, and the surrogate exploitation, namely the failure probability estimation based on IS. The selected approach is based on the following comment. In the context of IS for failure probability estimation using an ISD belonging to the parametric family $\{f_{\mathcal{N}(0,\gamma^2 I_d)},\ \gamma > 0\}$, one would seek for γ^* leading to an *optimal* ISD. A natural approach consists in finding the value γ^* that minimizes the variance estimator (Eq. 9), sometimes referred to as the *Variance Minimization Method* [26]. Inspiring from the latter, the procedure used in this work seeks to minimize the IS CoV (Eq. 10). The first step consists in generating $\mathcal{N}_\gamma$ samples $\{\mathbf{x}_i\}_i \overset{iid}{\sim} f_{\mathcal{N}(0,I_d)}$, while it is worthy pointing out that $\{\gamma\mathbf{x}_i\}_i \overset{iid}{\sim} f_{\mathcal{N}(0,\gamma^2 I_d)}$, for $\gamma > 0$. It is then possible to *define* the functions $\gamma \mapsto \text{CoV}(\gamma)$ and $\gamma \mapsto \text{CoV}^+(\gamma)$ returning the IS CoV obtained from the estimation of $p_{\tilde{f}}$ and $p_{\tilde{f}}^{(k)+}$ respectively, based on the IS samples $\{\gamma\mathbf{x}_i\}_i$ and corresponding surrogate evaluations. Note that for some given surrogate model and γ, $\text{CoV}(\gamma)$ or $\text{CoV}^+(\gamma)$ might be undefined, indicating that all samples $\{\gamma\mathbf{x}_i\}_i$ are considered to belong to the *safe* domain by the surrogate.

The chosen value γ^* is then set to

$$\gamma^* = \underset{\gamma \in [\gamma_{min}, \gamma_{max}]}{\arg\min}\ \text{CoV}(\gamma) \tag{11}$$

with $\gamma_{min},\ \gamma_{max} > 0$ (Table 1), under existence. Otherwise,

$$\gamma^* = \underset{\gamma \in [\gamma_{min}, \gamma_{max}]}{\arg\min}\ \text{CoV}^+(\gamma). \tag{12}$$

If the latter is not defined, γ^+ is set to a default value γ_0 (Table 1). The optimization steps (Eq. 11-12) are approximately solved, discretizing uniformly the interval $[\gamma_{min}, \gamma_{max}]$ in n_γ values. This operation requires at worst $2N_\gamma n_\gamma$ surrogate evaluations, and is performed anytime the metamodel is updated.

3.4. DoE Selection

Considering a set $\mathcal{S}$ of candidate points as described in Algorithm 1, we follow a strategy very similar to [1] to select the point(s) to be added to the DoE.

3.4.1. Single Sample Selection: Single eAK-MCS

Due to the Gaussian nature of the metamodel, to each sample $\mathbf{x} \in \mathbb{R}^d$ corresponds the so-called probability of misclassification $P_m(\mathbf{x})$ defined as the non-zero probability that the gaussian predictor $\hat{G}(\mathbf{x})$ exceeds the threshold u (safe) while the prediction mean $\mu_{\hat{G}}(\mathbf{x})$ is below u (failure) or vice versa. It can be written as [27]:

$$P_m(\mathbf{x}) = \Phi\left(-\frac{|\mu_{\hat{G}}(\mathbf{x}) - u|}{\sigma_{\hat{G}}(\mathbf{x})}\right) \tag{13}$$

The so-called U-function U is then defined as the *reliability index* linked to the probability of misclassification P_m [28, 7, 1]:

$$U(\mathbf{x}) = \frac{|\mu_{\hat{G}}(\mathbf{x}) - u|}{\sigma_{\hat{G}}(\mathbf{x})}. \tag{14}$$

Similarly to AK-MCS, the sample used to enrich the DoE among the candidates $\mathcal{S}$ is the one minimizing the U-function, thus maximizing its probability of being misclassified:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{S}} U(\mathbf{x}) = \arg\max_{\mathbf{x}\in\mathcal{S}} P_m(\mathbf{x}) \tag{15}$$

Note that P_m and U are associated to the threshold u, which is not clearly indicated with the notation for a sake of clarity.

3.4.2. Multiple Sample Selection

If parallel computing is available, and $(K+1)$ samples can be added simultaneously to the DoE, we propose the following strategy, inspired from [1] but different in steps 1, 2 and 3:

1. One sample $\mathbf{x}_0^*$ is selected among $\mathcal{S}$ following the single eAK-MCS selection (Eq. 15).
2. K samples $(\mathbf{x}_1^*, ..., \mathbf{x}_K^*)$ are simultaneously selected among the margin set $\mathcal{S}^{(k)}$ belonging to the LSM $\mathbb{M}_f^{(k)}$, using a clustering technique [1], the weighted K-means algorithm, detailed hereafter for a sake of clarity. If this method returns only $K_1 < K$ samples ($K_1 = 0$ possibly), then the very same method is applied to the extended set $\mathcal{S}$ to provide the remaining $K - K_1$ samples.
3. To avoid samples too close to each other (for preventing metamodel training issues), a filtering procedure is performed on the selected samples $\mathcal{X}^* = (\mathbf{x}_0^*, ..., \mathbf{x}_K^*)$, removing too close points: if $\|\mathbf{x}_i^* - \mathbf{x}_j^*\|_2 < TOL$, $\mathbf{x}_j^*$ is discarded from $\mathcal{X}^*$. It involves that the multiple selection strategy might return strictly less than $K + 1$ samples.

Weighted K-means Algorithm. Let $\mathcal{A} = (x_1, ..., \mathbf{x}_P)$ denoting a sample set, referring either to $\mathcal{S}$ or $\mathcal{S}^{(k)}$. In step 2, a weighted K-means clustering algorithm is used [29] accounting for the importance of the samples in $\mathcal{A}$. Following [1], the weights are set to the probability of misclassification (Eq. 13) of each sample $\mathbf{x} \in \mathcal{A}$, bounded on $[0, 0.5]$ by definition. The K samples are then selected as the clusters' centroids. More specifically, this clustering technique aims to minimize the total cluster variance defined as:

$$V_{\mathcal{A}} = \sum_{j=1}^{K} \sum_{l \backslash i_l = j} \|\mathbf{x}_j^* - \mathbf{x}_l\|_2 \tag{16}$$

where $i_l = \underset{j \in [\![1,K]\!]}{\arg\min} \|\mathbf{x}_j^* - \mathbf{x}_l\|_2$, indicates the index of the closest centroid $\mathbf{x}_{i_l}^*$ to a sample $\mathbf{x}_l$. $\{\mathbf{x}_j^*\}_j$ denote the K weigthed centroids defined as

$$\mathbf{x}_j^* = \frac{1}{\sum\limits_{l \backslash i_l = j} \omega_l} \sum_{l \backslash i_l = j} \omega_l \mathbf{x}_l \tag{17}$$

where $\omega_l = P_m(\mathbf{x}_l)$ [1] is the weight associated to the sample $\mathbf{x}_l$. Note that in AKMCSi [8], the chosen weights are $\omega_l = \dfrac{1}{U^2(\mathbf{x}_l)}$. $\mathbf{x} \mapsto \dfrac{1}{U^2(\mathbf{x})}$ and $\mathbf{x} \mapsto P_m(\mathbf{x})$ being characterized by the same monoticity, numeral experiments have not demonstrated clear advantages in choosing one or the other weight definition.

Setting $\omega_l = 1$ for all samples leads to the definition of the regular centroids, and by extension, to the regular K-means algorithm. The weighted K-means algorithm is summarized in Algorithm 2.

Remark. The set of candidate points $S^{(k)}$ belonging to the LSM $\mathcal{M}^{(k)}$ can contain less than K samples, possibly being empty. This fact explains why a point in S maximizing the U-function is also selected.

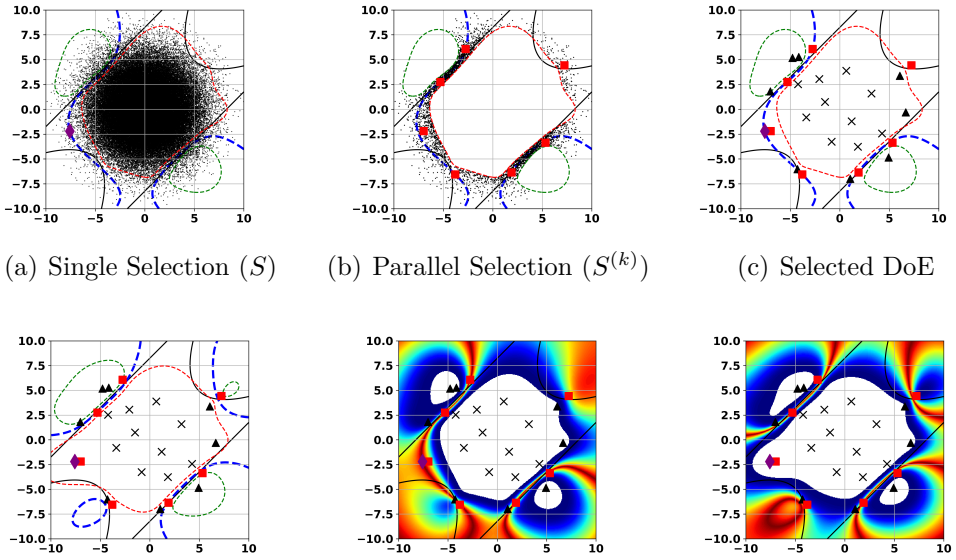
Algorithm 2: Weighted K-means Algorithm [1].

