RESEARCH PAPER

A new reliability method combining adaptive Kriging and active variance reduction using multiple importance sampling

Augustin Persoons[1] · Pengfei Wei[2] · Matteo Broggi[3] · Michael Beer[3]

Abstract

This article describes a new adaptive Kriging method combined with adaptive importance sampling approximating the optimal auxiliary by iteratively building a Gaussian mixture distribution. The aim is to iteratively reduce both the modeling and sampling errors simultaneously, thus avoiding limitations in cases of very rare failure events. At each iteration, a near optimal auxiliary Gaussian distribution is defined and new samples are drawn from it following the scheme of adaptive multiple importance sampling (MIS). The corresponding estimator is provided as well as its variance. A new learning function is developed as a generalization of the U learning function for MIS populations. A stopping criterion is proposed based on both the modeling error and the variance of the estimator. Results on benchmark problems show that the method exhibits very good performances on both efficiency and accuracy.

Keywords Reliability method · Adaptive Kriging · Multiple importance sampling · Extremely rare failure events · Variance reduction

1 Introduction

Industrial projects are subjected to uncertainties regarding the state of systems and their forecasted evolution. In turn, they induce uncertainties on the performances of the system and eventually the occurrence of a failure event. The probabilistic framework approaches such a problem by modeling input uncertainties as random variables defined by a joined probability distribution. Considering a black-box input–output relationship representing the system (called performance function), the output performance is simply a random variable, defined as the image of the input distributions through the performance function. Given a failure criterion associated with the random output variable(s), the failure probability is comprehensively defined as the (unknown) probability of occurrence of this event. The problem associated with the evaluation of such a probability is the central problem of structural reliability.

Computation methods for structural reliability problems stem in two main branches, the simulation and approximation methods. Sampling methods aim at estimating the failure probability by estimating the performance of a dedicated population of samples. Among the most famous simulation methods are the Monte Carlo method as well as advanced sampling methods such as: Importance sampling (Melchers 1990), Subset-simulations (Au and Beck 2001), Line sampling (Pradlwarter et al. 2005; Schuëller et al. 2004) or Directional sampling (Ditlevsen et al. 1986; Nie and Ellingwood 2000).

The approximation methods reduce the computation cost by resolving a simplified problem calibrated to minimize the approximation error, well known examples are the first and second order reliability methods (FORM and SORM) (Hasofer and Lind 1974; Lee et al. 2012; Lemaire 2013). For the interested readers, a general overview and discussion on FORM/SORM and their variants can be found in the work of Rackwitz (2001). The surrogate modeling approach is a type of approximation aiming at calibrating an emulator of the system which predicts its behavior without

✉ Augustin Persoons
augustin.persoons@kuleuven.be

1 KU Leuven, Jan De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium

2 School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University, Xi'an 710129, China

3 Institute for Risk and Reliability, Leibniz University Hannover, Callinstr. 34, Hannover, Germany

simulating the physics, and thus with only a fraction of the computation burden. The calibration of a surrogate model requires the definition and evaluation of a calibration dataset often called design of experiment (DoE). The advantage of this approach is the ability to estimate the performance of any population with the expense of evaluating the physical model only for the samples of the DoE. This general framework regroups many types of surrogate strategies, among the most popular are: support vector machines (Cortes and Vapnik 1995), polynomial chaos (Crestaux et al. 2009; Sudret 2008), Neural networks (Hurtado and Alvarez 2001) and Kriging (Matheron 1973; Bichon et al. 2008). In the context of structural reliability, Kriging-based methods have become increasingly popular in the past decade. This popularity can be related to one of their most appealing features, i.e., the estimation of their own localized prediction uncertainty in the shape of the posterior standard deviation. It allowed the development of the so-called adaptive Kriging approach which combine a sampling reliability method with an adaptive scheme to iteratively improve the calibration of the surrogate.

In an adaptive Kriging scheme both the sampling method and the surrogate induce an error, respectively a sampling and modeling error. The main goal is then to iteratively build a near-optimal DoE minimizing the modeling error. Almost all well-known reliability method has been combined with adaptive Kriging and the Literature of existing methods is very extensive. Among the most successful are Markov Chain Monte Carlo (MCMC, see Zhao et al. 2015; Song et al. 2019; Wei et al. 2019) and importance sampling. This overview is limited to the most common sampling-based strategies with a focus on importance-sampling approaches. For a wider overview of existing methods the reader can refer to Sudret (2012), Teixeira et al. (2021) and Moustapha et al. (2021). Most method that fall in that family minimize the modeling and sampling errors independently, for example by generating beforehand a suitable population, assumed to carry a negligible variance of the estimator. Well known examples of such methods are AK-MCS (Echard et al. 2011) or AK-IS (Echard et al. 2013).

