

RESEARCH ARTICLE

Efficient estimation of extreme quantiles using adaptive kriging and importance sampling

Nassim Razaaly¹ | Daan Crommelin^{2,3} | Pietro Marco Congedo¹

¹DeFI Team, CMAP - Centre de Mathématiques Appliquées - Ecole Polytechnique, Inria Saclay - Ile de France, Palaiseau, France

²CWI Amsterdam, Amsterdam, the Netherlands

³Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, the Netherlands

Correspondence

Nassim Razaaly, DeFI Team, CMAP - Centre de Mathématiques Appliquées - Ecole Polytechnique, Inria Saclay - Ile de France, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France.

Email: nassim.razaaly@inria.fr

SUMMARY

This study considers an efficient method for the estimation of quantiles associated to very small levels of probability (up to $O(10^{-9})$), where the scalar performance function J is complex (eg, output of an expensive-to-run finite element model), under a probability measure that can be recast as a multivariate standard Gaussian law using an isoprobabilistic transformation. A surrogate-based approach (Gaussian Processes) combined with adaptive experimental designs allows to iteratively increase the accuracy of the surrogate while keeping the overall number of J evaluations low. Direct use of Monte-Carlo simulation even on the surrogate model being too expensive, the key idea consists in using an importance sampling method based on an isotropic-centered Gaussian with large standard deviation permitting a cheap estimation of small quantiles based on the surrogate model. Similar to AK-MCS as presented in the work of Schöbi et al., (2016), the surrogate is adaptively refined using a parallel infill criterion of an algorithm suitable for very small failure probability estimation. Additionally, a multi-quantile selection approach is developed, allowing to further exploit high-performance computing architectures. We illustrate the performances of the proposed method on several two to eight-dimensional cases. Accurate results are obtained with less than 100 evaluations of J on the considered benchmark cases.

KEYWORDS

extreme quantile, importance sampling, kriging, multiple failure regions, quantile, rare event, tail probability

1 | INTRODUCTION

Quantile estimation is of fundamental importance in statistics as well as in design applications,¹ the main challenge being the number of observations required when these correspond to the output of expensive numerical simulations. Variance reduction techniques such as Importance Sampling (IS),² correlation-induction,³ and control variates^{4,5} have been proposed. The number of required observations is however large, especially for small quantiles.

Let us introduce a probabilistic model, considering a physical d -dimensional random vector \mathbf{Y} fully described by its probability density function (PDF) $f_{\mathbf{Y}}$, and a performance function $J : \mathbb{R}^d \rightarrow \mathbb{R}$ representing the system response. The problem of interest consists in the estimation of the quantile $q \in \mathbb{R}$ associated to a level of probability $\alpha \in]0, 1[$, with $\alpha \sim 0$:

$$\alpha = \mathbb{P}_{\mathbf{Y}}(J(\mathbf{Y}) < q), \quad (1)$$

$\mathbb{P}_{\mathbf{Y}}$ denoting the probability measure induced by \mathbf{Y} . It is assumed possible to recast the problem (1) using an isoprobabilistic transformation* T (eg, Rosenblatt/Nataf transform⁶⁻⁸), used to define the standard random vector \mathbf{X} and the performance function $G : \mathbb{R}^d \rightarrow \mathbb{R}$ in the standard space as

$$\mathbf{X} = T(\mathbf{Y}), \quad (2)$$

$$G(\mathbf{X}) = J(T^{-1}(\mathbf{X})). \quad (3)$$

We recall that $\mathbf{X} \sim \mathcal{N}(0, I_d)$ is the standard normal random vector of \mathbb{R}^d , described by its PDF $f_{\mathcal{N}(0, I_d)}$. Equation (1) consequently reads:

$$\alpha = \mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) < q) = \mathbb{E}_{\mathbf{X}}[\mathbb{1}_{G < q}(\mathbf{X})] = \int_{\mathbb{R}^d} \mathbb{1}_{G < q}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad (4)$$

with $\mathbb{P}_{\mathbf{X}}$ and $\mathbb{E}_{\mathbf{X}}$ denoting, respectively, the probability measure and the expectation operator induced by \mathbf{X} . $\mathbb{1}_{G < q}$ is the indicator function, equal to 1 if $G < q$, 0 otherwise. The problem of quantile estimation is closely linked to the one of estimating a failure probability α , where the model response G is associated to the critical value q . A typical approach to the latter estimation (Equation (4)) consists in resorting to a direct Monte-Carlo (MC) scheme. Its estimator reads

$$\hat{\alpha} = \mathbb{E}_{\mathbf{X}}[\mathbb{1}_{G < q}(\mathbf{X})] = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \mathbb{1}_{G < q}(\mathbf{x}_i), \quad (5)$$

where $(\mathbf{x}_1, \dots, \mathbf{x}_{N_{MC}}) \stackrel{\text{iid}}{\sim} \mathbf{X}$. This estimator is asymptotically unbiased and convergent. Its variance estimator reads:

$$\hat{\sigma}_{\hat{\alpha}}^2 = \frac{\hat{\alpha} - \hat{\alpha}^2}{N_{MC}}. \quad (6)$$

The accuracy of the failure probability estimate can then be measured directly with the coefficient of variation (CoV) estimator:

$$\hat{\delta}_{\hat{\alpha}} = \frac{\hat{\sigma}_{\hat{\alpha}}}{\hat{\alpha}} = \sqrt{\frac{1 - \hat{\alpha}}{\hat{\alpha} N_{MC}}}. \quad (7)$$

We recall the definition of the quantile, resorting to the cumulative density function (CDF) $F_G : x \mapsto \mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) < x)$ of the random variable $G(\mathbf{X})$:

$$q = F_G^{-1}(\alpha) = \text{Inf}\{x \in \mathbb{R} : F(x) \geq \alpha\}. \quad (8)$$

A MC estimator $\hat{F}_{G, N_{MC}}$ of the CDF F_G reads:

$$\hat{F}_{G, N_{MC}}(x) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \mathbb{1}_{G(\mathbf{x}_i) < x}. \quad (9)$$

A MC-based estimator of the quantile reads:

$$\hat{q} = \hat{F}_{G, N_{MC}}^{-1}(\alpha). \quad (10)$$

A more practical expression reads:

$$\hat{q} = G(\mathbf{x}_{[\alpha N_{MC}]}), \quad (11)$$

after having reordered the set $\{G(\mathbf{x}_1), \dots, G(\mathbf{x}_{N_{MC}})\}$ in ascending order.

*The proposed method is based on the failure probability estimation algorithm eAK-MCS requiring to work in the Standard Space.

We assume that the CDF F_G is differentiable in \hat{q} , and $f_G(\hat{q}) > 0$, where the PDF of $G(\mathbf{X})$ is defined by $f_G(x) = \frac{dF_G(x)}{dx}$ when it exists. The central limit theorem for quantile estimators (section 2.3.3 of Reference 9) provides an asymptotic measure of error \hat{q} :

$$\sqrt{N_{MC}}(\hat{q} - q) \rightarrow \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f_G^2(\hat{q})}\right), \quad (12)$$

the arrow \rightarrow denoting here the convergence in distribution as N_{MC} goes to infinity. This result permits to theoretically build Confidence Intervals (CI) for the true quantile; for example, the 95% CI reads $\left[\hat{q} \pm 1.96 \frac{1}{f_G(\hat{q})} \sqrt{\frac{\alpha(1-\alpha)}{N_{MC}}}\right]$. A major difficulty lies in the estimation of the PDF (under existence) at \hat{q} (10). In this study, we restrict ourselves to estimating the quality of the quantile estimator by considering the one of the associated failure probability $\alpha_{\hat{q}} = \mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) < \hat{q})$. The accuracy of the MC-based estimator $\hat{\alpha}_{\hat{q}}$ of $\alpha_{\hat{q}}$ is quantified by its CoV estimator $\sqrt{\frac{1-\hat{\alpha}_{\hat{q}}}{\hat{\alpha}_{\hat{q}}N_{MC}}} \sim \sqrt{\frac{1}{\alpha N_{MC}}}$. Consequently, if a target CoV of $\delta < 1\%$ is desired, a quantile of level 10^{-9} would require a number of simulation $N_{MC} \sim \frac{1}{\alpha\delta^2} \sim 10^{13}$ yielding a prohibitive computational cost for industrial cases, and likely very expensive even using a surrogate model for G .

Various methods have been proposed in the literature to estimate failure probabilities and the reader may refer to Reference 11 for a critical review. Sampling methods include the aforementioned MC characterized by a low convergence rate, IS^{12,13} relying on a prudent choice of the IS density, and subset simulation (SS).¹⁴

Surrogate-based methods rely on the substitution of the performance function by a metamodel, one that is 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,¹⁵ AK-IS,¹⁶ KAIS,¹⁷ and AK-SS.¹⁸

Several algorithms have been proposed to deal with very small failure probabilities ($10^{-5} - 10^{-9}$) and multiple failure regions: Meta-IS,¹⁹ MetaAK-IS²,²⁰ BSS,²¹ ASVR,²² 2SMART,²³ AK-MCSi,²⁴ GPSS,²⁵ AK-MCS-IS,²⁶ S4IS,²⁷ and SS-KK.^{28,29} Some other methods such as SORM³⁰ or AK-IS are suitable for very small failure probabilities, but rely on the existence of an assumed unique so-called most probable failure point.

However, the adaptation of a failure probability estimation algorithm to an algorithm for quantile estimation is not straightforward, especially in the context of extreme quantiles (ie, quantiles associated with very small probabilities). We propose here a methodology for such adaptation. This methodology employs a previously developed failure probability estimation procedure to estimate small quantiles when the performance function J is expensive to calculate (eg, output of an expensive-to-run finite element model). The failure probabilities and quantiles are defined under a probability measure that can be recast as a multivariate standard Gaussian law using an isoprobabilistic transformation, regardless of restrictive assumption on J .

We note that some studies are focused on quantile estimation in case of an expensive performance function, such as References 31-35. However, these are not adapted for extreme quantiles. By contrast, the methodology proposed here is well suited for estimation of extreme quantiles, as we demonstrate with several examples.

In particular, Schobi³⁴ proposed to formulate the problem of estimation of failure probability and quantile in an unified way, based on the popular AK-MCS¹⁵ algorithm, originally developed for the estimation of failure probability. A Gaussian-Process-based surrogate³⁶ (polynomial chaos kriging, PCK, in Reference 34) of J is adaptively refined until a convergence criterion is reached, making the most of the uncertainty prediction of the surrogate. More precisely, the refinement algorithm based on AK-MCS permits to select a batch of samples added to the so-called *Design of Experiment* (DoE, a.k.a. plan of experiment), where the threshold value is the quantile estimated from a MC population and the surrogate.