Input: samples $\mathcal{A} = (\mathbf{x}_1, ..., \mathbf{x}_P)$, TOL, $NMAX$, P_m, K
Output: K centroids $(\mathbf{x}_1^*, ..., \mathbf{x}_K^*)$

If $P < K$, return $\mathcal{A}$. Centroids Initialization $\mathcal{C}^{(0)} = (\mathbf{x}_1^{(0)}, ..., \mathbf{x}_K^{(0)})$: Regular K-means (or Random).
n=0 ; err=1 ;
while $n < NMAX$ *and* $err > TOL$ **do**
    $n \leftarrow n + 1$
    Assign a cluster i_l to each sample $\mathbf{x}_l$:
    $\forall l \in [1, P],\ i_l = \underset{j \in [\![1,K]\!]}{\arg\min} \|\mathbf{x}_j^* - \mathbf{x}_l\|_2$
    Update Weigthed centroids $\mathcal{C}^{(n)}$: Equation 17.
    Update Error: $err = \sum_{j=1}^{K} \|\mathbf{x}_j^{(n)} - \mathbf{x}_j^{(n-1)}\|_2$
Return $\mathcal{C}^{(n)}$

Illustrative example. To showcase the process of selecting multiple additional samples, we consider the two-dimensional four-branches series system example introduced in Subsubsection 4.1.2, characterized by $p_f \sim 5.596 \times 10^{-9}$ and initialized with a DoE of size 10. The process of selecting $K + 1 = 8$ samples (resp. $K + 1 = 64$) is illustrated in Figure 1 (resp. Figure 2).

In Figures 1(a) and 2(a), the small black dots indicate the candidate points S among which a single point minimizing the U-function is selected (purple diamond). The candidates points of $S^{(k)}$ belonging to the LSM $\mathcal{M}^{(k)}$ and extracted from S are indicated in small black dots in Figures 1(b) and 2(b), where selected samples (weighted centroids from the clustering technique Algorithm 2) are indicated in red squares. The set $S^{(k)}$ contains less points focused on the LSM (*exploitation*), while the point selected among the set S has an *exploratory* role.



(a) Single Selection (S) (b) Parallel Selection ($S^{(k)}$) (c) Selected DoE

(d) Metamodel Updated (e) Selected DoE(P_m) (f) Metamodel Updated(P_m)

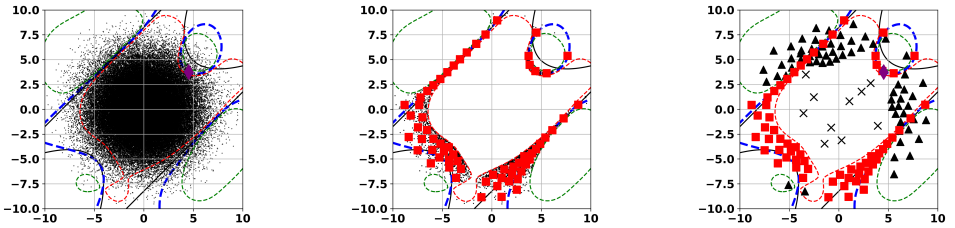
Figure 1: Illustration of the parallel refinement strategy: Selection of $K+1=8$ samples. Four-branch series system 2D (Subsection 4.1.2).Candidates points $S^{(k)}$ or S are represented by small black dots. LSS: The true LSS $\{G(\mathbf{x})=u\}$ is indicated by a black line, the *predicted* LSS $\{\mu_{\hat{G}}(\mathbf{x})=u\}$ by a dashed blue line, the lower LSS $\{\mu_{\hat{G}}(\mathbf{x})-k\sigma_{\hat{G}}(\mathbf{x})=u\}$ by a red dashed line and the upper LSS $\{\mu_{\hat{G}}(\mathbf{x})+k\sigma_{\hat{G}}(\mathbf{x})=u\}$ by a green dashed line. DoE: The initial DoE is indicated by grey crosses, the current DoE by black triangles, the point selected with single eAK-MCS by a purple diamond and selected points with weighted K-means by red squares. Contours of the probability of misclassification P_m are indicated in white when lower than 10^{-4}. Blue and red correspond respectively to 10^{-4} and 1.

3.5. Stopping Criterion

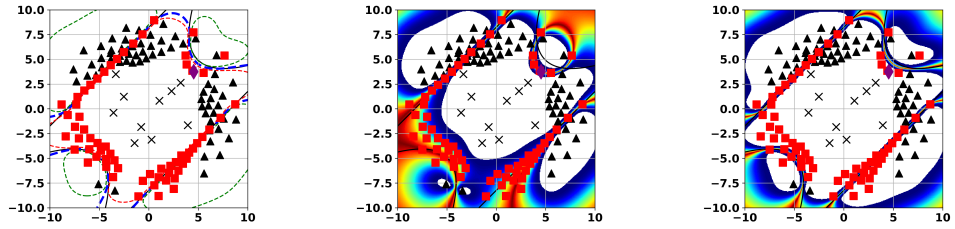
The stopping criterion adopted in this paper is the same as the one proposed in [1], focusing on the accuracy of the quantity of interest, consequently on the upper and lower bounds of the failure probability:

$$\text{Basic: } \frac{\hat{p}_{\tilde{f}}^{(\tilde{k})+} - \hat{p}_{\tilde{f}}^{(\tilde{k})-}}{\hat{p}_{\tilde{f}}} < \epsilon_{p_{\tilde{f}}}, \tag{18}$$

for two consecutive iteration steps, where $p_{\tilde{f}}$, $p_{\tilde{f}}^{(\tilde{k})+}$ and $p_{\tilde{f}}^{(\tilde{k})-}$ are estimated using IS based on the ISD $f_{\mathcal{N}(0,\gamma^{*2}I_d)}$ with N_γ^{IS} samples. Those estimations



(a) Single Selection (S) (b) Parallel Selection ($S^{(k)}$) (c) Selected DoE



(d) Metamodel Updated (e) Selected DoE(P_m) (f) Metamodel Updated(P_m)

Figure 2: Illustration of the parallel refinement strategy: Selection of $K+1=64$ samples. Four-branch series system 2D (Subsection 4.1.2). Legend settings in Figure 1.

can be performed with another ISD, such as the Gaussian mixture as introduced in [24]. However, it seems that in high dimension and for very small failure probabilities, the IS efficiency is reduced involving estimates subject to larger CoVs, thus, larger IS confidence intervals.

Modified versions of the *basic* stopping criterion (Eq. 18) reads:

$$\text{Conservative: } \frac{\hat{p}^{(\tilde{k})+}_{\tilde{f},max} - \hat{p}^{(\tilde{k})-}_{\tilde{f},min}}{p_{\tilde{f}}} < \epsilon_{p_{\tilde{f}}}, \tag{19}$$

$$\text{Fast: } \frac{\hat{p}^{(\tilde{k})+}_{\tilde{f},min} - \hat{p}^{(\tilde{k})-}_{\tilde{f},max}}{p_{\tilde{f}}} < \epsilon_{p_{\tilde{f}}}. \tag{20}$$

Note also that the value of k depends on the level of accuracy required and might be different from the one used in the process of generating candidate points selection (via LSM $\mathbb{M}_f^{(k)}$) or when tuning γ, for which we would recommend a large value (say $k=3$). A lower value denoted as $\tilde{k}$ is used when evaluating the stopping criterion.

3.6. eAK-MCS numerical settings

The tuning parameters mentioned in the method section are summarized in Table 1 with their suggested value, used in the numerical experiments (except when explicitly mentioned). The stopping criterion selected is systematically *Fast* (Eq. 20). r_{max}, the distance from the center above which candidate samples are discarded, is chosen so that a reliability problem involving a MPFP $\mathbf{x}^*$ s.t. $\| \mathbf{x}^* \| = r_{max}$ would lead to a negligible FORM failure probability estimate $p_f^{FORM} = 10^{-20}$.