Fewer methods propose to actively reduce both errors simultaneously. Dubourg et al. (2013) proposed to use the Kriging model to approximate the optimal auxiliary density, using MCMC sampling and *K*-means clustering in method called with Meta-IS. In a related method called metaAK-IS[2] Cadini et al. (2014), proposed to combine the work of Echard et al. (2013) and Dubourg et al. (2013). Several of these methods have in common to iteratively approximate the optimal auxiliary importance sampling distribution using the Kriging model. Balesdent et al. (2013) proposed to do so using the NAIS method (Morio 2011), Liu et al. (2020) used the adaptive-linked importance sampling method, Razaaly and Congedo (2018) using a Gaussian mixture

calibrated using *K*-mean clustering, Yang et al. (2018) with kernel-density-estimation or Zhang and Taflanidis (2018) using intermediate condition distribution similar to subset simulation. Those methods usually then require advanced sampling strategies such as MCMC to draw sample from the auxiliary. Other related work combining adaptive Kriging with advanced importance sampling include Xiong and Tan (2018), Yun et al. (2018), Zhao et al. (2015), and Zhang et al. (2020).

In this work a related method is proposed where the optimal importance sampling distribution is iteratively approximated with a Gaussian mixture distribution using a scheme that can be related to the adaptive importance sampling (AMIS) and thus avoiding the limitations associated with advanced sampling techniques. The Gaussian mixture is refined by iteratively adding new normal distributions defined as a near-optimal from a variance-reduction point of view. The population is also built iteratively by simply drawing and adding samples from the last Gaussian distribution. For convenience, the proposed method will be referred to as AK-AMIS for adaptive-Kriging and adaptive multiple importance sampling (MIS).

The paper starts describing the method with the failure probability estimator in Sect. 2. A dedicated learning function is developed followed by the description of the iterative auxiliary distribution, defined as a near-optimal from a variance reduction perspective. Two stopping parameters are proposed based respectively on the expected error and the expected variance of the estimator and are combined in a single stopping criterion. The method is then validated in Sect. 3 through three academic benchmark problems and an industrial one highlighting its performances and limitations. The results are discussed, and the performances of AK-AMIS are compared to well established and closely related adaptive reliability methods. The paper finally concludes in Sect. 4 with some lead to further improve the performances.

2 The AK-AMIS method

One limitation of most sampling-based adaptive Kriging approaches (such as AK-MCS or AK-IS) is that they rely on a given population of samples. The method then adaptively minimizes the surrogate modeling error associated with the estimator on that given population. Considering that surrogate approaches combined with sampling techniques induce two types of error, respectively the modeling and sampling error, the simple adaptive Kriging scheme focuses on minimizing only the first. The second is considered given and comparatively negligible if the population is large enough. However, the validity of this hypothesis is usually impossible to evaluate beforehand.

One way to tackle this limitation is, instead of reasoning with a given population and distribution, to iteratively refine them according to the kriging surrogate. Several methods discussed in the previous section proposed schemes to do so, focusing on approximating the optimal distribution and then sampling with advanced sample generation techniques. The approaches described in this work is a new one falling into that category, where the optimal distribution is approximated with a Gaussian mixture and the population is generated simultaneously. The auxiliary distribution is built iteratively, starting from the standard Gaussian distribution at the first iteration and adding a new one every iteration. The method then relies on iteratively selecting both the best sample to enrich the DoE and the best Gaussian distribution to enrich the auxiliary mixture distribution.

The Gaussian mixture auxiliary requires to use a dedicated estimator of the failure probability, see Sect. 2.1. The sequential DoE relies on the definition of a dedicated learning function and is discussed in Sect. 2.2. The enrichment of the population is based on the scheme of the AMIS as proposed by Cornuet et al. (2012), see Sect. 2.3. Finally, each part can be associated with a stopping criterion, two are proposed and briefly discussed in Sect. 2.4.