In a very similar fashion, Bichon³⁵ replaces the AK-MCS *adaptive sampling step* with efficient global reliability analysis (EGRA) based on single refinement. Both methods from References 34,35 illustrate the use of failure probability refinement algorithms for quantile estimation in the context of *moderate (not extreme)* quantile levels.

When dealing with quantiles of small level, such a surrogate-based approach becomes intractable due to the prohibitive large size of the MC population that needs to be evaluated with the surrogate itself, required to obtain a satisfactory evaluation of the quantile.

The same comment applies to the computation of small failure probability, where AK-MCS³⁴ becomes unaffordable. eAK-MCS³⁷ extends AK-MCS for very small failure probability, inheriting a similar refinement strategy and general properties. It requires, however, to map the input random vector to the standard space. Similar to AK-MCS-based quantile estimation,³⁴ the adaptive refinement algorithm eAK-MCS is used to adaptively refine the

surrogate for a threshold equal to the estimated quantile. An IS scheme whose Density (ISD) is an isotropic centered Gaussian with large SD permits to satisfactorily estimate a small quantile ($O(10^{-9})$) with a tractable number of surrogate evaluations ($O(10^7)$). Another contribution of this study is a simple technique to propose in a parallel manner several quantiles for which levels the surrogate will be refined, which can be also used in the context of *larger* quantiles ($O(10^{-2} - 10^{-4})$).

It is emphasized that the framework proposed here uses eAK-MCS as a forward model for small failure probability estimation whose refinement algorithm serves to propose suitable samples to refine the surrogate. In this framework, eAK-MCS could be replaced by any *similar* surrogate-based methodology[†], such as Meta-IS,¹⁹ MetaAK-IS^{2,20} or MetaAL-OIS.³⁸ The main contribution of this paper lies in adapting such a reliability analysis tool for extreme quantile estimation.

The paper is structured as follows: Section 2 presents some general concepts such as Gaussian Processes and the theory of IS. In Section 3, the proposed algorithm named QeAK-MCS is described. Numerical experiments illustrating the method are presented in Section 4 to illustrate its efficiency. Conclusions are drawn in Section 5.

2 | BASIC INGREDIENTS

In this section, we describe two main ingredients for building the method proposed in this paper, that is, Gaussian Processes in Section 2.1 and IS in Section 2.2. In Section 2.3, a IS-based quantile estimator based on a Gaussian isotropic ISD is derived.

2.1 | Gaussian processes and kriging

In this subsection, concise details about Gaussian Processes and *Simple/Universal Kriging* in the noiseless cases are given. For comprehensive details about Kriging, the reader may refer to well-known references;^{36,39} more concise descriptions can be found in Reference 10,40.

The objective is to construct an approximation \tilde{y} of a true (expensive to evaluate) unknown function $y : \mathbb{X} \subseteq \mathbb{R}^d \rightarrow \mathbb{Y} \subseteq \mathbb{R}$, based on a training set, a.k.a DoE of size $n \in \mathbb{N}^*$, $\mathcal{T} = \{\mathbf{x}_i, y_i\}_{i \in \llbracket 1, n \rrbracket}$, where $y_i = y(\mathbf{x}_i)$. The output of the deterministic computer experiment y is assumed to be a *realization* of a real-valued random process Y indexed over \mathbb{X} . The set of the observed outputs $\mathbf{y} = (y_1, \dots, y_n)$ then appear as respective realizations of the random process $\mathbf{Y} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$. For $\mathbf{x} \in \mathbb{X}$, we assume $Y(\mathbf{x})$ of the form $Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x})$, with $\mu(\mathbf{x})$ being the mean of $Y(\mathbf{x})$ and $Z(\cdot)$ being a zero-mean stationary gaussian process, fully characterized by its symmetric positive definite autocovariance function (a.k.a. covariance kernel). The specific case of a known mean function μ is equivalent to consider the kriging formulation, applying the affine map $y \leftarrow y - \mu$, known as simple kriging (SK) with $Y(\mathbf{x}) = Z(\mathbf{x})$. Assuming the mean μ as a linear combination of so-called *basis functions* leads to the so-called Universal Kriging (UK) formulation $Y(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} + Z(\mathbf{x})$, where $\mathbf{f} : \mathbb{X} \rightarrow \mathbb{R}^p$, $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_p(\mathbf{x}))$ is a map of $p \in \mathbb{N}^*$ user-defined basis functions and $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown vector of weights to be determined. Note that the particular case of UK where \mathbf{f} boils down to $\mathbf{x} \mapsto 1$ is known as ordinary kriging (OK), leading to $Y(\mathbf{x}) = \beta_0 + Z(\mathbf{x})$, with $\beta_0 \in \mathbb{R}$ being the unknown constant trend. Considering a given $\mathbf{x} \in \mathbb{X}$ (untried location), the objective of kriging is to derive a random predictor of the unknown random process Y at \mathbf{x} , based on the observations \mathbf{Y} . The so-called best linear unbiased predictor (BLUP, see Reference 39) $\hat{Y}(\mathbf{x})$ is considered. The Kriging mean predictor is defined as $\mu_{\hat{Y}}(\mathbf{x}) = \mathbb{E}[\hat{Y}(\mathbf{x})]$, while the Kriging variance predictor yields $\sigma_{\hat{Y}}^2(\mathbf{x}) = \mathbb{E}[(\hat{Y}(\mathbf{x}) - Y(\mathbf{x}))^2]$, \mathbb{E} refers to the probability measure induced by the random process Y . Details permitting to derive the expression of the BLUP $\hat{Y}(\mathbf{x})$ and the predictors can be found in References 36,39.

In this paper, SK is considered with the anisotropic Matérn covariance kernel with regularity $\frac{5}{2}$ (Library *GPpy* written by the *Sheffield Machine Learning* group⁴¹). Note that the algorithm presented in this paper is obviously also compatible with any GP-based metamodel, for instance with more sophisticated trends (eg, PC-Kriging³⁴) or kernels. In cases where an isoprobabilistic transform T is used to recast the *physical* problem into the *standard* one, it is recommended to build a surrogate in the *physical* space, namely for $J(\mathbf{y})$, obtaining: $\tilde{G}(\mathbf{x}) = \mu_{\tilde{G}}(\mathbf{x}) = \mu_j(\mathbf{y})$ and $\sigma_{\tilde{G}}(\mathbf{x}) = \sigma_j(\mathbf{y})$, where

[†]Suitable for very small failure probability and multiple failure regions, providing samples for a given threshold value.

$\mathbf{y} = T^{-1}(\mathbf{x})$. In this work, both are systematically built and compared by means of leave-one-out estimation.⁴² In the analytical examples treated here, the surrogate built in the physical space was systematically better.

2.2 | Importance Sampling Theory

The Importance Sampling (IS) method^{2,12,13} is one of the most well-known variance reduction techniques used for assessing small failure probabilities. In this subsection, we present the general concept of IS, and apply it in Section 2.3 for IS-based quantile estimation.

The idea consists in drawing samples following another distribution than the original one in order to populate more frequently the so-called failure domain. The failure probability estimate is then obtained as a weighted average of these draws. We consider the generic computation of

$$p_g = \mathbb{E}_{\mathbf{X}}[g(\mathbf{X})], \tag{13}$$

where g is a scalar function (eg, $\mathbb{1}_{\mu_g < u}$, given a critical value $u \in \mathbb{R}$). Let h be a proposal PDF (a.k.a. biasing/instrumental PDF or ISD) assumed to dominate $gf_{\mathbf{X}}$ in the absolutely continuous sense:

$$\forall \mathbf{x} \in D_{\mathbf{X}}, h(\mathbf{x}) = 0 \Rightarrow g(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) = 0. \tag{14}$$

Then, p_g may be rewritten as follows:

$$p_g = \int_{\mathbb{R}^d} g(\mathbf{x}) \frac{f_{\mathbf{X}}(\mathbf{x})}{h(\mathbf{x})} h(\mathbf{x}) d\mathbf{x} = \mathbb{E}_h \left[g(\mathbf{X}) \frac{f_{\mathbf{X}}(\mathbf{X})}{h(\mathbf{X})} \right]. \tag{15}$$

It easily leads to the *IS estimator*:

$$\hat{p}_g = \frac{1}{N_{\text{IS}}} \sum_{i=1}^{N_{\text{IS}}} g(\mathbf{x}_i) \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{h(\mathbf{x}_i)}, \tag{16}$$

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

$$\hat{\sigma}_g^2 = \frac{1}{N_{\text{IS}} - 1} \left(\frac{1}{N_{\text{IS}}} \sum_{i=1}^{N_{\text{IS}}} g(\mathbf{x}_i)^2 \frac{f_{\mathbf{X}}(\mathbf{x}_i)^2}{h(\mathbf{x}_i)^2} - \hat{p}_g^2 \right). \tag{17}$$

The corresponding CoV $\hat{\delta}_g$, quantifying the estimation accuracy, is defined (if $\hat{p}_g \neq 0$) as

$$\hat{\delta}_g = \frac{\hat{\sigma}_g}{\hat{p}_g}. \tag{18}$$

The accuracy of the approximation given by IS critically depends on the choice of the ISD h . In this study, the ISD is chosen as $\mathcal{N}(0, \gamma^2 I_d)$ where $\gamma \geq 1$ is a parameter which is defined using a rule of thumb as discussed in Section 3.5 (or can also be tuned following Reference 37). Note that a Gaussian mixture ISD with suitable empirical parameters might be used,³⁸ but those empirical parameters would depend on the critical value u , which represents the unknown quantile q here.

2.3 | Quantile estimation using IS

At each step of QeAK-MCS, several estimations of quantiles based on the surrogate model \tilde{G} are required. When α is very small ($\alpha < 10^{-5}$), an accurate estimation using the MC approach becomes expensive, possibly unfeasible, even using surrogate evaluations.