$\tilde{k}$	k	n_0	ϵ	r_{max}	N_C	γ_0	γ_{min}	γ_{max}	N_γ	n_γ	N_γ^{IS}
1	3	$5d$	10^{-5}	$-\Phi^{-1}(10^{-20}) \approx 9.2$	10^6	2.5	1	5	10^5	15	10^7
p	d_{min}	$K+1$	TOL	$NMAX$	α	$\epsilon_{p_{\tilde{f}}}$					
5	10^{-4}	8	10^{-4}	100	2	5%					

Table 1: Tuning Parameters

4. Numerical experiments

In this section, we illustrate the capabilities of the algorithm through its application to several test-cases. Only cases involving $p_f \sim 10^{-5} - 10^{-9}$ are considered here, to showcase the suitability of the proposed algorithm to deal with very small failure probabilities (unlike AK-MCS). Reference values are estimated either with Subset Simulation [25] or by means of IS with a Gaussian mixture ISD [24].

eAK-MCS is systematically compared against FORM and SORM, using respectively SQP (Python implementation) and the asymptotic formula of Hohenbichler and Rackwitz [30], if not explicitely mentioned. Gradient and Hessian estimations are performed by finite differences with a step size of 10^{-4} assuming one iteration.

The test cases are introduced in Subsection 4.1.

A first study involving three 2D and one 6D examples consists of the assessment of the eAK-MCS algorithm with the parallel refinement strategy, with $K+1 = 8$ samples being iteratively added to the DoE (Subsection 4.2). In Subsection 4.3, the same examples are then considered with a single refinement strategy, starting from an *unfavorable* DoE: it permits to investigate the capability of eAK-MCS to detect failure regions, even when the initial

surrogate does not. Finally, two more challenging 8D benchmark cases associated to a highly non-linear isoprobabilistic transformation T are studied in Subsection 4.4, underlying some limitations of the stopping criterion and the isotropic centered Gaussian used as the ISD in the exploitation step, while assessing the robustness of the surrogate refinement algorithm.

4.1. Academic Examples

4.1.1. Single Failure Region 2D

A first classic 2D example is taken from [9] [11]. This example is characterized by a low failure probability ($p_f \sim 3 \times 10^{-5}$), a very smooth limit state and a single failure region. The performance function in the standard space reads:

$$G(x_1, x_2) = \frac{1}{2}(x_1 - 2)^2 - \frac{3}{2}(x_2 - 5)^3 - 3 \tag{21}$$

where $X_1, X_2 \overset{\perp\!\!\!\perp}{\sim} \mathcal{N}(0,1)$. The probability of failure reads $p_f = \mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) < 0)$. The reference value is estimated as 2.874×10^{-5} with a CoV of 0.03% using IS with a Gaussian mixture ISD [24] based on 10^7 samples.

4.1.2. Four-branch series system 2D

This example is a variation of a classical structrural reliability test case [11, 10], where the threshold is modified to make p_f smaller. The performance function is defined as:

$$G(x_1, x_2) = \min \left\{ \begin{array}{c} 3 + \frac{(x_1 - x_2)^2}{10} - \frac{x_1 + x_2}{\sqrt{2}} \\ 3 + \frac{(x_1 - x_2)^2}{10} + \frac{x_1 + x_2}{\sqrt{2}} \\ x_1 - x_2 + \frac{7}{\sqrt{2}} \\ -(x_1 - x_2) + \frac{7}{\sqrt{2}} \end{array} \right\} \tag{22}$$

and $X_1, X_2 \overset{\perp\!\!\!\perp}{\sim} \mathcal{N}(0,1)$. The objective is to estimate $p_f = \mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) \leq u)$. For $u = -4$, the reference failure probability is $p_f \sim 5.596 \times 10^{-9}$, with a CoV of about 0.04% [25], based on 100 runs of Subset Simulation with sample size 10^7.

4.1.3. Deviation of a Cantilever Beam 2D

This example is taken from [25], where the deflection of the tip of a Cantilever Beam with a rectangular cross-section reads

$$f(y_1, y_2) = \frac{3L^4 y_1}{2E y_2^3}, \tag{23}$$

where $L = 6$, $E = 2.6 \times 10^4$, Y_1 and Y_2 are assumed independent, with $Y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $\mu_1 = 10^{-3}$, $\sigma_1 = 0.2\mu_1$, $\mu_2 = 0.3$ and $\sigma_2 = 0.1\mu_2$. The failure probability reads $p_f = \mathbb{P}_{\mathbf{Y}}(-f(\mathbf{Y}) < -\frac{L}{325})$. The reference value is $p_f \sim 3.937 \times 10^{-6}$, with a CoV of about 0.03% [25], based on 100 runs of Subset Simulation with sample size 10^7 .

4.1.4. Response of a non-linear Oscillator 6D

This example is taken from [25]. It consists of a non-linear undamped single degree of freedom system [7]. In particular, the performance function is given as follows:

$$J(\mathbf{y}) = J(c_1, c_2, m, r, t_1, F_1) = 3r - \left| \frac{2F_1}{m\omega_0^2} sin\left(\frac{\omega_0 t_1}{2}\right) \right| \tag{24}$$

with $\omega_0 = \sqrt{\frac{c_1 + c_2}{m}}$. The six random variables (assumed independent) are listed in Table 2, where the variability of F_1 is modified w.r.t. [7] in order to make the failure probability $p_f = \mathbb{P}_{\mathbf{Y}}(J(\mathbf{Y}) < 0)$ smaller. The reference value is $p_f \sim 1.514 \times 10^{-8}$, with a CoV of about 0.04% [25], based on 100 runs of Subset Simulation with sample size 10^7 .

Table 2: Response of a non-linear oscillator: Random Variables [25].

Variable[a]	P.D.F.	Mean	Standard Deviation
m	Normal	1	0.05
c_1	Normal	1	0.1
c_2	Normal	0.1	0.01
r	Normal	0.5	0.05
F_1	Normal	0.45	0.075
t_1	Normal	1	0.2

[a] Variables are independent

4.1.5. Borehole-function 8D

This example is taken from [1]. This benchmark function describes the water flow through a borehole:

$$v(\mathbf{y}) = \frac{2\pi T_u (H_u - H_l)}{\ln(\frac{r}{r_w})\big(1 + \frac{T_u}{T_l} + \frac{2LT_u}{\ln(\frac{r}{r_w}) r_w^2 K_w}\big)} \tag{25}$$

with $\mathbf{y} = (r_w, r, T_u, H_u, T_l, H_l, L, K_w)$, $v(\mathbf{y})$ is the fluid water flow measured in m^3/year, r_w is the radius of the borehole, r the radius of influence, T_u the transmissivity of the upper aquifer, H_u the potentiometric head of the upper aquifer, T_l the trasmissibity of the lower aquifer, H_l the potentiometric head of the lower quifer, L the length of the borehole and K_w the hydraulix condutivity of the soil.

The eight independent random variables are listed in Table 3. The threshold value is modified w.r.t. [1] to make the failure probability $p_f = \mathbb{P}_{\mathbf{Y}}(-v(\mathbf{Y}) < -300)$ smaller. The reference values is $p_f \sim 8.732 \times 10^{-9}$, estimated with a CoV of 0.20% using IS with a Gaussian mixture ISD [24] based on 10^7 samples.

Table 3: Borehole-functions 8D: Random Variables [1].

Variable[a]		Distribution	Parameters
r_w	[m]	Uniform	(0.05,0.15)[b]
r	[m]	Lognormal	(7.71,1.0056)[c]
T_u	[m^2/year]	Uniform	(63070,115600)[b]
H_u	[m]	Uniform	(990,1110)[b]
T_l	[m^2/year]	Uniform	(63.1,116)[b]
H_l	[m]	Uniform	(700,820)[b]
L	[m]	Uniform	(1120,1680)[b]
K_w	[m^2/year]	Uniform	(9855,12045)[b]

[a] Variables are independent.
[b] Minimum and maximum of the distribution.
[c] Mean and standard deviation of the natural logarithm of the distribution.

4.1.6. A two-degree-of-freedom damped oscillator 8D

This eight-dimensional structural reliability example was first proposed in [31] and studied in [10, 14, 15]. It is used here to demonstrate the ro-

bustness of eAK-MCS for this case characterized by a highly non-linear LSS. This example considers the reliability assessmeent of a two d.o.f. primary-secondary system under a white noise base acceleration. The performance function is defined, in the physical space, as

$$\begin{aligned} J(\mathbf{y}) &= J(m_p, m_s, k_p, k_s, \xi_p, \xi_s, F_S, S_0) \\ &= F_S - 3k_s \left(\frac{\pi S_0}{4\xi_s \omega_s^3} \frac{\xi_a \xi_s}{\xi_p \xi_s (4\xi_a^2 + \theta^2) + \kappa \xi_a^2} \frac{(\xi_p \omega_p^3 + \xi_s \omega_s^3)\omega_p}{4\xi_a \omega_a^4} \right)^{\frac{1}{2}}, \end{aligned} \tag{26}$$

with $\omega_p = \sqrt{\frac{k_p}{m_p}}$, $\omega_s = \sqrt{\frac{k_s}{m_s}}$, $\omega_a = \frac{\omega_p + \omega_s}{2}$, $\xi_a = \frac{\xi_p + \xi_s}{2}$, $\kappa = \frac{m_s}{m_p}$ and $\theta = \frac{\omega_p - \omega_s}{2}$. The eight independent random variables are listed in Table 4. The reference value reading $p_f = \mathbb{P}_{\mathbf{Y}}(J(\mathbf{Y}) < 0)$ is estimated using IS with a Gaussian mixture ISD [24] based on 10^7 samples. It results in $p_f \sim 3.781 \times 10^{-7}$ with a CoV of 0.09%, consistent with previous studies [10, 14].