2.1 Multi-IS estimator

In the proposed approach, the auxiliary population is adaptively refined, meaning that the population consists of an aggregate of sub-populations each from a different auxiliary distribution. The failure probability could be estimated with each sub-population directly using the IS estimator. If the auxiliary distributions are strictly identical the overall estimator would be the average of the individual IS estimators and if the distributions are strictly non-overlapping it would be the sum. However, in the general case, the MIS estimator requires a dedicated definition. Such an estimator has been proposed by Veach and Guibas (1995) and popularized by Owen and Zhou (2000). The estimator used in AK-AMIS and detailed in this section is a particular case the first and is referred to as the *multi-sample model* (with the *balance heuristic*) in the field of computer graphics (Víctor Elvira et al. 2019; Sbert et al. 2016). Notice that in the context of AMIS, the set of auxiliary distribution is built sequentially. Therefore, each time a set of new samples is generated, the weight all samples should be updated, not only the weight of the new samples. The procedure could thus be seen as a single importance sampling where the auxiliary population is drawn from a mixture made of the sum of the auxiliary distributions (Cornuet et al. 2012).

Consider a d-dimensional input vector $\mathbf{X}$ with a joint probability distribution function f. The reliability problem can be formulated as the estimation of the probability

$P_f = P(g(\mathbf{X}) \leq 0)$ with $g : \mathbb{R}^d \to \mathbb{R}$ the performance function of the system such that a negative value indicates the failure event. After k iteration(s), the population S_k is composed of $\sum_{i=1}^{k} n_i \neq 0$ samples, each set of n_i samples are drawn from an auxiliary distribution $h_i : \mathbb{R}^d \to \mathbb{R}^{*+}$. The failure probability can then be defined as follows:

$$
\begin{aligned}
P_f &= \int_{\mathbb{R}^d} I(X) f(X) \mathrm{d}X \\
&= \int_{\mathbb{R}^d} I(X) f(X) \frac{\sum_{i=1}^{k} n_i h_i(X)}{\sum_{p=1}^{k} n_p h_p(X)} \mathrm{d}X \\
&= \sum_{i=1}^{k} n_i \int_{\mathbb{R}^d} I(X) \frac{f(X)}{\sum_{p=1}^{k} n_p h_p(X)} h_i(X) \mathrm{d}X
\end{aligned}
\tag{1}
$$

with I the classification function such that $I(X) = 1$ if $g(X) \leq 0$, i.e., X is failing and $I(X) = 0$ otherwise. Each integral can be approximated by a sum using the Monte Carlo estimator to define the following MIS estimator:

$$
\begin{aligned}
\widehat{P}^{\mathrm{MIS}} &= \frac{1}{N_k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} I(X_{j,i}) \frac{f(X_{j,i})}{\frac{1}{N_k} \sum_{p=1}^{k} n_p h_p(X_{j,i})} \\
&= \frac{1}{N_k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} I(X_{j,i}) \frac{f(X_{j,i})}{w_k(X_{j,i})}
\end{aligned}
\tag{2}
$$

with $X_{j,i}$ the jth sample drawn for the ith distribution, $N_k = \sum_{i=1}^{k} n_i$ the total number of samples in the population and $w_k(X) = \frac{1}{N_k} \sum_{i=1}^{k} n_i h_i(X)$ the mixture of the auxiliary distributions. This latest expression can be simplified further into the classical IS estimator by noticing that the aggregated population is equivalent to a population of the same size directly drawn from the mixture w_k (see, e.g., Cornuet et al. 2012; Victor Elvira et al. 2015):

$$
\widehat{P}^{\mathrm{MIS}} = \frac{1}{N_k} \sum_{i=1}^{N_k} I(X_i) \frac{f(X_i)}{w_k(X_i)}.
\tag{3}
$$

This formulation holds for any set of (non-zero) auxiliary distribution h_i. In the context of the AK-AMIS method, we will adopt the simple scheme of the mean-shift importance sampling and only consider Gaussian distributions with unit standard deviation such that $h_k = \varphi(\mathrm{M}_k, 1)$ with φ the probability distribution function of the standard normal distribution. In addition, the set of auxiliary distributions starts with the actual joint distribution of the input parameters, i.e., $h_1 = f$.