In the literature, several Variance Reduction Techniques have been investigated for quantile estimation, including control variates,^{43,44} IS,² and IS combined with stratified sampling.⁴⁵ In this study, we propose to resort to the IS procedure using $h = f_{\mathcal{N}(0, \gamma^2 I_d)}$ as the ISD is preferred. We consider the generic problem of estimating the quantile q of level $\alpha \in]0, 1[$ associated to a scalar function g (eg, $g(\mathbf{x}) = \mu_{\hat{G}}(\mathbf{x})$):

$$\alpha = \mathbb{P}_{\mathbf{X}}(g(\mathbf{X}) < q). \quad (19)$$

A set of N_{IS} samples $(\mathbf{x}_1, \dots, \mathbf{x}_{N_{\text{IS}}}) \stackrel{\text{iid}}{\sim} f_{\mathcal{N}(0, \gamma^2 I_d)}$ is generated. The IS estimator (Equation (16)) permits to implicitly evaluate a quantile estimator \hat{q} :

$$\hat{q} = \text{Argmin} \left\{ q \in \mathbb{R} \text{ s.t. } \frac{1}{N_{\text{IS}}} \sum_{i=1}^{N_{\text{IS}}} \mathbb{1}_{g < q}(\mathbf{x}_i) \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{h(\mathbf{x}_i)} > \alpha \right\} .. \quad (20)$$

The sample set is re-ordered so $(g(\mathbf{x}_1), \dots, g(\mathbf{x}_{N_{\text{IS}}}))$ is sorted in ascending order. Note that the ISD $h = f_{\mathcal{N}(0, \gamma^2 I_d)}$ does not depend of \hat{q} . The implicit estimation of \hat{q} is rewritten as:

$$\alpha = \sum_{i=1}^{N_{\text{IS}}} \mathbb{1}_{g < \hat{q}}(\mathbf{x}_i) y_i, \quad (21)$$

where $y_i = \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{N_{\text{IS}} h(\mathbf{x}_i)}$. Let c_i denote:

$$c_i = \sum_{j=1}^i y_j, \quad \forall i \in \llbracket 1, N_{\text{IS}} \rrbracket. \quad (22)$$

It is then easy to show that $\exists ! l \in \llbracket 1, N_{\text{IS}} - 1 \rrbracket$ s.t. $c_l \leq \alpha < c_{l+1}$, if $c_1 \leq \alpha < c_{N_{\text{IS}}}$. A simple estimator of \hat{q} would then simply read:

$$\hat{q} = g(\mathbf{x}_1), \text{ if } \alpha < c_1 \quad (23)$$

$$= g(\mathbf{x}_{N_{\text{IS}}}), \text{ if } \alpha \geq c_{N_{\text{IS}}} \quad (24)$$

$$= g(\mathbf{x}_l), \text{ if } c_{l+1} \leq \alpha < c_l, l \in \llbracket 2, N_{\text{IS}} \rrbracket. \quad (25)$$

In order to slightly improve the accuracy of this quantile estimator, a linear interpolation[‡] is performed:

$$\hat{q} = g(\mathbf{x}_k) + (\alpha - c_k) \frac{g(\mathbf{x}_{k+1}) - g(\mathbf{x}_k)}{c_{k+1} - c_k}, \quad (26)$$

with

$$k = 1, \text{ if } \alpha < c_1 \quad (27)$$

$$= N_{\text{IS}} - 1, \text{ if } \alpha \geq c_{N_{\text{IS}}} \quad (28)$$

$$= l, \text{ if } c_1 \leq \alpha < c_{N_{\text{IS}}}. \quad (29)$$

Due to numerical robustness considerations, the linear estimator (Equation (26)) is replaced by the simple estimator (Equation (25)) if $|c_{k+1} - c_k| < 10^{-14}$. This IS-based quantile estimation is the key of the extension of eAK-MCS for extreme quantile estimation, since it allows one to significantly decrease the CPU cost associated to the estimation of an extreme quantile, even based on a surrogate model. The extension is described in the next section.

[‡]This step can be replaced by a higher order (eg, quadratic) interpolation. Numerical experiments do not suggest its relevance.

3 | THE QeAK-MCS ALGORITHM

In this section, we propose a new method for the estimation of extreme quantiles, denoted as the QeAK-MCS algorithm. It is aimed at the problem of estimating the quantile q of the performance function G , given the level α , as in Equation (4). The algorithm builds a GP-based surrogate of G , refining it iteratively considering the estimated quantile as the critical value of the associated failure probability estimation.

QeAK-MCS is similar to AK-MCS-based quantile estimation,³⁴ however it uses IS instead of MC (as is done in AK-MCS), thereby extending the approach from eAK-MCS³⁷ to quantile estimation. The main steps of QeAK-MCS can be summarized as follows:

- (1) *Initial DoE*: An experimental design \mathcal{X} is generated by Latin-Hypercube Sampling (LHS) (see Section 3.1).
- (2) *IS population*: A set of N_{IS} sample points $\mathcal{S} = (\mathbf{x}_1, \dots, \mathbf{x}_{N_{\text{IS}}}) \stackrel{\text{iid}}{\sim} f_{\mathcal{N}(0, \gamma^2 I_d)}$ is generated.
- (3) *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 $\{\mathcal{X}, \mathcal{Y}\}$ (see Section 2.1).
- (4) *Critical values/quantiles selection*: The surrogate-based quantile estimate \hat{q} is obtained using IS. A set of K_q quantiles $(\hat{q}_1, \dots, \hat{q}_{K_q})$ around \hat{q} is selected, based on the surrogate information (see Section 3.2).
- (5) *eAK-MCS-based samples selection*: for each \hat{q}_l , $l \in \llbracket 1, K_q \rrbracket$, K_p samples $(\mathbf{x}_1^l, \dots, \mathbf{x}_{K_p}^l)$ are selected following the refinement step of the algorithm eAK-MCS, for the failure probability $\mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) < \hat{q}_l)$. Details are given Subsection 3.3 for the sake of self-completeness of the paper.
- (6) *Filtering procedure and surrogate update*: An a posteriori filtering procedure is performed on the selected samples $\mathcal{X}^* = \{\mathbf{x}_j^l\}_{j \in \llbracket 1, K_q \rrbracket, l \in \llbracket 1, K_p \rrbracket}$, removing too close points: for a, b distinct samples of \mathcal{X}^* , if $\|a - b\|_2 < \text{TOL}$, a is discarded from \mathcal{X}^* . It permits to avoid samples too close to each other (thereby preventing metamodel training issues). The selected samples \mathcal{X}^* are added to the experimental design \mathcal{X} , and Step 3 is applied.
- (7) *Stopping criterion*: If a stopping criterion is satisfied (see Section 3.4), the enrichment stops. Otherwise the algorithm goes back to step 4.

3.1 | Initial DoE

An initial design of size $n_0 = 5d^{19,21,40}$ is generated as described in Reference 21. A compact subset $\mathbb{X}_0 = \prod_{i=1}^d [q_\delta^i, q_{1-\delta}^i]$ is constructed, where q_δ^i and $q_{1-\delta}^i$ are, respectively, the quantiles of order δ and $1 - \delta$ of the i th input variable. Working in the standard space, \mathbb{X}_0 reads $[\Phi(\delta), \Phi(-\delta)]^d$, Φ denoting the CDF of the univariate normal Gaussian law. 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. This rescaled design is the initial DoE \mathcal{X} .

3.2 | Critical values/quantiles selection

At each step of the refinement algorithm, K_q quantiles are proposed, to serve afterward as critical values for the eAK-MCS refinement algorithm.

First, the surrogate-based quantile \hat{q} is estimated using IS (Equation (26)):

$$\alpha = \mathbb{P}_{\mathbf{X}}(\mu_{\hat{G}}(\mathbf{X}) < \hat{q}). \quad (30)$$

Note that if $K_q = 1$, the selected quantile is simply \hat{q} . Bounds \hat{q}^- , \hat{q}^+ of the quantile estimate \hat{q} are derived based on the surrogate Gaussian nature, using in particular the predictive standard deviation:

$$\alpha = \mathbb{P}_{\mathbf{X}}(\mu_{\hat{G}}(\mathbf{X}) + k\sigma_{\hat{G}}(\mathbf{X}) < \hat{q}^+), \quad (31)$$

$$\alpha = \mathbb{P}_{\mathbf{X}}(\mu_{\hat{G}}(\mathbf{X}) - k\sigma_{\hat{G}}(\mathbf{X}) < \hat{q}^-), \quad (32)$$

where $k > 0$ sets the confidence level.

3.2.1 | Linear selection

The approach followed in this paper consists in choosing linearly K_q quantiles in $[\hat{q}^-, \hat{q}^+]$:

$$\hat{q}_l = \hat{q}^- + (l-1) \frac{\hat{q}^+ - \hat{q}^-}{K_q - 1}, \quad \forall l \in \llbracket 1, K_q \rrbracket. \quad (33)$$

Two other approaches, implemented but not studied in this paper, might be considered. Based on the experience of the authors, there is no clear advantage of one selection criterion over another. The strategies Linear- α and Linear- k involve a slight additional computational burden, negligible though since based on surrogate evaluations. They are provided in Appendix A1, and might be combined during the sampling refinement.

3.3 | eAK-MCS selection

In this subsection, we recall the basics of the refinement selection of eAK-MCS (see Reference 37 for details). Given a critical value u , K_p samples are selected for the refinement of the performance function surrogate \hat{G} . For each sample $\mathbf{x} \in \mathbb{R}^d$, the so-called probability of misclassification $P_m^u(\mathbf{x})$ is defined as Reference 46:

$$P_m^u(\mathbf{x}) = \Phi \left(-\frac{|\mu_{\hat{G}}(\mathbf{x}) - u|}{\sigma_{\hat{G}}(\mathbf{x})} \right). \quad (34)$$

The popular U-function^{15,34} is defined as:

$$U^u(\mathbf{x}) = \frac{|\mu_{\hat{G}}(\mathbf{x}) - u|}{\sigma_{\hat{G}}(\mathbf{x})}. \quad (35)$$

The eAK-MCS refinement strategy reads as follows:

- (1) 1 sample \mathbf{x}_0^* is selected among S following the single eAK-MCS selection: $\mathbf{x}^* = \arg \min_{\mathbf{x} \in S} U^u(\mathbf{x})$.
- (2) $K_p - 1$ samples $(\mathbf{x}_1^*, \dots, \mathbf{x}_{K_p-1}^*)$ are simultaneously selected among the set $\mathbb{M}^{\bar{k}}(u) = \{\mathbf{x} \in S : \mu_{\hat{G}}(\mathbf{x}) - \bar{k}\sigma_{\hat{G}}(\mathbf{x}) < u\} \setminus \{\mathbf{x} \in S : \mu_{\hat{G}}(\mathbf{x}) + \bar{k}\sigma_{\hat{G}}(\mathbf{x}) < u\}$, using a weighted K-means clustering technique^{34,37} where the weights are chosen as $P_m^u(\mathbf{x})$ for each sample $\mathbf{x} \in \mathbb{M}^{\bar{k}}(u)$. If this method returns only $K_1 < K_p - 1$ samples ($K_1 = 0$ possibly), then the very same method is applied to the full IS population S to provide the remaining $K_p - 1 - K_1$ samples.