Table 4: Two d.o.f. damped oscillator: Random Variables [10].

Variable[a]	P.D.F.	Mean	CoV
m_p	Lognormal	1.5	10%
m_s	Lognormal	0.01	10%
k_p	Lognormal	1	20%
k_s	Lognormal	0.01	20%
ξ_p	Lognormal	0.05	40%
ξ_s	Lognormal	0.02	50%
F_S^b	Lognormal	27.5	10%
S_0	Lognormal	100	10%

[a] Variables are independent.

4.2. Parallel Refinement Study

In this Subsection, the test cases listed in Subsubsections 4.1.1 4.1.2 4.1.3 4.1.4 considering $K+1 = 8$ samples added at each iteration of the refinement procedure are investigated.

To assess the statistical significance of the proposed method (due to its stochastic nature), each test case is studied based on 50 independent runs. The number of calls N_{calls} to the performance function, the number of iterations N_{iter}, the estimation of the failure probability $\hat{p}_f$, and the final relative

$\epsilon^{rel}_{p_f}$ error, a.k.a. relative absolute bias $\epsilon^{rel}_{p_f} = \dfrac{|\hat{p}_{\tilde{f}} - p_f^{ref}|}{p_f^{ref}}$ are consequently random variable illustrated by their empirical average. The initial DoE is supposed to be evaluated in one iteration.

Note that the CoV indicated for eAK-MCS is the one empirically estimated from the replications: the lowest it is, the more likely the method is to return an estimation $p_{\tilde{f}}$ close to its asymptotic average. Differently, the IS-based CoV (Equation 10) is an indicator of the efficiency of the surrogate exploitation step, applying the IS method on the metamodel, hence, independently of the latter.

Note that in the context of multiple independent runs, the CoV of $\hat{p}_{\tilde{f}}$ estimated from realizations of $\hat{p}_{\tilde{f}}$ (which is different from the IS-based CoV estimation (Equation 10) for a single run), is an indicator of the robustness of the method. The lowest is the CoV, the more likely the method will return an estimation $\hat{p}_{\tilde{f}}$ close to its asymptotic average. It is then dependent of the metamodel accuracy at the end of the refinement algorithm (or equivalently the DoE), and the IS-based CoV estimation (Equation 10).

The tuning parameters are the ones provided in Table 1. For each case, figures showing the average relative absolute bias $\epsilon^{rel}_{p_f}$, and the average estimate $\hat{p}_{\tilde{f}}$ as a function of the number of performance function calls are provided, where additional samples are added even after the stopping criterion is met, for the sake of illustration. As previously mentioned, the parallel strategy refinement might propose strictly less than $K+1=8$ samples. For two-dimensional examples, we have also provided an illustration of the final DoE and the refined metamodel, when the convergence criterion is satisfied, based on a single run. Whenever it is possible, results also are compared against other methods in the literature. The Four-branch series system 2D, the Cantilever Beam 2D and the non-linear Oscillator 6D are compared against BSS as reported in [25], based on 100 independent runs, for different accuracy settings (see [25] for details) and a *single* enrichment strategy. It explains the vast range of CoV and $\epsilon^{rel}_{p_f}$ for those cases estimated with BSS.

Single Failure Region 2D. In Table 5, we compare the results against the following methods: Crude MC, FORM, SORM, FORM+IS, AK-IS, MetaAK-IS^2, based on single runs. The proposed algorithm performs well w.r.t. other metamodel based methods (AK-IS, MetaAK-IS^2) with reasonable accuracy, concerning the number of performance function calls, while it outperforms

them regarding the number of iterations, as expected. An illustration is provided in Figure 3(a), showing the refined metamodel for a single run. The DoE selected at the last iteration are well clustered around the true LSS, the latter being correctly estimated by the predicted LSS, matching the upper/lower LSS. This is further illustrated in the failure probability history Figure 3(c) where the $2-\sigma$ confidence interval of $\hat{p}_{\tilde{f}}$ rapidly merges and converges to the reference value. After 4 iterations, the average absolute relative bias $\epsilon_{p_f}^{rel}$ is below 1%. The low empiral CoV $\sim 1\%$ quantifies the high robustness of the method for this test case characterized by a single MPFP and a slighly curved LSS. It explains why it is also well treated by SORM while FORM significantly overestimates p_f.

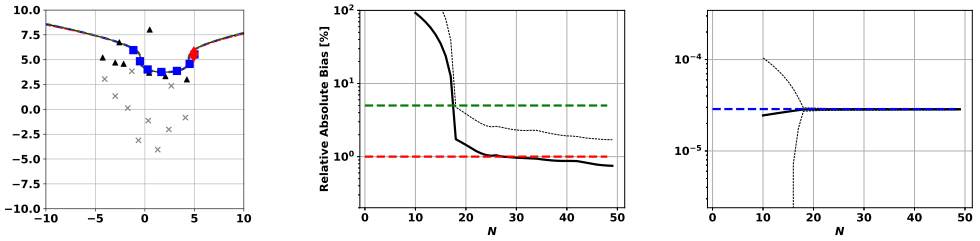
The low value of $\epsilon_{p_f}^{rel}$ can be explained by both the good predictability of the surrogate, and the efficient metamodel exploitation quantified by the IS-based CoV (below 0.40%).

Table 5: Results of the Single Failure Region 2D with the parallel strategy ($K+1=8$).

Method	N_{calls}	N_{iter}	$\hat{p}_f$	CoV	$\epsilon_{p_f}^{rel}$
Reference[a]	10^7	-	2.874×10^{-5}	0.03%	-
FORM[b]	19	-	4.21×10^{-5}	-	98.5%
SORM[c]	19+5	-	2.83×10^{-5}	-	1.64%
FORM + IS[b]	$19 + 10^4$	$10 + 10^4$	2.86×10^{-5}	2.39%	0.48%
AK-IS[b]	26	17	2.86×10^{-5}	2.39%	0.48%
MetaAK-IS^2 [b]	28	19	2.87×10^{-5}	2.39%	0.14%
eAK-MCS[c]	26.5	3.1	2.851×10^{-5}	1.02%	1.01%

[a] IS with a gaussian mixture as ISD [24]
[b] Reproduced from [11], single run, single refinement strategy.
[c] Result obtained assuming the MPFP is obtained from FORM [11], with $N_{calls} = N_{FORM} + N_{SORM}$.
[d] Initial DoE size: 10. K+1=8 samples iteratively added. Based on 50 independent runs.

Four-branch series system 2D. This two-dimensional example characterized by four failure domains is subject to a very low failure probability ($\sim 5.6 \times 10^{-9}$). The results are presented in Table 6 and Figure 4. eAK-MCS stops after ~ 62 calls (~ 8 iterations) on average, while BSS uses between 50 and 80 calls. The empirical CoV is small, 1.57%, and the accuracy satisfactory,



(a) Metamodel (b) Relative Absolute Bias History (c) $\hat{p}_f$ History

Figure 3: Results of the Single Failure Region 2D with the parallel strategy ($K+1=8$). (a) Legend Settings in Figure 1. (b) Green and red thick dashed lines indicate respectively 5% and 1% relative error. The y-axis is logarithmic. The average relative absolute bias and the associated $2-\sigma$ confidence intervall are represented respectively in black thick line and black thin dashed lines. (c) The average *predicted* failure probability $\hat{p}_{\tilde{f}}$ and the associated $2-\sigma$ confidence intervall are represented respectively in thick plain and thin dashed black lines. (b-c) based on 50 runs.

with an average relative error of 1.20%. IS is efficient for that case, the IS-based average CoV being lower than 0.60%.

As seen in Figure 4 (a), the true LSS is well estimated by the predicted LSS in the region characterized by high density of input distribution, where the ability of the surrogate to classify samples into the safe/unsafe domain is the most sensitive. Note also that in that zone, the upper/lower predicted LSS match the predicted LSS, indicating high predictability of the surrogate and the presence of DoE clustered in that zone. Figures 4 (b) (c) illustrate respectively the average relative absolute bias and the failure probability history as a function of the number of performance calls. One can note that after, 26 function calls (3 iterations), the average predicted failure probability $\hat{p}_{\tilde{f}}$ is already of the same order of magnitude of the reference value, and the convergence is then rather fast, with an average absolute relative bias lower that 1% after $\sim$ 70 function calls ($\sim$8 iterations).

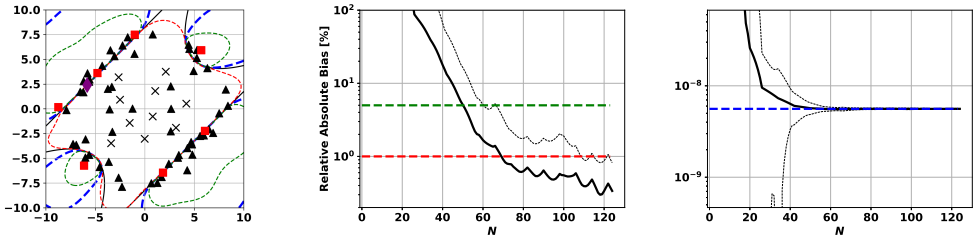
As expected, FORM/SORM present large biases due to the presence of multiple MPFPs of similar Hasofer-Lind reliability indexes. Moreover, they provided the same p_f estimation, since the MPFP found is associated to a linear LSS.