The variance of this estimator is defined as follows:

$$\text{var}\left(\widehat{P}^{\text{MIS}}\right) = \frac{1}{N}\text{var}\left(I(X)\frac{f(X)}{w_k(X)}\right) \tag{4}$$

thus:

$$\text{var}\left(\widehat{P}^{\text{MIS}}\right) = \frac{1}{N_k}\left(E\left(I^2(X)\frac{f^2(X)}{w_k^2(X)}\right) - E^2\left(I(X)\frac{f(X)}{w_k(X)}\right)\right)$$

$$= \frac{1}{N_k}\left(E\left(I^2(X)\frac{f^2(X)}{w_k^2(X)}\right) - P_f^2\right). \tag{5}$$

2.2 Definition of the learning function

In the context of reliability analysis, the only modeling errors that have an influence on the failure probability estimator are the misclassification of samples. The posterior probability of misclassification $\Phi\left(\frac{-|\widehat{\mu}(X)|}{\widehat{\sigma}(X)}\right)$ is a good indicator of the individual contribution of each sample to the overall modeling error, which is base principle behind U learning function of the AK-MCS algorithm.

In the framework of the AK-AMIS method, the learning function should also reflect that each sample is associated with a weight that affects its contribution to the estimated failure probability. It is thus proposed to consider another related quantity of interest which is more suited for the MIS population previously described: the expected absolute failure probability error induced by the Kriging misclassifications: $E\left(\left|\Delta P_f\right|\right)$.

In this section, we are only interested in reducing the modeling error, let us then consider a given population $S_k = \left\{X_i, i = [1, \dots, N_k], N_k = \sum_{i=1}^k n_i\right\}$. For each sample, the Kriging-based classification function $\widehat{I}(X_i)$ is considered as a random variable holding the uncertainty related to the Kriging prediction, i.e.:

$$\forall i \in [1, N_k] \begin{cases} P\left(\widehat{I}(X_i) = 1\right) = \Phi\left(\frac{-\widehat{\mu}(X_i)}{\widehat{\sigma}(X_i)}\right), \\ P\left(\widehat{I}(X_i) = 0\right) = 1 - P\left(\widehat{I}(X_i) = 1\right). \end{cases} \tag{6}$$

The expected absolute prediction error can then be developed as follows:

$$E\left(\left|\Delta P_f\right|\right) = E\left(\left|\widehat{P}^{\text{MIS}} - \widehat{P}_{\text{krig}}^{\text{MIS}}\right|\right)$$

$$= E\left(\left|\frac{1}{N_k}\sum_{i=1}^N \frac{f(X_i)}{w_k(X_i)}\left(I(X_i) - \widehat{I}(X_i)\right)\right|\right) \tag{7}$$

with $\widehat{P}_{\text{krig}}^{\text{MIS}}$ the MIS estimator associated with the Kriging surrogate classification function. Notice that since the population is given, the Kriging posterior is the only random variable in Eq. (7). Unfortunately, this quantity does not allow us to capture the contribution of each individual sample to the overall error, which is the goal of the learning function. It is therefore proposed to consider instead the following upper bound of the absolute error:

$$E\left(\left|\Delta P_f\right|\right) \le E\left(\frac{1}{N_k}\sum_{i=1}^N \frac{f(X_i)}{w_k(X_i)}\left|I(X_i) - \widehat{I}(X_i)\right|\right)$$

$$= \frac{1}{N_k}\sum_{i=1}^N \frac{f(X_i)}{w_k(X_i)}E\left(\left|I(X_i) - \widehat{I}(X_i)\right|\right)$$

$$= \frac{1}{N_k}\sum_{i=1}^N \frac{f(X_i)}{w_k(X_i)}\Phi\left(\frac{-\left|\widehat{\mu}(X_{j,i})\right|}{\widehat{\sigma}(X_{j,i})}\right) \tag{8}$$

$$= \frac{1}{N_k}\sum_{i=1}^N L_k(X_i)$$

with $L_k(X) = \frac{f(X)}{w_k(X)}\Phi\left(\frac{-|\widehat{\mu}(X)|}{\widehat{\sigma}(X)}\right)$ the proposed learning function for the AK-AMIS algorithm.