3.4 | Stopping criterion

The stopping criterion adopted is the same as the one proposed in Reference 34, focusing on the accuracy of the quantity of interest, hence on the upper and lower bounds of the quantile:

$$\frac{\tilde{q}^+ - \tilde{q}^-}{q_{\text{ref}}} < \epsilon, \quad (36)$$

for two consecutive iteration steps, where the quantiles bounds are estimated by:

$$\alpha = \mathbb{P}_X(\mu_{\hat{G}}(\mathbf{X}) + \tilde{k}\sigma_{\hat{G}}(\mathbf{X}) < \tilde{q}^+), \quad (37)$$

$$\alpha = \mathbb{P}_X(\mu_{\hat{G}}(\mathbf{X}) - \tilde{k}\sigma_{\hat{G}}(\mathbf{X}) < \tilde{q}^-), \quad (38)$$

the parameter \tilde{k} setting the confidence interval for the stopping criterion. Note that \tilde{k} is in practice different from k used in the selection of quantiles, and \bar{k} used for the selection of samples in Section 3.3. q_{ref} is a positive scalar value permitting

TABLE 1 Tuning parameters

\tilde{k}	k	\bar{k}	n_0	δ	K_q	K_p	TOL	ϵ	γ	N_{IS}
1	3	2	$5d$	10^{-5}	3	3	10^{-4}	5%	Equation (39)	10^7

to normalize the quantile. In industrial cases, experts are likely to be able to provide such normalization constant. In the analytical benchmark functions studied here, q_{ref} is replaced by σ_G , the standard deviation of $G(\mathbf{X})$ under the distribution of \mathbf{X} . It is evaluated beforehand using MC, with a large sample size, on the analytical function G . It is suggested in Reference 34 to replace q_{ref} by the surrogate-based estimate of the standard deviation of the performance function. According to the authors experience, this approach can be misleading since at a given iteration step, this estimate might be highly inaccurate and severely overestimated, leading the refinement algorithm to stop prematurely.

3.5 | Quantile 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 otherwise).

For eAK-MCS,³⁷ the choice of γ in the ISD has a large impact on the efficiency of the method, and an automatic tuning method was proposed based on the evaluation of the failure probability. Here, since the level α is fixed (can be interpreted as a failure probability, for a critical value equal to \hat{q}), a less elaborate tuning method only based on α is sufficient to obtain satisfactory results. If $\alpha > 10^{-3}$, a MC-based method is enough to obtain an accurate quantile (CoV of 1%) with a reasonable number of samples (10^7), so $\gamma = 1$ is fine. For smaller α , a linear law in $\log_{10}(\alpha)$ is chosen, assuming that if $\alpha = 10^{-9}$, $\gamma = 2.5$ permits satisfactory results. The rule of thumb for tuning γ is consequently defined as:

$$\gamma = \max\left(1, \frac{1}{4} (1 - \log_{10}\alpha)\right). \tag{39}$$

3.6 | Illustrative example

To showcase the process of selecting multiple additional samples, we consider the example discussed in more detail in Section 4.2, initialized with a DoE of size 10. In the process of selecting $K = K_q K_p = 18$ samples at each refinement step, $K_q = 3$ quantiles ($\hat{q}_1, \hat{q}_2, \hat{q}_3$) are estimated (Section 3.2). Note that ($\hat{q}_1 = \hat{q}^-, \hat{q}_2 = \hat{q}, \hat{q}_3 = \hat{q}^+$), for $K_q = 3$. Then, $K_p = 6$ samples are selected using the eAK-MCS refinement considering \hat{q}_i as the critical value, as shown in Figures 1, 2, and 3,

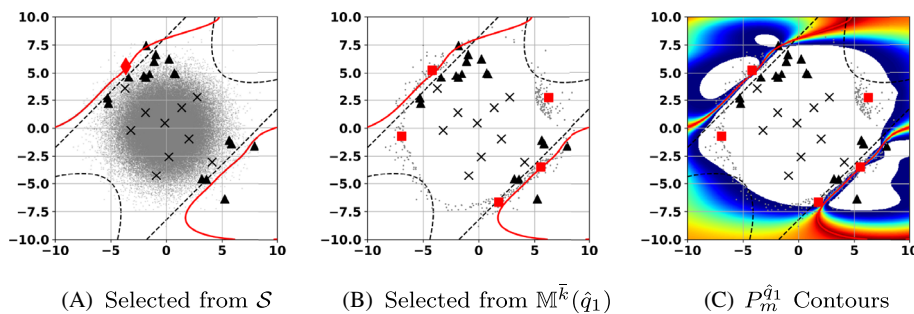


FIGURE 1 Refinement illustration for $\hat{q}_1 = \hat{q}^- = -4.84$. Black crosses and triangles represent, respectively, the initial DoE, and the selected design of experiment (DoE) at the first refinement step. The dashed black line represents the level $G = q$ based on the true performance function G and the exact quantile $q = -4$, the red line represents the level $\mu_{\hat{c}} = \hat{q}_1$ based on the surrogate, and the DoE selected based on the level $\mu_{\hat{c}} = \hat{q}_1$ is indicated in red (diamond/squares). The diamond in (A) represents the point selected from the importance sampling population S shown with grey dots in (A). The squares in (B) and (C) represent the points selected from $\hat{M}^{\bar{k}}(\hat{q}_1)$ shown with grey dots in (B) and (C), with $\bar{k} = 2$. In (C), the contours of the probability of misclassification with respect to the critical level \hat{q}_1 are shown, with blue ~ 0 and red ~ 1 . For clarity, zones where $P_m < 10^{-4}$ are indicated in white. A, Selected from S ; B, Selected from $\hat{M}^{\bar{k}}(\hat{q}_1)$; C, $P_m^{\hat{q}_1}$ contours [Color figure can be viewed at wileyonlinelibrary.com]

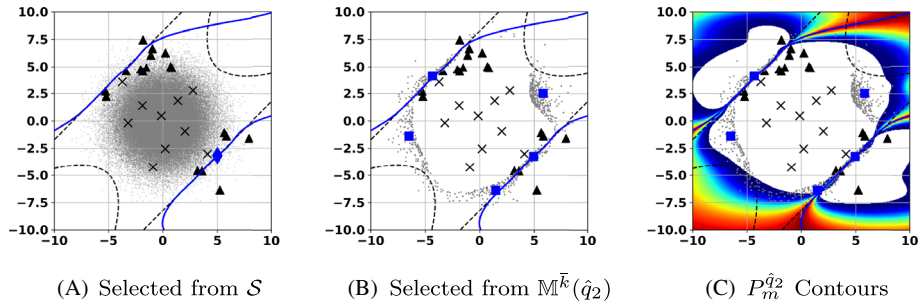


FIGURE 2 Refinement illustration for $\hat{q}_2 = -3.99$. The blue line represents the level $\mu_{\hat{G}} = \hat{q}_2$ and the design of experiment selected based on this level is indicated in blue (diamond/squares). See Figure 1 for legend settings. A, Selected from \mathcal{S} ; B, Selected from $\mathbb{M}^{\bar{k}}(\hat{q}_2)$; C, $P_m^{\hat{q}_2}$ contours [Color figure can be viewed at wileyonlinelibrary.com]

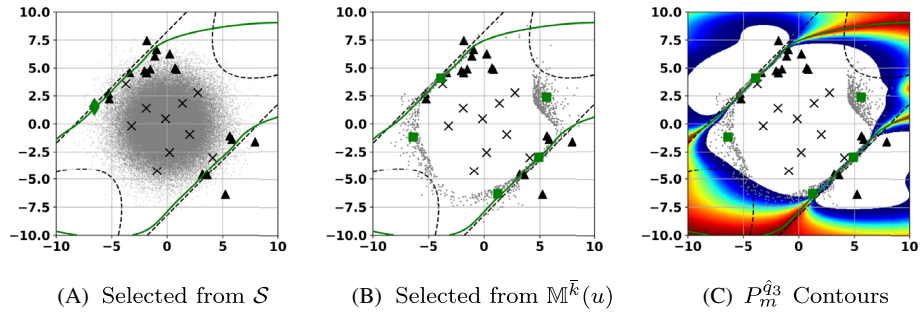


FIGURE 3 Refinement illustration for $\hat{q}_3 = \hat{q}^+ = -3.66$. The green line represents the level $\mu_{\hat{G}} = \hat{q}_3$ and the design of experiment selected based on this level is indicated in green (diamond/squares). See Figure 1 for legend settings. A, Selected from \mathcal{S} ; B, Selected from $\mathbb{M}^{\bar{k}}(u)$; C, $P_m^{\hat{q}_3}$ contours [Color figure can be viewed at wileyonlinelibrary.com]

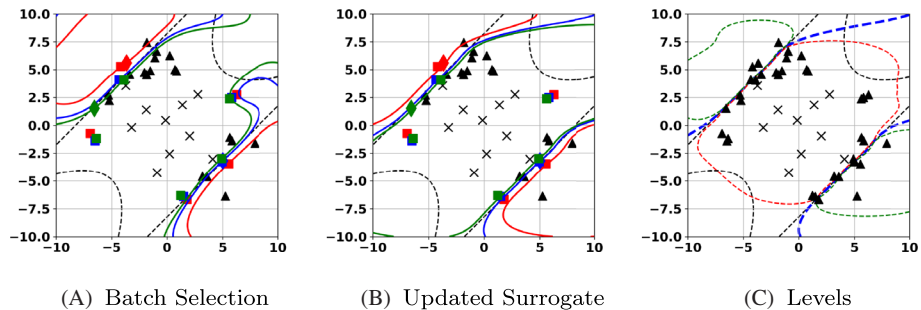


FIGURE 4 Illustrative example: Design of experiment and surrogate. Settings are as in Figure 1, with red, blue, and green corresponding, respectively, to \hat{q}_1 , \hat{q}_2 , \hat{q}_3 . (C) represent the contours for $\mu_{\hat{G}} = q$, $\mu_{\hat{G}} + 2\sigma_{\hat{G}} = q$ and $\mu_{\hat{G}} - 2\sigma_{\hat{G}} = q$ represented, respectively, in blue, green, and red dashed lines. A, Batch selection; B, Updated surrogate; C, Levels [Color figure can be viewed at wileyonlinelibrary.com]

respectively, $i = 1$, $i = 2$ and $i = 3$. For the sake of illustration, the second refinement step is shown, starting from a DoE composed of the initial DoE, and the samples added to the experimental design after the first refinement step. The size of the IS population is set here to $N = 10^5$.

Figure 4A,B shows the DoE refinement step, respectively, before and after the metamodel update. Figure 4C represents the true limit-state surface associated to the exact quantile $q = -4$, and the surrogate levels associated to the quantile estimates \hat{q}_1 , \hat{q}_2 , and \hat{q}_3 .

4 | NUMERICAL EXPERIMENTS

In this section, we consider the application of QeAK-MCS[§] to several test-cases, involving $\alpha \sim 10^{-4}$ to 10^{-9} , to showcase the suitability of the proposed algorithm to deal with small to extreme quantiles. Three 2D, one 6D, and one 8D examples are investigated.