Deviation of a Cantilever Beam 2D. This two-dimensional example is characterized by a single failure region, with $p_f \sim 4 \times 10^{-6}$. The results are

Table 6: Results of the Four-branch series system 2D with the parallel strategy ($K+1=8$).

Method	N_{calls}	N_{iter}	$\hat{p}_f$	CoV	$\epsilon_{p_f}^{rel}$
Reference[a]	10^7	-	5.596×10^{-9}	0.04%	-
BSS [25]	$\sim 50-80$	$\sim 41-71$	-	$\sim 0.5-0.01$	$\sim 0.01-5\%$
FORM[b]	??	??	2.79×10^{-9}	-	50.0%
SORM[b]	??+5	$?? + 1$	2.79×10^{-9}	-	50.0%
eAK-MCS[c]	61.9	7.5	5.579×10^{-9}	1.57%	1.20%

[a] Based on 100 independent runs, for different accuracy settings and a single refinement strategy. [25].
[b] $N_{calls} = N_{FORM} + N_{SORM}$.
[c] Initial DoE size: 10. K+1=8 samples iteratively added. Based on 50 independent runs.



(a) Metamodel (b) Relative Absolute Bias History. (c) $\hat{p}_f$ History

Figure 4: Results of the Four-branch series system 2D with the parallel strategy ($K+1=8$). Legend settings in Figure 3.

presented in Table 7 and Figure 5. The eAK-MCS algorithm stops after $\sim$ 41 calls in average, significantly larger than for BSS ($\sim$ 25), but still with fewer iterations ($\sim$ 5 against $\sim$ 15). Figure 5 (b) shows that the average relative absolute bias is below 1% after around 48 calls. For the sake of illustration, we have truncated the history to 50 calls. Indeed, for few runs characterised by very accurate metamodels, the enrichment algorithm proposes candidates that are too close to the existing DoE to be accepted, and the algorithm stops.

The empirical CoV is small, 2% and the accuracy is satisfactory, with an average relative error lower of 1.21%. IS performs well, with a average IS-based CoV lower than 0.5%. It can be noticed that the surrogate detects spurious predicted LSS during the refinement step, which explains the additional computational burden. For the single run illustrated Figure 5 (a), a

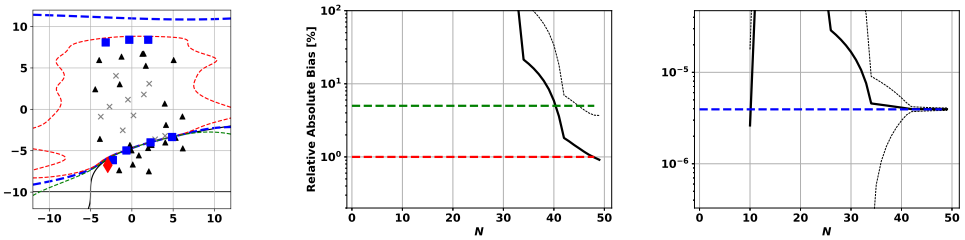
second artificial failure region is finally so far from the origin (center of the standard space distribution) that it has no impact on the estimated failure probability.

Both FORM and SORM perform well, due to the presence of a single MPFP and an almost linear LSS.

Table 7: Results of the Deviation of a Cantilever Beam 2D with the parallel strategy ($K+1=8$).

Method	N_{calls}	N_{iter}	$\hat{p}_f$	CoV	$\epsilon_{p_f}^{rel}$
Reference[a]	10^7	-	3.937×10^{-6}	0.03%	-
BSS [25]	$\sim 22-25$	$\sim 13-16$	-	$\sim 0.5-0.01$	$\sim 0.1-5\%$
FORM[b]	??	??	4.19×10^{-6}	-	6.49%
SORM[b]	??+5	?? + 1	3.88×10^{-6}	-	1.26%
eAK-MCS[b]	41.2	4.9	3.949×10^{-6}	2.07%	1.21%

[a] Based on 100 independent runs, for different accuracy settings and a single refinement strategy. [25].
[b] $N_{calls} = N_{FORM} + N_{SORM}$.
[c] Initial DoE size: 10. K+1=8 samples iteratively added. Based on 50 independent runs.



(a) Metamodel (b) Relative Absolute Bias History. (c) $\hat{p}_f$ History

Figure 5: Results of the Deviation of a Cantilever Beam 2D with the parallel strategy ($K+1=8$). Legend settings in Figure 3.

Response of a Nonlinear Oscillator 6D. The results for this six-dimensional example characterized by $p_f \sim 1.5 \times 10^{-8}$ are presented in Table 8 and Figure 6. They are compared against FORM/SORM, BSS [25] and AK-MCSi [8]. eAK-MCS requires on average ~ 45 calls, significantly less than AK-MCSi, similarly to BSS for the less costly settings, but with a meager number of iterations (~ 3). Note however that for the most expensive ones,

BSS would require in average 180 calls. The robustness and accuracy for this higher dimensional test case are acceptable (empirical CoV $\sim 8.3\%$, average relative error $\sim 8.9\%$) compared to previous two-dimensional test-cases but still satisfactory considering the low computational cost.

This increased inaccuracy stems mainly from the *average* efficiency of the IS (with the ISD $\mathcal{N}(0, \gamma^2 I_d)$) to reduce the variance in this higher dimensional case. Indeed, the CoV of the IS estimation of $\hat{p}_{\tilde{f}}$ is in average ~ 2.25 % (while below 0.5% for the two-dimensional cases), and is entirely independent of the quality of the surrogate (or the DoE). The second reason is the choice of the stopping criterion (*Fast*, Eq. 20) which is likely to stop prematurely, especially here with significant IS confidence intervals.

Figure 6(a)show that the average relative error reaches around 2% after ~ 75 calls and then oscillates to stay between $\sim 1.5\%$ and 2.5%, confirming respectively that the algorithm stopped prematurely, and that the IS estimation is not accurate enough to provide an average relative error lower than 1%. The quality of the refinement algorithm itself (or equivalently, of the surrogate) is further confirmed in Figure 6 (b) indicating that the reference value is however contained the noise associated to the IS-based CoV.

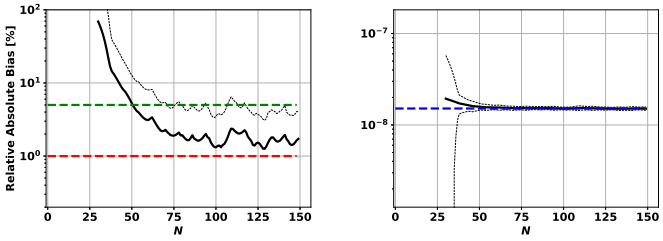
Table 8: Results of Response of a Nonlinear Oscillator 6D with the parallel strategy ($K+1=8$).

Method	N_{calls}	N_{iter}	$\hat{p}_f$	CoV	$\epsilon_{p_f}^{rel}$
Reference[a]	10^7	-	1.514×10^{-8}	0.04%	-
BSS [25]	$\sim 45-180$	$\sim 36-171$	-	$\sim 0.5-0.01\%$	$\sim 0.01-10\%$
AK-MCSi[b] [8]	77	68	1.44×10^{-8}	$< 5\%$	3.4%
FORM[c]	??	??	1.56×10^{-8}	-	3.14%
SORM[c]	??+27	$?? + 1$	8.44×10^{-9}	-	44.28%
eAK-MCS[d]	44.7	2.8	1.633×10^{-8}	8.35%	8.91%

[a] Based on 100 independent runs, for different accuracy settings and a single refinement strategy. [25].
[b] Single refinement strategy [8].
[b] $N_{calls} = N_{FORM} + N_{SORM}$.
[d] Initial DoE size: 10. K+1=8 samples iteratively added. Based on 50 independent runs.

4.3. Sequential Study: Failure regions detection

In this Subsection, numerical experiments are conducted with an *unfavorable* initial DoE, namely *too* focused on the center of the standard space, corresponding to a large ϵ: $\epsilon = 10^{-2}$, against $\epsilon = 10^{-5}$ in default numerical



(a) Relative Absolute Bias History

(b) $\hat{p}_f$ History

Figure 6: Results of Response of a Nonlinear Oscillator 6D with the parallel strategy ($K+1=8$). Legend settings in Figure 3.

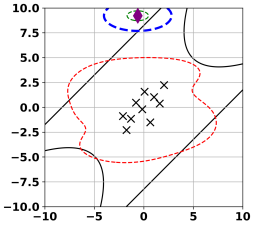
settings; in order to provide some hints about the ability of eAK-MCS to detect failure regions, even if the surrogate does not.

eAKMCS is used without parallel refinement, with a single run, adressing the test cases introduced in Subsubsections 4.1.2 4.1.3 4.1.4.