This expression is however prone to overly prioritize areas of small posterior variance. This behavior is due to the observed tendency of Kriging models to underestimate (respectively overestimate) the variance in regions of high (respectively low) variance in noise-free problems. In methods considering a given population this behavior doesn't affect the performances significantly since the method eventually eliminates all affected samples. However, in the proposed approach, new samples are iteratively added which could lead the algorithm to focus only on areas already well calibrated (exploitation) and ignore areas poorly calibrated (exploration). The solution chosen to counteract this behavior is to introduce a hyperparameter p which is simply a power affected to the posterior standard deviation in the expression of the learning function:

$$L_k^p(X) = \frac{f(X)}{w_k(X)}\Phi\left(\frac{-|\widehat{\mu}(X)|}{\widehat{\sigma}^p(X)}\right). \tag{9}$$

From the authors experience, the value $p = 3$ yields satisfactory results and is the value used in the applications presented in Sect. 3. As any hyperparameter unrelated to interpretable metrics, the choice of p is at least partly arbitrary and any hypothetical optimal value would be problem dependent. The choice proposed in this article is based on the limited empirical knowledge available to the authors, a short presentation of them and their interpretation is proposed in Appendix 1.

The best candidate for enriching the DoE X^* is then defined as the sample satisfying:

$$X_k^* = \underset{X \in S_k}{\text{argmax}}\left(L_k^p(X)\right). \tag{10}$$

2.3 Adaptive multiple importance sampling strategy

As previously mentioned, this algorithm approaches the optimal IS auxiliary by iteratively building a Gaussian mixture, however the proposed framework is presented in a broader context and would apply for a mixture of any distribution. The goal is to incrementally reduce the variance of the MIS estimator as described in Eq. (5). The purpose of this section is thus to find a way to iteratively define a set of near-optimal auxiliary distributions from a variance reduction perspective without additional calls to the performance function. The proposed AK-AMIS method uses the general scheme described in Cornuet et al. (2012) which reflects in the way the estimator is updated as described in Sect. 2.1. However, the sequential definition of the auxiliary distribution presented in this section differs from the AMIS method (Cornuet et al. 2012) and instead of minimizing a statistical distance metrics, it is proposed to directly minimize the variance of the estimator associated with the Kriging surrogate.

Considering k previously computed iterations, the goal is to estimate the next near-optimal auxiliary distribution h_{k+1}. Based on the expression of the variance as given in Eq. (5), the optimal mixture w_{k+1} is reached when

$$\operatorname{var}\left(\widehat{P}^{\text{MIS}}\right) = \frac{1}{N_{k+1}}\left(E\left(I^2(X)\frac{f^2(X)}{w_{k+1}^2(X)}\right) - P_f^2\right) = 0 \quad (11)$$

thus

$$E\left(\frac{I^2(X)}{P_f^2}\frac{f^2(X)}{w_{k+1}^2(X)}\right) = 1. \quad (12)$$

One solution for an optimal w_{k+1} is thus (see, e.g., Morio 2011):

$$w_{k+1}(X) = \frac{I(X)f(X)}{P_f}. \quad (13)$$

By definition of the mixture distribution, we have:

$$\frac{1}{N_{k+1}}\sum_{i=1}^{k+1} n_i h_i(X) = \frac{I(X)f(X)}{P_f} \quad (14)$$

thus, the optimal new auxiliary h_{k+1}^* can be defined as:

$$h_{k+1}^*(X) = \frac{N_{k+1}}{n_{k+1}}\frac{I(X)f(X)}{P_f} - \sum_{i=1}^{k}\frac{n_i}{n_{k+1}}h_i(X). \quad (15)$$

Or

$$h_{k+1}^*(X) = \frac{N_k + n_{k+1}}{n_{k+1}}\frac{I(X)f(X)}{P_f} - \frac{N_k}{n_{k+1}}w_k(X). \quad (16)$$

Defining such an optimal distribution faces the same issue as the classical importance sampling method, which is that neither P_f nor I are known. It is then proposed to chose h_{k+1}^* as a near-optimal solution in a similar way as the classical mean-shift importance sampling, i.e., a normal distribution with unit standard deviation and mean M_{k+1} such that:

$$M_{k+1}^{\text{opt}} = \underset{X \in \mathbb{R}^d}{\arg\max}\left(\frac{N_k + n_{k+1}}{n_{k+1}}\frac{I(X)f(X)}{P_f} - \frac{N_k}{n_{k+1}}w_k(X)\right). \quad (17)$$

Considering that w_k, N_k and n_{k+1} are strictly positive, we reach the following equivalent expression:

$$M_{k+1}^{\text{opt}} = \underset{X \in \mathbb{R}^d}{\arg\max}\left(\frac{I(X)f(X)}{P_f w_k(X)}\right). \quad (18)$$