Very accurate estimation of extreme quantiles of an analytical function can be a complex task. To alleviate this issue, test-cases available in the literature in the context of failure probability estimation are considered: the reference quantile is then taken as the critical value of the problem of interest, while the fixed quantile level α is set to an accurate estimate of the failure probability (CoV < 0.20%), using either SS²¹ or IS with a Gaussian mixture.³⁷

To the best of the knowledge of the authors, no other method is suitable for extreme quantile estimation as discussed in the Introduction. This makes comparison of our proposed method to other approaches from the literature difficult. Notwithstanding, we present an example (Section 4.5) where we compare QeAK-MCS with the AK-MCS approach from Reference 34, in case of *moderate* quantiles.

A parallel refinement strategy is adopted, with $K_q K_p = 9$ samples added at each iteration step ($K_q = 3$, $K_p = 3$), to showcase the suitability of QeAK-MCS for parallel environment. Note that the exploratory nature of the refinement strategy suggest the use of $K_q, K_p > 1$, even considering sequential refinement[¶]. The other tuning parameters are the ones provided in Table 1, except when explicitly mentioned. The initial DoE is assumed to be evaluated in 1 iteration.

Because of the stochastic nature of the proposed method, we assess the statistical significance of the results for each test case with 50 independent runs (replications). Thus, the number of calls N_{calls} to the performance function, the number of iterations N_{iter} , the quantile estimate \hat{q} , and the final *relative* error $\epsilon_q = \frac{|\hat{q} - q|}{q_{\text{ref}}}$ are random variables, and we show their average over the 50 realizations resulting from the independent runs. Furthermore, we show the CoV of \hat{q} (also computed from the 50 realizations), as it is an indicator of the robustness of the method. The lower the CoV is, the more likely is the algorithm to return an estimate of \hat{q} which is close to its asymptotic average over many independent runs.

We point out that the accuracy of the estimates obtained with QeAK-MCS is determined by two ingredients of the algorithm. One is the quality of the refined GP-based surrogate of the performance function, the other is the accuracy of the quantile estimation given the surrogate. Thus, there are two main sources of error: (i) the difference between the true performance function and its surrogate and (ii) the statistical error in the estimation of quantiles based on the surrogate with IS.

Note that for some analytical examples studied here, the true quantile q is zero, so assessing the classic relative error is not suitable. Instead, q_{ref} is chosen here as the standard deviation of the performance function σ_G , estimated on the analytical using MC with 10^7 samples. The quantity $\hat{\hat{q}}$, defined as $\hat{\hat{q}} = \frac{\hat{q} - q}{q_{\text{ref}}}$, is therefore introduced as a normalized measure of the quantile estimate.

The average CoV of the estimation of the failure probability $\mathbb{E}[\text{CoV}[\hat{\alpha}]]$ based on IS (Equation (18)) is also given in order to assess the quality of the IS procedure based on the ISD $f_{\mathcal{N}(0, \gamma^2 I_d)}$, that is, the accuracy of the estimated quantile \hat{q} based on the sampling procedure (not the quality of the surrogate).

For each case, figures showing the average *relative* error ϵ_q , and the average normalized quantile estimate $\hat{\hat{q}}$ 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. The $2\text{-}\sigma$ confidence interval is represented with black dashed lines. For two-dimensional (2D) examples, an illustration of the final DoE and refined metamodel is also provided, when the convergence criterion is satisfied, based on a single run.

Remark. The parallel strategy refinement might propose strictly less than $K_q K_p$ samples due to the *a posteriori* filtering procedure. A postprocessing step involving a linear interpolation procedure on ϵ_q and \hat{q} is required at the end of the 50 runs.

[§]The in-house software is implemented in Python 3.6, using the package GPy for the surrogate modeling.

[¶]This comment is based on authors experience, and is not supported by rigorous numerical experiments.

4.1 | Single failure region 2D

A first classic 2D example is taken from References 16,20. This example is characterized by a low failure probability ($\alpha \sim 3 \times 10^{-5}$), a very smooth limit state and a single failure region. The performance function in the standard space reads:

$$G(x_1, x_2) = \frac{1}{2}(x_1 - 2)^2 - \frac{3}{2}(x_2 - 5)^3 - 3. \quad (40)$$

where $X_1, X_2 \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. The probability of failure reads $\alpha = \mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) < q) = 2.874 \times 10^{-5}$, estimated with a CoV of 0.03% using IS with a Gaussian mixture ISD³⁸ based on 10^7 samples, for $q = 0$.

4.1.1 | Results

In Table 2, we compare the reference with the results of QeAK-MCS, based on 50 independent runs. For all runs, the QeAK-MCS algorithm stops after one refinement step (nine samples added to the initial DoE), yielding a very accurate estimate of the quantile, the average relative error being below 0.05%.

An illustration is provided in Figure 5A, showing the refined metamodel for a single run. The DoE selected is well clustered around the True LSS $G = q$, showing that both the quantile \hat{q} and the surrogate are well estimated. This is further illustrated in the quantile history in Figure 5C where the $2 - \sigma$ confidence interval of \hat{p}_f rapidly merges and converges to the reference value. The relative error history, Figure 5B, demonstrates the convergence behavior of the method in that case, where the mean error is below 0.01% after 60 calls. This convergence behavior is confirmed by the very low mean CoV of the failure probability estimate based on IS $\mathbb{E}[\text{CoV}[\hat{\alpha}]]$, which assesses the efficiency of the ISD in that low dimensional case. The method in that case is also very robust, with $\text{CoV}[\hat{q}] < 0.05\%$, which is further demonstrated with the $2 - \sigma$ confidence interval rapidly merging to the reference value Figure 5B.

The original function being a second-order polynomial easily captured by the metamodel, this example is rather simple, also because it is characterized by a single failure region and low dimensionality. The low value of $\epsilon_{pf}^{\text{rel}}$ can be explained

Method	$\mathbb{E}[N_{\text{calls}}]$	$\mathbb{E}[N_{\text{iter}}]$	$\mathbb{E}[\hat{q}]$	$\text{CoV}[\hat{q}]$	$\mathbb{E}[\epsilon_q]$	$\mathbb{E}[\text{CoV}[\hat{\alpha}]]$
Reference	10^7	-	0	-	0	$< 0.04\%$
QeAK-MCS	19.0	2.0	-1.023×10^{-3}	0.03%	0.02%	0.44%

TABLE 2 Results of the single failure region two dimensional

Note: Solving $\mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) \leq \hat{q}) = 2.874 \times 10^{-5}$, with $q_{\text{ref}} = \sigma_G = 121.334$.

^aReference: Importance sampling (IS) with a Gaussian mixture as importance sampling density,³⁶ 50 independent runs. Coefficient of variation (CoV) $< 0.04\%$

^bInitial design of experiment size: 10. Nine samples iteratively added. Based on 50 independent runs.

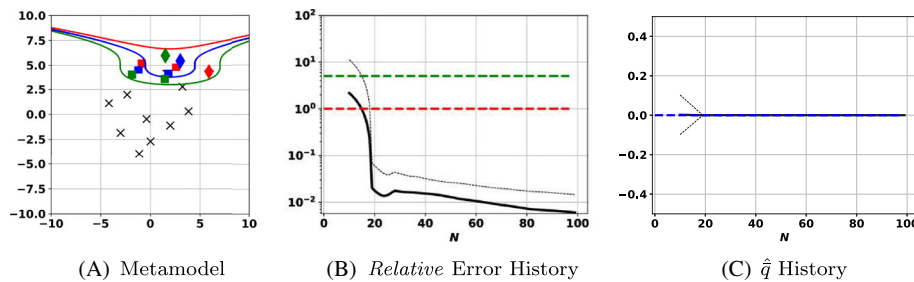


FIGURE 5 Results of the single failure region two dimensional. (A) Legend settings in Figure 4. (B) Green and red thick dashed lines indicate, respectively, 5% and 1% relative error. The y-axis is logarithmic. The average relative error bias and the associated $2 - \sigma$ confidence interval are represented, respectively, in black thick line and black thin dashed lines. (C) The normalized predicted quantile estimate \hat{q} and the associated $2 - \sigma$ confidence interval are represented respectively in black thick line and black thin dashed lines. A, Metamodel; B, relative error history; C, \hat{q} history [Color figure can be viewed at wileyonlinelibrary.com]

by both the good quality of the surrogate, and the accuracy of the quantile based on the IS sampling method, illustrated by the low $\mathbb{E}[\text{CoV}[\hat{\alpha}]]$.

4.2 | Four-branch series system 2D

This example is a variation of a classical structural reliability test case.¹⁹⁻²¹ The performance function is defined as:

$$G(x_1, x_2) = \min \left\{ \begin{array}{l} 3 + \frac{(x_1 - x_2)^2}{10} - \frac{x_1 + x_2}{\sqrt{2}} \\ 3 + \frac{(x_1 - x_2)^2}{10} + \frac{x_1 + x_2}{\sqrt{2}} \\ x_1 - x_2 + \frac{7}{\sqrt{2}} \\ -(x_1 - x_2) + \frac{7}{\sqrt{2}} \end{array} \right\}, \tag{41}$$

and $X_1, X_2 \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. The failure probability is $\alpha = \mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) \leq q) = 5.596 \times 10^{-9}$, estimated with a CoV of about 0.04%,²¹ based on 100 runs of SS with sample size 10^7 , for $q = -4$. This 2D example characterized by four failure domains and an extreme quantile level represents a very challenging case, albeit low dimensional.

4.2.1 | Results

The results are presented in Table 3 and Figure 6. QeAK-MCS stops after ~ 99 calls (~ 11 iterations) in average, resulting in a very accurate and robust result, with respectively a mean relative error of 0.57% and a CoV of \hat{q} of 0.82%.

As seen in Figure 6A, 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 and the presence of DoE clustered in that zone. Figure 6B,C shows, respectively, the average relative error and the quantile estimate history as a function of the number of performance calls, with corresponding 2- σ confidence interval. One can note that after on average 40 function calls (4-5 iterations), the average predicted quantile \hat{q} is already of the same order of magnitude as the reference value, and the convergence is then rather fast, with a mean relative error lower than 1% after ~ 90 function calls (\sim eight iterations). The average CoV associated with the IS estimate of the corresponding failure probability $\mathbb{P}(\mu_{\hat{G}} < \hat{q})$ is still low for this 2D case, about 0.95 %, illustrating the efficiency of the IS method with the Gaussian ISD $\mathcal{N}(0, \gamma^2 I_d)$ in this case. This translates into an accurate estimate of the quantile, for a given surrogate accuracy, which is confirmed by the convergence trend of the relative error which decreases as the surrogate is refined.