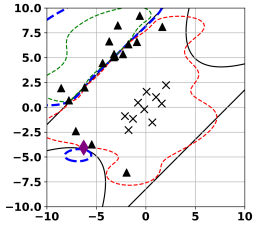
For each case, figures showing the relative absolute bias $\epsilon_{p_f}^{rel}$, and the estimates $\hat{p}_{\tilde{f}}^{(\tilde{k})+}$, $\hat{p}_{\tilde{f}}^{(\tilde{k})-}$ and $\hat{p}_{\tilde{f}}$ as a function of the number of performance function calls are provided. For two-dimensional examples, an illustration of the metamodel refinement is provided.

Four-branch series system 2D. Figures 7 (a), (b), (c) and (d) show that all the four failure domains are detected after respectively 11, 28, 56 and 100 calls of the original model. It illustrates a *behaviour* discussed in the original AK-MCS version [7], namely that the single refinement strategy tends to focus on each failure branch individually then *explore* other branches when sufficiently refined. Even though the initial DoE is *unfavorable*, all failure domains are successfully detected.

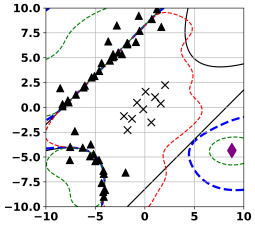
Deviation of a Cantilever Beam 2D. Figure 8 (a) shows that the failure domain is detected after 14 performance function evaluations. The final metamodel is obtained iteratively with a total of 23 function calls, the same order of magnitude required for BSS. Note the unexpected convergence of the failure probability history, from whose we can notice that the reference value is globally not contained within the estimated failure probability range: it highlights a lack of predictability of the selected surrogate, whose features could be modified accordingly, while validating the robustness of the method.



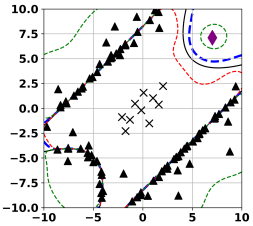
(a) First Failure Zone: $N = 11$.



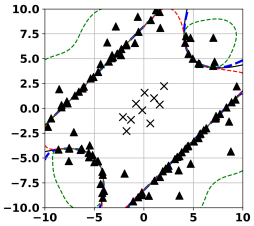
(b) Second Failure Zone: $N = 28$.



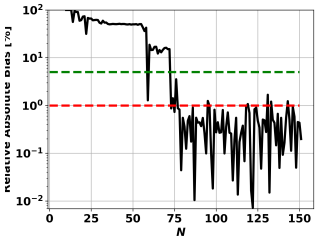
(c) Third Failure Zone: $N = 56$.



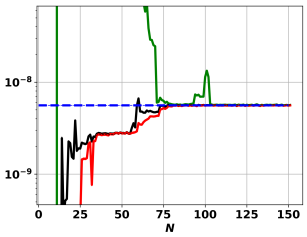
(d) Fourth Failure Zone: $N = 100$.



(e) Final: $N = 112$.



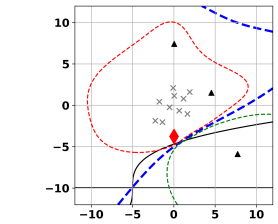
(f) Relative Absolute Bias History.



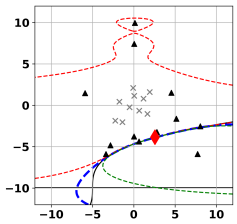
(g) $\hat{p}_f$ History

Figure 7: Results of the Four-branch series system 2D with the single refinement strategy, based on a single run and *unfavorable* initial DoE ($\epsilon = 10^{-2}$). Legend settings in Figure 3 (a-f). (g) The *predicted* failure probability $\hat{p}_{\tilde{f}}$, its lower and upper bounds $\hat{p}_{\tilde{f}}^{(\tilde{k})-}$, $\hat{p}_{\tilde{f}}^{(\tilde{k})+}$ are indicated respectively in black, red and green thick lines. Associated 2-σ)-IS confidence intervals are indicated in thin dashed lines, however not visible for small CoV.

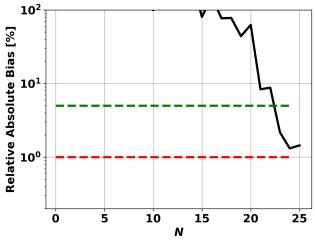
Response of a Nonlinear Oscillator 6D. The Absolute Relative Bias history represented in Figure 9 (a) shows that the relative error is below 5% after around 35-50 performance calls, corresponding to around 5-20 refinement



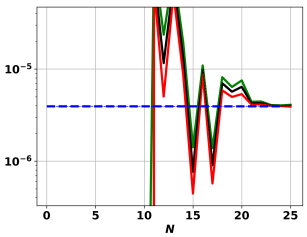
(a) First Failure Zone: $N = 14$.



(b) Final: $N = 23$.



(c) Relative Absolute Bias History.



(d) $\hat{p}_f$ History

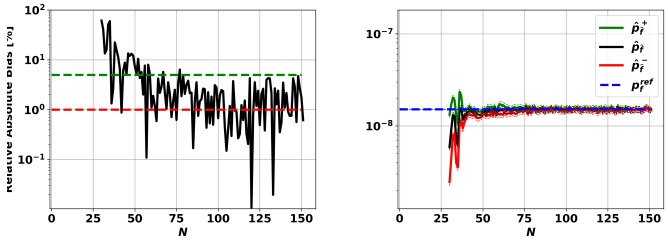
Figure 8: Results of the Deviation of a Cantilever Beam 2D with the single refinement strategy, based on a single run and *unfavorable* initial DoE ($\epsilon = 10^{-2}$). Legend settings in Figure 7.

steps. Note the oscillatory behaviour of the relative absolute bias. Those oscillations stem from the significant IS-based CoV, as mentioned in the previous Subsection. Again, Figure 9(b) shows that the reference value lies within the confidence interval of $\hat{p}_{\tilde{f}}$.

It illustrates the capability of eAK-MCS to detect failure regions even under *unfavorable* initial DoE.

4.4. eAK-MCS limits

In this Subsection, the two main weaknesses of eAK-MCS as presented in this article are discussed through its application to the 8D cases (Subsubsections 4.1.5 4.1.6) involving a non-linear isoprobabilistic transform T. The robustness of the eAK-MCS refinement algorithm, main contribution of the present study, is however assessed.



(a) Relative Absolute Bias History.

(b) $\hat{p}_f$ History

Figure 9: Results of Response of a Nonlinear Oscillator 6D with the single refinement strategy, based on a single run and *unfavorable* initial DoE ($\epsilon = 10^{-2}$). Legend settings in Figure 7.

Borehole-function 8D. This eight-dimensional case characterized by a very small failure probability ($p_f \sim 9 \times 10^{-9}$) is considered in the following: it is based on a single run and a parallel strategy adding $K + 1 = 8$ samples at each iteration, starting from an initial DoE of size 40 and a maximal DoE size of 200 samples.

The relative absolute bias and predicted failure probability histories are presented in Figure 10, respectively. Conversely to the lower dimensional cases studied in the two previous Subsections and characterized by linear iso-probabilistic transformations, we observe in Figure 10(b) that the IS-based CoV is very large (40%-50%). It reveals the inadequacy of the centered isotropic ISD to accurately exploit the surrogate by IS (see [32]). Additionally, we observe that the reference value is contained within the 2-σ IS confidence interval, after around 45 LSF evaluations. It explains the few peaks of very low relative absolute biases (Figure 10(a)), corresponding to favorable realizations of the predictive failure probability $p_{\tilde{f}}$. We note however, a very oscillatory relative absolute bias, mostly above 10%. This IS ineffiency the oscillatory nature of the predicted failure probabilities (Figure 10).

In order to partially adress this issue, we resort to a Gaussian mixture ISD [24] in place of the isotropic centered gaussian distribution in the IS-based surrogate exploitation step, for the very same DoE, represented in Figure 11. The IS-based CoV is greatly reduced ($\sim 0.30\% - 0.40\%$), showcasing the suitability of this IS, as observed in Figure 11(b). The latters highlights that, rapidly, the obtained predicted failure probabilities are of the same order of

magnitude as the reference value. This is confirmed in Figure 11(a), which shows that the relative error bias is most of the time below 10%, and below 5% after around 120 LSF calls. Besides few peaks, Figure 11(b) depicts a fairly stable predicted failure probability, close to the reference value, while the bounds $\hat{p}_{\tilde{f},min}$ and $\hat{p}_{\tilde{f},max}$ do not clearly converge towards $p_{\tilde{f}}$.

This test case suggests that in higher dimension, IS based on the isotropic centered Gaussian ISD is not adapted [32]; hence, the surrogate exploitation step should be performed by a more accurate technique, such as IS using a more suitable ISD [24], SS... Moreover, it also shows limit of the stopping criterion introduced in Subsection 3.5, depending mainly on the *predictive* accuracy of the GP-based surrogate. Indeed, in that case the algorithm would stop either prematurily or until the LSF budget is exhausted, depending on the ISD choice.