In order to avoid extra evaluations of the performance function, the classification function and failure probability are replaced by their surrogate counterpart:

$$M_{k+1}^{\text{opt}} = \underset{X \in \mathbb{R}^d}{\arg\max}\left(\frac{\widehat{I}(X)f(X)}{\widehat{P}_{\text{krig}}^{\text{MIS}} w_k(X)}\right). \quad (19)$$

However, doing so induces a major issue that needs to be addressed: because of the behavior of $\widehat{I}$, all samples outside of the failure domain have a zero-value and cannot be sorted. Therefore, at the initial stage, when the Kriging model is badly calibrated, $\widehat{I}$ may be equivalent to the null function over $\mathbb{R}^d$ thus preventing the algorithm from converging. To overcome this limitation, the classification function is replaced by the function $\Phi\left(\frac{-\widehat{\mu}(X)}{\widehat{\sigma}(X)}\right)$, which can be interpreted as the probability, according to the Kriging posterior, that the sample falls in the failure domain. In this case, it acts as a sigmoid-like function following the trend of the classification function while providing a sortable output value. A related issue occurs with $\widehat{P}_{\text{krig}}^{\text{MIS}}$ which may be equal to zero at initial stages, the same strategy of replacing $\widehat{I}$ with $\Phi\left(\frac{-\widehat{\mu}(X)}{\widehat{\sigma}(X)}\right)$ is therefore used in the $\widehat{P}_{\text{krig}}^{\text{MIS}}$ estimator. M_{k+1} is finally defined by solving the optimization problem defined in Eq. (20).

$$M_{k+1}^{\text{opt}} = \underset{X \in \mathbb{R}^d}{\arg\max}\left(\frac{\Phi\left(\frac{-\widehat{\mu}(X)}{\widehat{\sigma}(X)}\right)f(X)\frac{1}{w_k(X)}}{\sum_{i=1}^{N}\Phi\left(\frac{-\widehat{\mu}(X_i)}{\widehat{\sigma}(X_i)}\right)f(X_i)\frac{1}{w_k(X_i)}}\right). \quad (20)$$

It is further proposed to choose the last sample added to the DoE (X^*) as the starting point of the optimization algorithm. The reason being that, because of the

computation error, the term $\Phi\left(\frac{-\hat{\mu}(X)}{\hat{\sigma}(X)}\right)$ may be equal to zero for all samples in areas that are both far from the limit-state and well calibrated. The optimization process might then be unable to start if the initial point of the optimization algorithm is chosen in such an area. By the definition of the learning function, X^* is a good starting point candidate to avoid this issue. In practice the optimization problem is solved using a global optimization strategy and the sequential quadratic programming algorithm.

The behavior of $\Phi\left(\frac{-\hat{\mu}(X)}{\hat{\sigma}(X)}\right)$ is illustrated in Fig. 1, in the context of application of the algorithm to the 4-modes benchmark problem presented in Sect. 3.1, for the early and late stages of the algorithm. Choosing this function instead of an actual sigmoid function or any other analytical equivalent allows us to avoid defining additional hyper-parameters.

Finally, it is necessary to define the number of new samples drawn from the last auxiliary distribution: n_{k+1}. The development presented in this paper did not focus on defining an adaptive population size for the iterative enrichment. It is proposed to use the simple scheme of iteratively adding a fixed number of new samples $n_1 = \cdots = n_{k+1} = n$. The choice of this parameter has an influence on the performance of the method, but it is found that this influence quickly decreases as the number of samples increases. Choosing a relatively low number of samples is more conservative since it ensures the variance-based stopping criterion (see Sect. 2.4) does not converge prematurely but can needlessly delay the termination of the algorithm. Increasing this number may in turn result in a significant computation time especially for the calibration of the Kriging model and the evaluation of the estimator. A few thousands of samples per dimension seems a good compromise for all the problems the method has been tested on.

2.4 Two stopping criteria

The enrichment of the DoE and the population are aimed at respectively minimizing the modeling error and the variance of the estimator. It is thus proposed to define a stopping criterion for each process and stop the algorithm when both are satisfied simultaneously.