TABLE 3 Results of the four-branch series system two dimensional

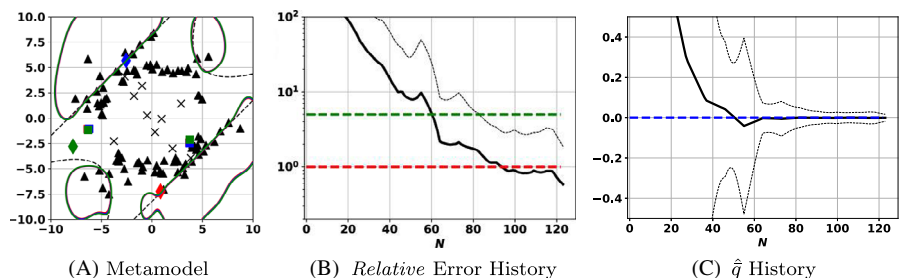
Method	$\mathbb{E}[N_{\text{calls}}]$	$\mathbb{E}[N_{\text{iter}}]$	$\mathbb{E}[\hat{q}]$	CoV $[\hat{q}]$	$\mathbb{E}[\epsilon_q]$	$\mathbb{E}[\text{CoV}[\hat{\alpha}]]$
Reference	10^7	-	-4	-	0	< 0.05%
QeAK-MCS	98.4	10.9	-3.999	0.82%	0.57%	0.95%

Note: Solving $\mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) \leq \hat{q}) = 5.596 \times 10^{-9}$, with $q_{\text{ref}} = \sigma_G = 0.6265$.

^aReference: Subset simulation,²¹ 100 independent runs. Coefficient of variation (CoV) < 0.05%.

^bInitial design of experiment size: 10. Nine samples iteratively added. Based on 50 independent runs.

FIGURE 6 Results of the four-branch series system two dimensional. Legend Settings in Figure 5. A, Metamodel; B, relative error history; C, \hat{q} history [Color figure can be viewed at wileyonlinelibrary.com]



4.3 | Deviation of a Cantilever Beam 2D

This 2D example characterized by a single failure region and a small quantile level $\alpha \sim 4 \times 10^{-6}$ is taken from Reference 21, where the deflection of the tip of a Cantilever Beam with a rectangular cross-section reads

$$f(x_1, x_2) = \frac{3L^4 x_1}{2Ex_2^3}, \quad (42)$$

where $L = 6$, $E = 2.6 \times 10^4$, X_1 and X_2 are assumed independent, with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $\mu_1 = 10^{-3}$, $\sigma_1 = 0.2\mu_1$, $\mu_2 = 0.3$ and $\sigma_2 = 0.1\mu_2$. The failure probability reads $\alpha = \mathbb{P}_{\mathbf{X}}(-f(\mathbf{X}) < q) = 5.596 \times 10^{-9}$, estimated with a CoV of about 0.03%,²¹ based on 100 runs of SS with sample size 10^7 , for $q = -\frac{L}{325}$.

4.3.1 | Results

The results are presented in Table 4 and Figure 7. The QeAK-MCS algorithm stops after ~ 29 calls on average (three iterations), yielding a mean relative error lower than 1.2% at the end of the refinement algorithm, whose robustness is quantified by the final low CoV of $\hat{q} < 1.5\%$. Figure 7B shows that the average relative error is below 10% after one refinement step (~ 19 calls), which explains the sharp decrease to 0 of the mean normalized \hat{q} in Figure 7C. For illustration purposes, the history is truncated to 35 calls. Indeed, for few runs characterized 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 IS-based quantile estimation shows good efficiency, the average CoV of the failure probability estimated with IS being low $\mathbb{E}[\text{CoV}[\hat{\alpha}]] \sim 0.51\%$. The final metamodel is represented in Figure 7A.

4.4 | Response of a nonlinear oscillator 6D

This example is taken from Reference 21. It consists of a nonlinear undamped single degree of freedom system.¹⁵ In particular, the performance function is given as follows:

$$G(c_1, c_2, m, r, t_1, F_1) = 3r - \left| \frac{2F_1}{m\omega_0^2} \sin\left(\frac{\omega_0 t_1}{2}\right) \right|, \quad (43)$$

Method	$\mathbb{E}[N_{\text{calls}}]$	$\mathbb{E}[N_{\text{iter}}]$	$\mathbb{E}[\hat{q}]$	CoV $[\hat{q}]$	$\mathbb{E}[\epsilon_q]$	$\mathbb{E}[\text{CoV}[\hat{\alpha}]]$
Reference	10^7	-	$\sim -1.84615 \times 10^{-2}$	-	0	$< 0.04\%$
QeAK-MCS	28.8	3.1	-1.84631×10^{-2}	1.44%	1.17%	0.51%

TABLE 4 Results of the deviation of a cantilever beam two dimensional

Note: Solving $\mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) < -\frac{6}{325}) = 5.596 \times 10^{-9}$, with $q_{\text{ref}} = \sigma_G = 1.1501 \times 10^{-3}$.

^aReference: Subset simulation,²¹ 100 independent runs. Coefficient of variation (CoV) $< 0.04\%$.

^bInitial design of experiment size: 10. Nine samples iteratively added. Based on 50 independent runs.

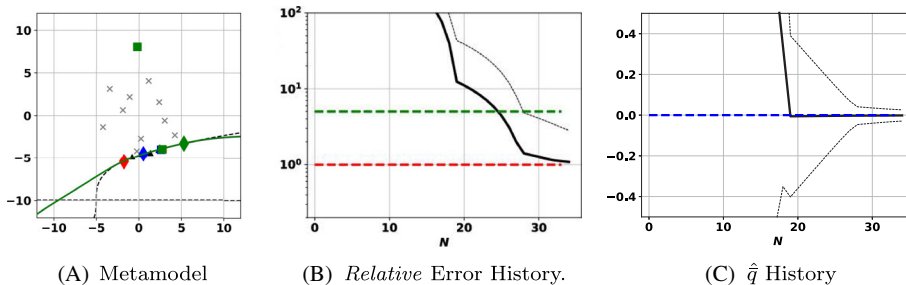


FIGURE 7 Results of the deviation of a cantilever beam two dimensional. Legend settings in Figure 5. A, Metamodel; B, Relative error history; C, \hat{q} history [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Random variables²¹ of Example 5

Variable	Probability density function	Mean	SD
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

^aVariables are independent.

TABLE 6 Results of response of a nonlinear oscillator six dimensional.

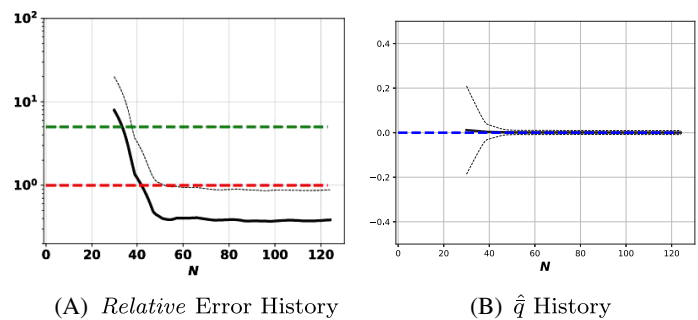
Method	$\mathbb{E}[N_{\text{calls}}]$	$\mathbb{E}[N_{\text{iter}}]$	$\mathbb{E}[\hat{q}]$	CoV $[\hat{q}]$	$\mathbb{E}[\epsilon_q]$	$\mathbb{E}[\text{CoV}[\hat{\alpha}]]$
Reference ^a	-	-	0	-	0	< 0.05%
QeAK-MCS ^b	41.2	2.2	-2.46×10^{-4}	1.39%	1.01%	2.52%

Note: Solving $\mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) \leq \hat{q}) = 1.514 \times 10^{-8}$, with $q_{\text{ref}} = \sigma_G = 0.18267$.

^aReference: Subset sampling,²¹ 100 independent runs. Coefficient of variation (CoV) < 0.05%.

^bInitial design of experiment size: 30. Nine samples iteratively added. Based on 50 independent runs.

FIGURE 8 Results of response of a nonlinear oscillator six dimensional. Legend settings in Figure 5. A, Relative error history; B, \hat{q} history [Color figure can be viewed at wileyonlinelibrary.com]



with $\omega_0 = \sqrt{\frac{c_1+c_2}{m}}$. The six random variables (assumed independent) are listed in Table 5, where the variability of F_1 is modified with respect to Reference 15 in order to decrease the failure probability, to $\alpha = \mathbb{P}_{\mathbf{X}}(f(\mathbf{X}) < q) = 1.514 \times 10^{-8}$. It is estimated with a CoV of about 0.04%,²¹ based on 100 runs of SS with sample size 10^7 , for $q = 0$.

4.4.1 | Results

The results for this six-dimensional example characterized by $\alpha \sim 1.5 \times 10^{-8}$ are presented in Table 6 and Figure 8. This example, due to the higher dimensionality and the very low level of the quantile sought, is rather challenging. Interestingly, QeAK-MCS performs very well here, the algorithm stopping in average after ~ 41 calls (~ 2.2 iterations), with approximately 1 or 2 refinement steps only. The estimated quantile returned is subjected to a mean relative error of $\sim 1\%$, in a robust way with a CoV on \hat{q} of $\sim 1.4\%$, which is remarkable. A look at the history of the mean relative error and \hat{q} in Figure 8 A,B, respectively, shows that the estimated quantile rapidly converges to the solution. The mean relative error reaches a plateau as the surrogate is refined, contrary to the other 2D examples for which it kept decreasing. This behavior can be explained by the deteriorated efficiency of the Gaussian ISD for the IS quantile estimation of the surrogate, for which the mean CoV of the failure probability estimator $\mathbb{E}[\text{CoV}[\hat{\alpha}]]$ is rather large, 2.52%, limiting the accuracy of the quantile based on the surrogate, independently of the accuracy of the surrogate itself. This tendency confirms the quality of the proposed refinement algorithm and of the final surrogate, even if the IS procedure cannot extract the quantile very accurately in this higher dimensional case. This behavior also confirms the tendency observed in Reference 37 for the very same example, for which the surrogate itself seemed accurate, but the Gaussian ISD could not permit to extract a highly accurate failure probability.

Variable ^a		Distribution	Parameters
r_w	(m)	Uniform	(0.05,0.15)
r	(m)	Lognormal	(7.71,1.0056) ^a
T_u	(m ² /year)	Uniform	(63070,115 600) ^b
H_u	(m)	Uniform	(990,1110) ^a
T_l	(m ² /year)	Uniform	(63.1,116) ^b
H_l	(m)	Uniform	(700,820) ^c
L	(m)	Uniform	(1120,1680)
K_w	(m ² /year)	Uniform	(9855,12 045) ^a

^aVariables are independent.

^bMinimum and maximum of the distribution.

^cMean and standard deviation of the natural logarithm of the distribution.