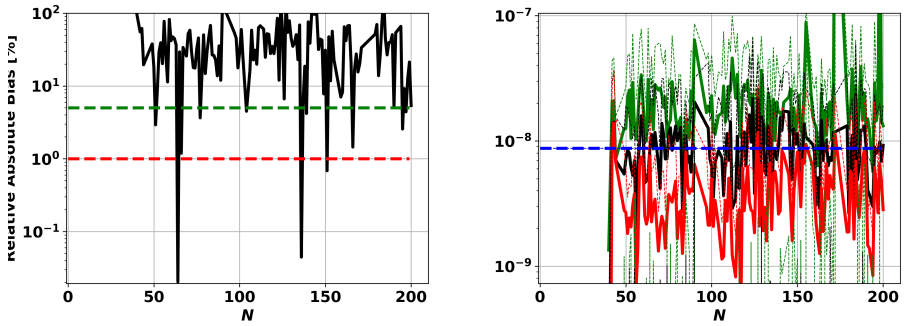
The Table 9 summarizes the results obtained, and compared against FORM/SORM. FORM totally overestimates p_f while SORM provides an acceptable result associated to a bias of 13.9%. eAK-MCS outperforms both, with a relative absolute bias below 3%. It is worth mentioning that an adapted stopping criterion would have permitted to stop earlier the algorithm.

Table 9: Results of Borehole-function 8D with the parallel strategy ($K+1=8$).

Method	N_{calls}	N_{iter}	$\hat{p}_f$	CoV	$\epsilon_{p_f}^{rel}$
Reference[a]	10^7	-	8.732×10^{-9}	0.20%	-
FORM[b]	??	??	3.11×10^{-7}	-	3462%
SORM[b]	??+44	$?? + 1$	7.51×10^{-9}	-	13.9%
eAK-MCS[c]	199	21	8.978×10^{-9}	0.30%	2.81%

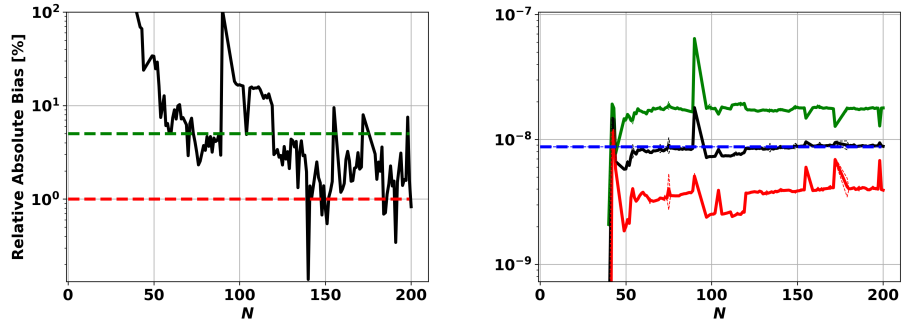
[a] IS with a gaussian mixture as ISD [24]
[b] $N_{calls} = N_{FORM} + N_{SORM}$.
[c] Initial DoE size: 40. K+1=8 samples iteratively added. Based on a single run. Algorithm stopped when the DoE size is above 200.

A two-degree-of-freedom damped oscillator 8D. This eight-dimensional example studied in [10, 14, 15] is compared against FORM, SORM, [2]SMART, Meta-IS and ASVR. eAK-MCS sharing more similarities with Meta-IS, the case is investigated with an initial DoE size of 32, adding $K + 1 = 16$ samples at each iteration, based on a single run, the LSF budget is fixed to 600. Convergence histories are reported in Figure 12.



(a) Relative Absolute Bias History. (b) $\hat{p}_f$ History

Figure 10: Results of the Borehole-Function with the parallel strategy ($K+1=8$), based on a single run. Legend settings in Figure 7.

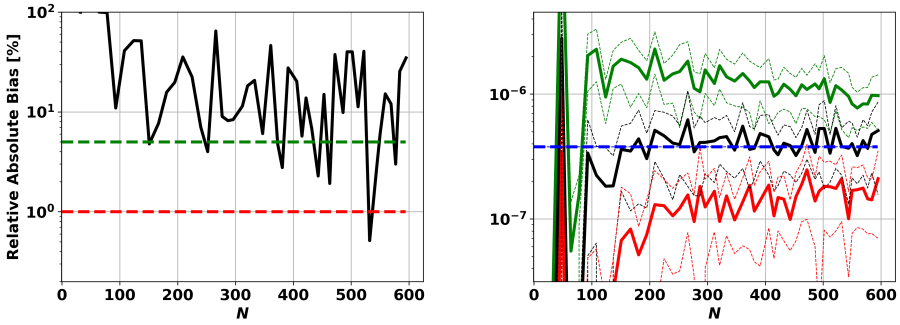


(a) Relative Absolute Bias History. (b) $\hat{p}_f$ History

Figure 11: Results of the Borehole Function ($p_f = \mathbb{P}_{\mathbf{Y}}(-v(\mathbf{Y}) < -290) \sim 5.795 \times 10^{-7}$) with the parallel strategy ($K+1=8$), based on a single run. Exploitation step performed with Gaussian Mixture ISD [24]. Legend settings in Figure 7.

Similar difficulties to the previous case are encountered: the stopping criterion do not seem adapted, while IS associated to the isotropic centered Gaussian ISD lacks of efficiency (IS-based CoV $\sim$ 20%. Again, a Gaussian mixture ISD [24] is used to exploit the surrogate, considering the very same DoE (Figure 13). We observe that after around 200 LSF calls, the predicted failure probability is rather close to the reference vallue, slighly overestimating it, and presenting some oscillations. Predicted failure probability bounds very slowly converge to $p_{\tilde{f}}$. Results are summarized in Table 10. The single

failure domain having the shape of *a needle pointing towards the origin* [23], explains why FORM overestimates p_f while SORM provides a fairly good estimation. In terms of accuracy and number of iterations, Meta-IS outperforms eAK-MCS, ASVR providing also a more accurate result with the same order of magnitude of LSF calls. Though, it is worth mentioning that a more suitable stopping criterion, or expert judgement (which is meaningful in the context of reliability analysis) would very likely lead to stop the algorithm between 200 and 300 LSF calls. In that case, eAK-MCS seems less accurate than other method, while it seems to rapidly provide a failure probability reasonably estimated. It could also be explained by the lowest quality of the surrogate w.r.t. to the ones of Meta-IS or ASVR.



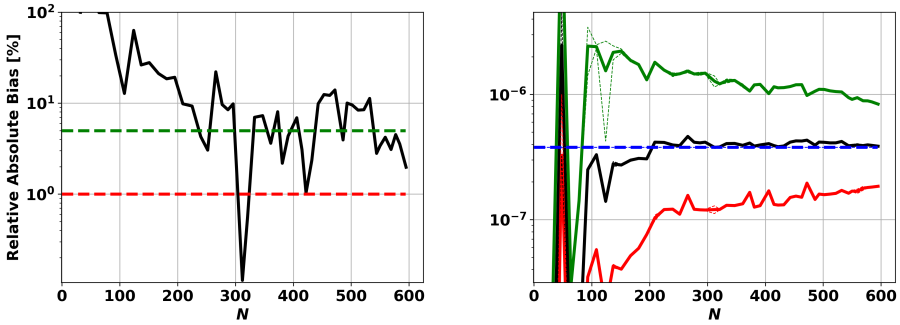
(a) Relative Absolute Bias History.

(b) $\hat{p}_f$ History

Figure 12: Results of the two d.o.f. Oscillator 8D with the parallel strategy ($K+1=16$), based on a single run. Legend settings in Figure 7.

5. Conclusion

This paper proposes an extension of AK-MCS as presented in [1] to make it suitable for very low failure probabilities. It uses a centered uncorrelated gaussian distribution to sample candidate points and use the IS method to estimate p_f. A procedure is proposed to tune its standard deviation adaptively. Moreover, the original multipoint refinement strategy inherited from [1] is slightly modified, enabling the use of available high-performance computing resources. The performance of the proposed algorithm is assessed and illustrated through some benchmark analytical functions, showcasing



(a) Relative Absolute Bias History. (b) $\hat{p}_f$ History

Figure 13: Results of the two d.o.f. Oscillator 8D with the parallel strategy ($K+1=16$), based on a single run. Exploitation step performed with Gaussian Mixture ISD [24]. Legend settings in Figure 7.

Table 10: Results of the two d.o.f. oscillator 8D with the parallel strategy ($K+1=16$), based on a single run.

Method	N_{calls}	N_{iter}	$\hat{p}_f$	CoV	$\epsilon_{p_f}^{rel}$
Reference[a]	10^7	-	3.781×10^{-7}	0.09%	0
FORM[b]	2727	303	3.91×10^{-6}	-	934 %
SORM[b]	2727+44	303+1	3.70×10^{-7}	-	2.14%
^{2}SMART[c]	4011	-	3.66×10^{-7}	9.6%	3.20%
Meta-IS^2[,c]	480+200	28+1	3.76×10^{-7}	$< 5\%$	0.55%
ASVR[d]	648	-	3.81×10^{-7}	1.6%	0.8%
eAK-MCS[e]	595	49	3.870×10^{-7}	0.34%	2.35%

[a] IS with a gaussian mixture as ISD [24].
[b] Reproduced from [14]. ^{2}SMART performed with 50 independent runs, the CoV being empirically estimated from the replications.
[c] Reproduced from [10]. $N_{calls} = m + N_{corr}$, m being the DoE size used to build the Kriging surrogate, N_{corr} the number of calls required to estimate the correction factor. Initial DoE size: 32. 16 samples iteratively added. Single run.
[d] Reproduced from [15], based on 20 independent runs. CoV empirically estimated from the replications.
[e] Initial DoE size: 32. K+1=16 samples iteratively added. Single run. Algorithm stops when the DoE size exceeds 600.

very satisfactory performances. Unlike AK-MCS, eAK-MCS requires to recast the problem in standard space. eAK-MCS, like most surrogate-based procedures, is a victim of the curse of dimensionality, and its efficiency is closely linked to the surrogate's ability to fit the original model. eAK-MCS is globally robust to the choice of the tuning parameters (Table 1), which provide excellent results for a broad range of cases, including the ones presented in this work.

eAK-MCS is straightforward to implement, with a clear parallelization strategy, and the possibility to easily monitor bounds of the predicted failure probability at each iteration.

The two eight dimensional examples characterized by a non-linear iso-probabilistic transformation T outlined two limitations of eAK-MCS as presented in this study. The first one is related to the stopping criterion depending only on the predicted accuracy of the GP-based surrogate, through the estimated bounds of the failure probability; In the context of reliability assessment, it has finally not an enormous impact since the user can stop at anytime the refinement, modify some parameters and re-run the procedure without losing any information. The second concerns the inadaptability of the isotropic centered Gaussian ISD for the surrogate exploitation step. As mentioned previously, it could be replaced by any suitable sampling strategy, as illustrated with the use of a Gaussian mixture ISD [24]. The surrogate refinement strategy is however assessed for those two challenging cases.

Studies with *unfavorable* initial DoE shows the ability of eAK-MCS to detect multiple failure domains, even if, unlike BSS, it aims at refining directly the *Limit-State Surface* $\{G = u\}$. This latter feature could be exploited in AK-MCS based derived methods such as quantile estimation [1] or in (quantile-based) optimization under uncertainty [16] where the refinement algorithm for a LSS of the form $\{G = u\}$ is part of more extensive procedure, the associated failure probability of this LSS being relatively high though (say $> 10^{-4}$). It makes eAK-MCS particularly attractive, so the adaptation of the mentioned algorithms for LSSs associated with small failure probability could be facilitated. The adaptation of the algorithm [1] for very small quantile estimation using eAK-MCS is indeed under progress.

References

[1] R Schöbi, Bruno Sudret, and Stefano Marelli. Rare event estimation using polynomial-chaos kriging. *ASCE-ASME Journal of Risk*

and Uncertainty in Engineering Systems, Part A: Civil Engineering, 3(2):D4016002, 2016.

[2] Jean-Marc Bourinet. *Reliability analysis and optimal design under uncertainty-Focus on adaptive surrogate-base d approaches*. PhD thesis, Université Clermont Auvergne, 2018.

[3] Armen Der Kiureghian and Taleen Dakessian. Multiple design points in first and second-order reliability. *Structural Safety*, 20(1):37–49, 1998.

[4] R. Y. Rubinstein. *Simulation and the Monte-Carlo method*. New York: Wiley, 1981.

[5] G. S. Fishman. *Monte-Carlo: concepts, algorithms and applications*. New York: Wiley, 1996.

[6] S. K. Au and J. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Prob. Eng. Mech.*, 16:263–277, 2001.

[7] B. Echard, N. Gayton, and M. Lemaire. Ak-mcs: An active learning reliability method combining kriging and monte carlo simulation. *Reliability Engineering and System Safety*, 33:145–154, 2011.

[8] Nicolas Lelièvre, Pierre Beaurepaire, Cécile Mattrand, and Nicolas Gayton. Ak-mcsi: A kriging-based method to deal with small failure probabilities and time-consuming models. *Structural Safety*, 73:1–11, 2018.

[9] B. Echard, N. Gayton, M. Lemaire, and N. Relun. A combined inportance sampling and kriging reliability method for small failure probabilities with time-demanding numerical models. *Reliability Engineering and System Safety*, 111:232–240, 2013.

[10] V. Dubourg, B. Sudret, and F. Deheeger. Metamodel-based importance sampling for structural reliability analysis. *Probabilistic Engineering Mechanics*, 33:47–57, 2013.

[11] F. Cadini, F. Santos, and E. Zio. An improved adaptive kriging-based importance technique for sampling multiple failure regions of low probability. *Reliability Engineering and System Safety*, 131:109–117, 2014.

[12] H. Zhao, Z. Yue, Y. Liu, Z. Gao, and Y. Zhang. An efficient reliability method combining adaptive importance sampling and kriging metamodel. *Applied Mathematical Modelling*, 39:1853–66, 2015.

[13] X. Huang, J. Chen, and H. Zhu. Assessing small failure probabilities by akss: An active learning method combining kriging and subset simulation. *Structural Safety*, 59:86–95, 2016.

[14] J-M Bourinet, François Deheeger, and Maurice Lemaire. Assessing small failure probabilities by combined subset simulation and support vector machines. *Structural Safety*, 33(6):343–353, 2011.

[15] J-M Bourinet. Rare-event probability estimation with adaptive support vector regression surrogates. *Reliability Engineering & System Safety*, 150:210–221, 2016.

[16] Maliki Moustapha, Bruno Sudret, Jean-Marc Bourinet, and Benoît Guillaume. Quantile-based optimization under uncertainties using adaptive kriging surrogate models. *Structural and multidisciplinary optimization*, 54(6):1403–1421, 2016.

[17] Régis Lebrun and Anne Dutfoy. A generalization of the nataf transformation to distributions with elliptical copula. *Probabilistic Engineering Mechanics*, 24(2):172–178, 2009.

[18] Regis Lebrun and Anne Dutfoy. An innovating analysis of the nataf transformation from the copula viewpoint. *Probabilistic Engineering Mechanics*, 24(3):312–320, 2009.

[19] Maurice Lemaire. *Structural reliability*. John Wiley & Sons, 2013.

[20] T. Santner, B. Williams, and W. Notz. *The design and analysis of computer experiments*. New York: Springer, 2013.

[21] C. E. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.

[22] GPy. GPy: A gaussian process framework in python. http://github.com/SheffieldML/GPy, since 2012.

[23] V. Dubourg. *Adaptive surrogate models for reliability analysis and reliability-based design optimization*. PhD Thesis, 2011.

[24] Nassim Razaaly and Pietro Marco Congedo. Novel algorithm using active metamodel learning and importance sampling: application to multiple failure regions of low probability. *Journal of Computational Physics*, 368:92–114, 2018.

[25] Julien Bect, Ling Li, and Emmanuel Vazquez. Bayesian subset simulation. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):762–786, 2017.

[26] Dirk P Kroese, Thomas Taimre, and Zdravko I Botev. *Handbook of monte carlo methods*, volume 706. John Wiley & Sons, 2013.

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

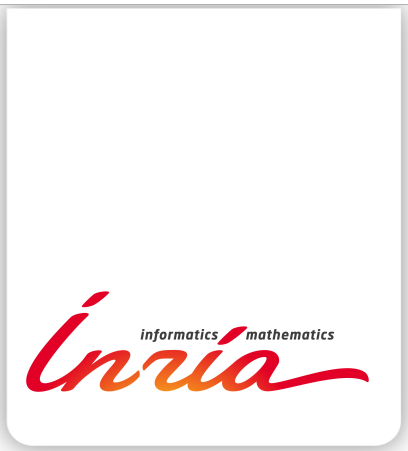
[28] F. Cadini, A. Gioletta, and E. Zio. Improved metamodel-based importance sampling for the performance assessment of radiactive waste repositories. *Reliability Engineering and System Safety*, 134:188–197, 2015.

[29] Mohammed J Zaki, Wagner Meira Jr, and Wagner Meira. *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press, 2014.

[30] Michael Hohenbichler and Ruediger Rackwitz. Improvement of second-order reliability estimates by importance sampling. *Journal of Engineering Mechanics*, 114(12):2195–2199, 1988.

[31] Armen Der Kiureghian and Mario De Stefano. Efficient algorithm for second-order reliability analysis. *Journal of engineering mechanics*, 117(12):2904–2923, 1991.

[32] Siu-Kui Au and JL Beck. Important sampling in high dimensions. *Structural safety*, 25(2):139–163, 2003.



RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

1 rue Honoré d'Estienne d'Orves

Bâtiment Alan Turing

Campus de l'École Polytechnique

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399