TABLE 7 Borehole-functions eight dimensional: Random variables³⁴

4.5 | Borehole-function 8D

This eight-dimensional example characterized by non-Gaussian input distributions and moderate dimensionality is taken from Reference 34. The associated benchmark function describes the water flow through a borehole:

$$v(\mathbf{y}) = \frac{2\pi T_u(H_u - H_l)}{\ln\left(\frac{r}{r_w}\right) \left(1 + \frac{T_u}{T_l} + \frac{2LT_u}{\ln\left(\frac{r}{r_w}\right)r_w^2 K_w}\right)}, \quad (44)$$

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 m³/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 transmissivity of the lower aquifer, H_l the potentiometric head of the lower aquifer, L the length of the borehole and K_w the hydraulic conductivity of the soil.

The eight independent random variables are listed in Table 7. Two quantile levels are investigated in the present work:

- *Moderate quantile level* ($\alpha \sim 10^{-4}$): The failure probability reads $\alpha = \mathbb{P}_{\mathbf{Y}}(v(\mathbf{Y}) > q) = 1.000 \times 10^{-4}$, estimated with a CoV of 0.16% using IS with a Gaussian mixture ISD³⁸ based on 10^7 samples, for $q = 260.13$. This case is studied in Reference 34.
- *Extreme quantile level* ($\alpha \sim 10^{-8}$): The failure probability reads $\alpha = \mathbb{P}_{\mathbf{Y}}(v(\mathbf{Y}) > q) = 8.732 \times 10^{-9}$, estimated with a CoV of 0.20% using IS with a Gaussian mixture ISD³⁸ based on 10^7 samples, for $q = 300$.

4.5.1 | Results

Far a sake of comparison with AK-MCS³⁴ based on both OK and PCK, we set the initial DoE size $n_0 = 12$ and use $K_q K_p = 6$ samples iteratively added with $K_q = 2$ and $K_p = 3$. The results are presented in Table 8 and Figures 9 and 10, corresponding respectively to the *moderate* and *extreme* quantile levels. In both cases, QeAK-MCS stops after few performance function calls ($\sim 24 - 35$) corresponding to \sim three to five iterations and yielding to a satisfactory final quantile estimate with an average relative error below 3.5%. The low CoV $[\hat{q}]$ below 1% indicates a high robustness/replicability of the QeAK-MCS algorithm. Observing the relative error and normalized quantile history curves (in Figures 9 and 10) permits to appreciate the fast convergence of the method: after 30 to 40 performance function calls, the relative error stops decreasing reaching a plateau above 1% in both cases, slightly higher for the *extreme level* case. It shows that the adaptive strategy performs well and it would have been possible to slightly improve performances with around one additional iteration, the refinement having stopped shortly prematurely. The average IS-based CoV associated to failure probability estimation is rather large in both cases ($\sim 7\% - 22\%$), significantly higher than in the previous test-cases. This loss of efficiency of IS based on the isotropic Gaussian is likely to be due to the moderate larger dimensionality, the nonlinear isoprobabilistic transform, and the low levels of quantile α . Interestingly, such large CoV $[\hat{\alpha}]$ have a moderate detrimental effect on the quantile estimate, contrary to failure probability estimation (see Reference 37).

TABLE 8 Results of the borehole-function eight dimensional

Method	$\mathbb{E}[N_{\text{calls}}]$	$\mathbb{E}[N_{\text{iter}}]$	$\mathbb{E}[\hat{q}]$	CoV $[\hat{q}]$	$\mathbb{E}[\epsilon_q]$	$\mathbb{E}[\text{CoV}[\hat{\alpha}]]$
Moderate quantile level $\alpha \sim 10^{-4}$						
Reference ^b	-	-	260.13	-	0	< 0.16%
QeAK-MCS ^c	24.1	3.2	261.04	0.67%	3.15%	6.91%
AK-MCS (OK) ^d	210	34	~ 260	-	<1%	<10%
AK-MCS (PCK) ^c	72	11	~ 260	-	<1%	<10%
Extreme quantile level $\alpha \sim 10^{-8}$						
Referenced	-	-	300	-	0	< 0.20%
QeAK-MCSb	34.9	4.9	300.59	0.68%	3.46%	21.81%

Reference: Importance sampling with a Gaussian mixture as importance sampling density.³⁸

^aSolving $\mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) \geq \hat{q}) = 1.000 \times 10^{-4}$, with $q_{\text{ref}} = \sigma_G = 45.69$.

^bInitial design of experiment size: 12. $K_q K_p = 6$ samples iteratively added, with $K_q = 2$ and $K_p = 3$. Based on 50 independent runs.

^cResults taken from Reference 34, based on a single run. Initial DoE size: 12. Six samples iteratively added. OK and PCK refer respectively to Ordinary kriging and polynomial chaos kriging. Note a confusion in the quantiles provided in Reference 34.

^dSolving $\mathbb{P}_{\mathbf{X}}(G(\mathbf{X}) \geq \hat{q}) = 8.732 \times 10^{-9}$, with $q_{\text{ref}} = \sigma_G = 45.69$.

FIGURE 9 Results of borehole eight dimensional, with moderate quantile level $\alpha = 10^{-4}$. Legend settings in Figure 5. A, Relative error history; B, \hat{q} history [Color figure can be viewed at wileyonlinelibrary.com]

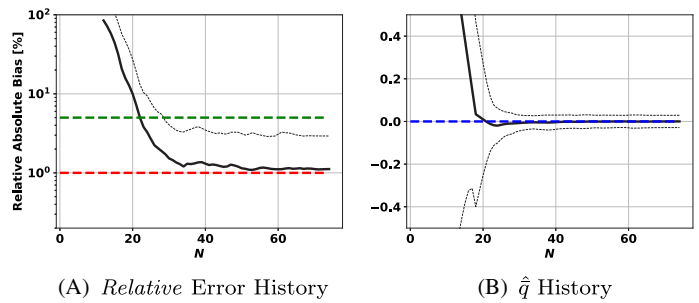
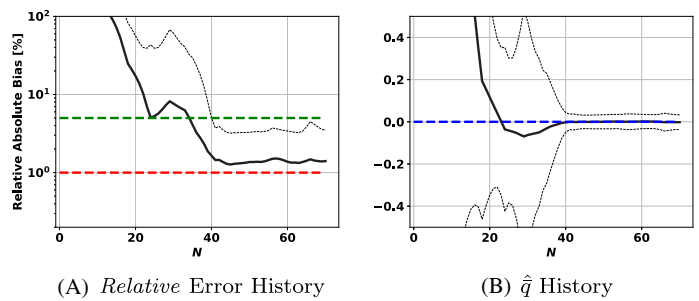


FIGURE 10 Results of borehole eight dimensional, with extreme quantile level $\alpha \sim 10^{-8}$. Legend settings in Figure 5. A, Relative error history; B, \hat{q} history [Color figure can be viewed at wileyonlinelibrary.com]



QeAK-MCS is compared to AK-MCS³⁴ (based on a single run) in the case of *Moderate quantile level*, with the same initial DoE size and considering six samples added at each refinement. Very surprisingly, QeAK-MCS significantly outperforms AK-MCS (OK and PCK) in terms of number of performance function evaluations. It seems that the multiple quantile selection and the adaptive refinement of eAK-MCS provide a significant beneficial impact on QeAK-MCS. We outline that QeAK-MCS is based on SK metamodeling, and AK-MCS (PCK) makes use of PCK which is expected to overperform SK³⁴ in terms of surrogate accuracy. This striking difference in terms of performances might be the result of different causes. The convergence criterion used in Reference 34 might be over-conservative (as a convergence history curve might permit to conclude) while we observe on Figure 9A that QeAK-MCS stopped slightly prematurely: we notice in Table 8 that QeAK-MCS provide less-accurate quantile estimates (mean relative error of 3.15%), while after around 36 performance functions calls (five iterations), the mean

relative error stabilizes slightly over 1%. Moreover, it seems that for some benchmark cases, the quantile estimation with such kind of methodology (QeAK-MCS, AK-MCS³⁴ for quantile estimation) may have a *smoothing* effect that simplifies the analysis with respect to failure probability estimation[#]. Additionally, it might be possible that there are some beneficial effects either due to the construction of the initial DoE, or the failure probability refinement strategy (eAK-MCS^{||})

5 | CONCLUSION

This paper proposes an extension of AK-MCS as presented in Reference 34 to make it suitable for the estimation of extreme quantiles. It uses a centered uncorrelated Gaussian distribution to sample candidate points and use the IS method to estimate the quantile. The refinement algorithm of eAK-MCS for failure probability estimation permits to select candidate points for the DoE, based on the quantiles estimates. A parallel quantile selection (also suitable for quantile estimation in the scope of Reference 34) is also proposed.

The performance of the proposed algorithm is assessed and illustrated through some benchmark analytical functions, showing very satisfactory accuracy and robustness, in less than 100 evaluations of the original model in average, for the considered examples. To the best of the knowledge of the authors, there is no algorithm in the literature able to estimate such extreme quantiles accurately, especially when the number of performance evaluations is limited.

The first limitation of the method is inherent to the choice of the GP-surrogate: low dimensionality and moderate size of the DoE. It also depends on its ability to fit the performance function J . The second limitation is in the physical input distribution, that needs to be accurately mapped onto the standard space. The last limitation concerns the accuracy of the IS-based quantile estimation: It strongly depends on the ability of the Gaussian ISD to reduce the variance when estimating the failure probability $\hat{\alpha}$ based on the estimated quantile. The same problem occurs even when evaluating a reference value with an analytical function based on a given small quantile level α . This IS quantile estimation is expected to deteriorate as the dimension increases, as identified in the 6D-oscillator and the 8D-borehole examples: the relative error reaches a stagnation value as the number of samples increases. Indeed, the IS CoV of α is significantly larger than in the 2D cases considered. This represents the main axis of improvement for the present method: Improving the IS-based quantile estimator accuracy, by improving the quality of the ISD.

This work also illustrates an important aspect of the eAK-MCS method, as an extension of AK-MCS for the estimation of very small failure probabilities: The possibility to adapt AK-MCS-based algorithms (eg, quantile-based optimization^{??} or quantile estimation³⁴) for such context. Here, the key of the adaptation of Reference 34 for extreme quantile is an IS-based quantile estimator from the surrogate, requiring a reasonable number of evaluations. A last contribution of this paper is the proposition of a parallel quantile selection that could also be applied for quantile estimation using the classic AK-MCS.³⁴

Future work would aim to improve the efficiency/accuracy of the IS quantile estimation, and apply the refinement step proposed in QeAK-MCS for design under uncertainty or reliability-based optimization where an extreme quantile is involved that needs to be estimated efficiently.

REFERENCES

1. Law AM, Kelton WD, Kelton WD. *Simulation Modeling and Analysis*. Vol 2. New York, NY: McGraw-Hill; 1991.
2. Glynn PW. Importance sampling for Monte Carlo estimation of quantiles. *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation*. Publishing House of St. Petersburg University; 1996: 180-185.
3. Avramidis AN, Wilson JR. Correlation-induction techniques for estimating quantiles in simulation experiments. *Oper Res*. 1998;46(4):574-591.
4. Hsu JC, Nelson BL. Control variates for quantile estimation. *Manag Sci*. 1990;36(7):835-851.
5. Hesterberg TC, Nelson BL. Control variates for probability and quantile estimation. *Manag Sci*. 1998;44(9):1295-1312.

[#] The nonlinear oscillator 6D case, studied in reliability context (eAK-MCS³⁷) and quantile estimation (Section 4.4) supports this behavior observed for some benchmark examples.

^{||} Some comparison against AK-MCSi (very similar to AK-MCS, but based on a larger MC population) in the case of the nonlinear oscillator 6D tends to support this claim.

6. Lebrun R, Dutfoy A. A generalization of the Nataf transformation to distributions with elliptical copula. *Probabilistic Eng Mech.* 2009;24(2):172-178.
7. Lebrun R, Dutfoy A. An innovating analysis of the Nataf transformation from the copula viewpoint. *Probabilistic Eng Mech.* 2009;24(3):312-320.
8. Lemaire M. *Structural Reliability*. John Wiley & Sons; 2013.
9. Serfling RJ. *Approximation theorems of mathematical statistics*. (Vol, 162) New York, NY: John Wiley & Sons; 2009.
10. Chu F, Nakayama MK. Confidence intervals for quantiles when applying variance-reduction techniques. *ACM Trans Model Comput Simul.* 2012;22(2):10.
11. Bourinet JM. *Reliability Analysis and Optimal Design Under Uncertainty-Focus on Adaptive Surrogate-Based Approaches*. Clermont, France: Université Clermont Auvergne; 2018.
12. Rubinstein RY. *Simulation and the Monte-Carlo Method*. New York, NY: Wiley; 1981.
13. Fishman GS. *Monte Carlo: Concepts, Algorithms and Applications*. New York, NY: Wiley; 1996.
14. Au SK, Beck J. Estimation of small failure probabilities in high dimensions by subset simulation. *Prob Eng Mech.* 2001;16:263-277.
15. Echard B, Gayton N, Lemaire M. AK-MCS: an active learning reliability method combining kriging and Monte Carlo simulation. *Reliab Eng Syst Saf.* 2011;33:145-154.
16. Echard B, Gayton N, Lemaire M, Relun N. A combined importance sampling and kriging reliability method for small failure probabilities with time-demanding numerical models. *Reliab Eng Syst Saf.* 2013;111:232-240.
17. Zhao H, Yue Z, Liu Y, Gao Z, Zhang Y. An efficient reliability method combining adaptive importance sampling and kriging metamodel. *Appl Math Model.* 2015;39:1853-1866.
18. Huang X, Chen J, Zhu H. Assessing small failure probabilities by AK-SS: an active learning method combining kriging and subset simulation. *Struct Saf.* 2016;59:86-95.
19. Dubourg V, Sudret B, Deheeger F. Metamodel-based importance sampling for structural reliability analysis. *Probabilistic Eng Mech.* 2013;33:47-57.
20. Cadini F, Santos F, Zio E. An improved adaptive kriging-based importance technique for sampling multiple failure regions of low probability. *Reliab Eng Syst Saf.* 2014;131:109-117.
21. Bect J, Li L, Vazquez E. Bayesian subset simulation. *SIAM/ASA J Uncertainty Quant.* 2017;5(1):762-786.
22. Bourinet JM. Rare-event probability estimation with adaptive support vector regression surrogates. *Reliab Eng Syst Saf.* 2016;150:210-221.
23. Bourinet JM, Deheeger F, Lemaire M. Assessing small failure probabilities by combined subset simulation and support vector machines. *Struct Saf.* 2011;33(6):343-353.
24. Lelièvre N, Beaupaire P, Mattrand C, Gayton N. AK-MCSi: a kriging-based method to deal with small failure probabilities and time-consuming models. *Struct Saf.* 2018;73:1-11.
25. Hristov P, DiazDelaO F, Farooq U, Kubiak K. Adaptive Gaussian process emulators for efficient reliability analysis. *Appl Math Model.* 2019;71:138-151.
26. Chen W, Xu C, Shi Y, Ma J, Lu S. A hybrid kriging-based reliability method for small failure probabilities. *Reliab Eng Syst Saf.* 2019;189:31-41.
27. Liu WS, Cheung SH, Cao WJ. An efficient surrogate-aided importance sampling framework for reliability analysis. *Adv Eng Softw.* 2019;135:102687.
28. Cui F, Ghosn M. Implementation of machine learning techniques into the subset simulation method. *Struct Saf.* 2019;79:12-25.
29. Ling C, Lu Z, Feng K, Zhang X. A coupled subset simulation and active learning kriging reliability analysis method for rare failure events. *Struct Multidiscip Optim.* 2019;60(6):2325-2341.
30. Der Kiureghian A, Dakessian T. Multiple design points in first and second-order reliability. *Struct Saf.* 1998;20(1):37-49.
31. Oakley J. Estimating percentiles of uncertain computer code outputs. *J R Stat Soc Ser C Appl Stat.* 2004;53(1):83-93.
32. Jala M, Lévy-Leduc C, Moulines É, Conil E, Wiart J. Sequential design of computer experiments for the assessment of fetus exposure to electromagnetic fields. *Technometrics.* 2016;58(1):30-42.
33. Labopin-Richard T, Picheny V. Sequential design of experiments for estimating percentiles of black-box functions. *Stat Sin.* 2018;28:853-877.
34. Schöbi R, Sudret B, Marelli S. Rare event estimation using polynomial-chaos kriging. *ASCE-ASME J Risk Uncertain Eng Syst Part A: Civil Eng.* 2016;3(2):D4016002.
35. Bichon B, McFarland J. Inverse Reliability Analysis with EGRA; 2011. <https://arc.aiaa.org/doi/abs/10.2514/6.2011-1908>.
36. Rasmussen CE, Williams C. *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press; 2006.
37. Razaaly N, Congedo PM. Extension of AK-MCS for the efficient computation of very small failure probabilities. INRIA Saclay-Equipe DEFI; 2019. RR-9286. <https://hal.inria.fr/hal-02265311>.
38. Razaaly N, Congedo PM. Novel algorithm using active metamodel learning and importance sampling: application to multiple failure regions of low probability. *J Comput Phys.* 2018;368:92-114.
39. Santner T, Williams B, Notz W. *The Design and Analysis of Computer Experiments*. New York, NY: Springer; 2013.
40. Dubourg V. Adaptive Surrogate Models for Reliability Analysis and Reliability-Based Design Optimization [PhD thesis]. Clermont, France: Université Blaise Pascal; 2011.
41. GPy. GPy: A Gaussian Process Framework in Python; Since 2012. <http://github.com/SheffieldML/GPy>.
42. Dubrule O. Cross validation of kriging in a unique neighborhood. *J Int Assoc Math Geol.* 1983;15(6):687-699.

43. Glasserman P, Heidelberger P, Shahabuddin P. Variance reduction techniques for estimating value-at-risk. *Manag Sci.* 2000;46(10):1349-1364.
44. Bect J, Ginsbourger D, Li L, Picheny V, Vazquez E. Sequential design of computer experiments for the estimation of a probability of failure. *Stat Comput.* 2012;22(3):773-793.
45. Cadini F, Gioietta A, Zio E. Improved metamodel-based importance sampling for the performance assessment of radioactive waste repositories. *Reliab Eng Syst Saf.* 2015;134:188-197.
46. Moustapha M, Sudret B, Bourinet JM, Guillaume B. Quantile-based optimization under uncertainties using adaptive kriging surrogate models. *Struct Multidiscip Optim.* 2016;54(6):1403-1421.

How to cite this article: Razaaly N, Crommelin D, Congedo PM. Efficient estimation of extreme quantiles using adaptive kriging and importance sampling. *Int J Numer Methods Eng.* 2019;1–20. <https://doi.org/10.1002/nme.6300>

APPENDIX A: CRITICAL VALUES/QUANTILES SELECTION: ADDITIONAL CRITERIA

A.1 Linear $-\alpha$ Selection

Similar linear selection is conducted, considering the surrogate-based bounds α^- , α^+ associated to the failure probability with \hat{q} as critical value, calculated using the IS estimator for failure probability (Equation (16)):

$$\alpha^- = \mathbb{P}_X(\mu_{\hat{G}}(\mathbf{X}) + k\sigma_{\hat{G}}(\mathbf{X}) < \hat{q}). \quad (\text{A1})$$

$$\alpha^+ = \mathbb{P}_X(\mu_{\hat{G}}(\mathbf{X}) - k\sigma_{\hat{G}}(\mathbf{X}) < \hat{q}). \quad (\text{A2})$$

Correspondingly, the K_q quantiles are chosen based on levels linearly selected in $[\alpha^-, \alpha^+]$. Thus:

$$\alpha_l = \alpha^- + (l - 1) \frac{\alpha^+ - \alpha^-}{K_q - 1}. \quad (\text{A3})$$

$$\alpha_l = \mathbb{P}_X(\mu_{\hat{G}}(\mathbf{X}) < \hat{q}_l), \quad \forall l \in \llbracket 1, K_q \rrbracket. \quad (\text{A4})$$

A.2 Linear $-k$ Selection

Here, a linear discretization of the range $[-k, k]$ defined by the parameter k setting the confidence interval is chosen. Defining:

$$k_l = -k + (l - 1) \frac{2k}{K_q - 1}, \quad (\text{A5})$$

$$\alpha_l = \mathbb{P}_X(\mu_{\hat{G}}(\mathbf{X}) + k_l\sigma_{\hat{G}}(\mathbf{X}) < \hat{q}), \quad \forall l \in \llbracket 1, K_q \rrbracket, \quad (\text{A6})$$

the K_q quantiles are chosen to satisfy

$$\alpha_l = \mathbb{P}_X(\mu_{\hat{G}}(\mathbf{X}) < \hat{q}_l). \quad (\text{A7})$$

In order to always include \hat{q} in the selection, the following replacement is performed once $(\hat{q}_1, \dots, \hat{q}_{K_q})$ is selected:

- if K_q is odd, $\hat{q}_{\frac{K_q+1}{2}} = \hat{q}$.
- if K_q is even, a is randomly sampled with $a \sim \mathcal{U}[0, 1]$:
 - if $a \leq \frac{1}{2}$, $\hat{q}_{\frac{K_q}{2}} = \hat{q}$.
 - if $a > \frac{1}{2}$, $\hat{q}_{\frac{K_q}{2}+1} = \hat{q}$.