The first criterion is directly related to the learning function. The most intuitive way of introducing such a criterion would be to set a threshold L_{stop} for the learning function and consider that the algorithm converged if $L(X^*) \leq L_{\text{stop}}$. However, the convergence of this parameter is usually quite noisy and may lead to premature termination. Therefore, the proposed criterion uses another related and more stable quantity: $E\left(\left|\Delta P_f\right|\right)$, as defined in Eq. (8). The first stopping criterion can then be rephased as follows:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} L_k^1(X_{j,i}) \leq E_{\text{stop}} \tag{21}$$

with E_{stop} a stopping threshold that can then be adjusted to a desired level of precision. This criterion is however affected by the inaccuracies mentioned in Sect. 2.2 associated with regions of high and low posterior variance. It is therefore proposed to use the conservative approach of evaluating both the expected error related to the unaltered Kriging surrogate and considering the power hyperparameters. The error-based criterion is then only satisfied when both errors are below the threshold:

$$E\left(\left|\Delta P_f\right|\right) = \max\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} L_k^1(X_{j,i}), \sum_{i=1}^{k}\sum_{j=1}^{n_i} L_k^p(X_{j,i})\right) \leq E_{\text{stop}}. \tag{22}$$

The second termination criterion is based on the variance of the estimator as given by Eq. (5). However, since

Fig. 1 Illustration of the behavior of $\Phi\left(\frac{-\hat{\mu}(X)}{\hat{\sigma}(X)}\right)$ compared to the classification function on the 4-modes benchmark problem at **a** early and **b** late stage of the AK-AMIS algorithm
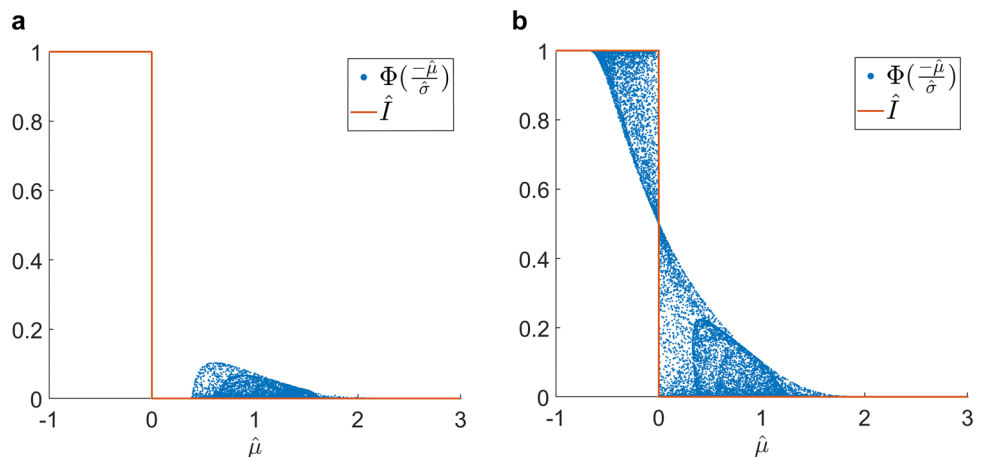
the classification function I is unknown is it replaced by the Kriging surrogate $\hat{I}$:

$$\widehat{\text{var}}\left(\hat{P}^{\text{MIS}}\right) = \frac{1}{N_k}\left(E\left(\hat{I}^2(\boldsymbol{X})\frac{f^2(\boldsymbol{X})}{w_k^2(\boldsymbol{X})}\right) - \left(\hat{P}^{\text{MIS}}_{\text{krig}}\right)^2\right). \quad (23)$$

The related stopping criterion is then simply satisfied when the corresponding coefficient of variation crosses a threshold value CoV_{stop} corresponding to a desired level of confidence, i.e.

$$\frac{\sqrt{\widehat{\text{var}}(\hat{P}^{\text{MIS}}_k)}}{\hat{P}^{\text{MIS}}_{\text{krig}}} \leq \text{CoV}_{\text{stop}}. \quad (24)$$

Using both criteria simultaneously provides a much more robust termination of the algorithm than using them individually. The reason is that unexpected termination of the error-based criterion can occur when very few samples exist in badly calibrated regions. While premature termination of the variance-based criterion requires the Kriging surrogate to significantly underestimate the actual variance associated with the given population. Since the variance is estimated empirically this event requires that a significant part of the population falls in badly calibrated areas. These events are clearly mutually exclusive, therefore, if one is fooled the other one has to "work as intended". Since these events are expected to occur only at early stages when the Kriging model is very badly calibrated, we can expect the later to acknowledge either that the modeling error is too high or that the population is too scarce to terminate. With this stopping strategy the complete flowchart of the algorithm can be drawn as illustrated in Fig. 2.
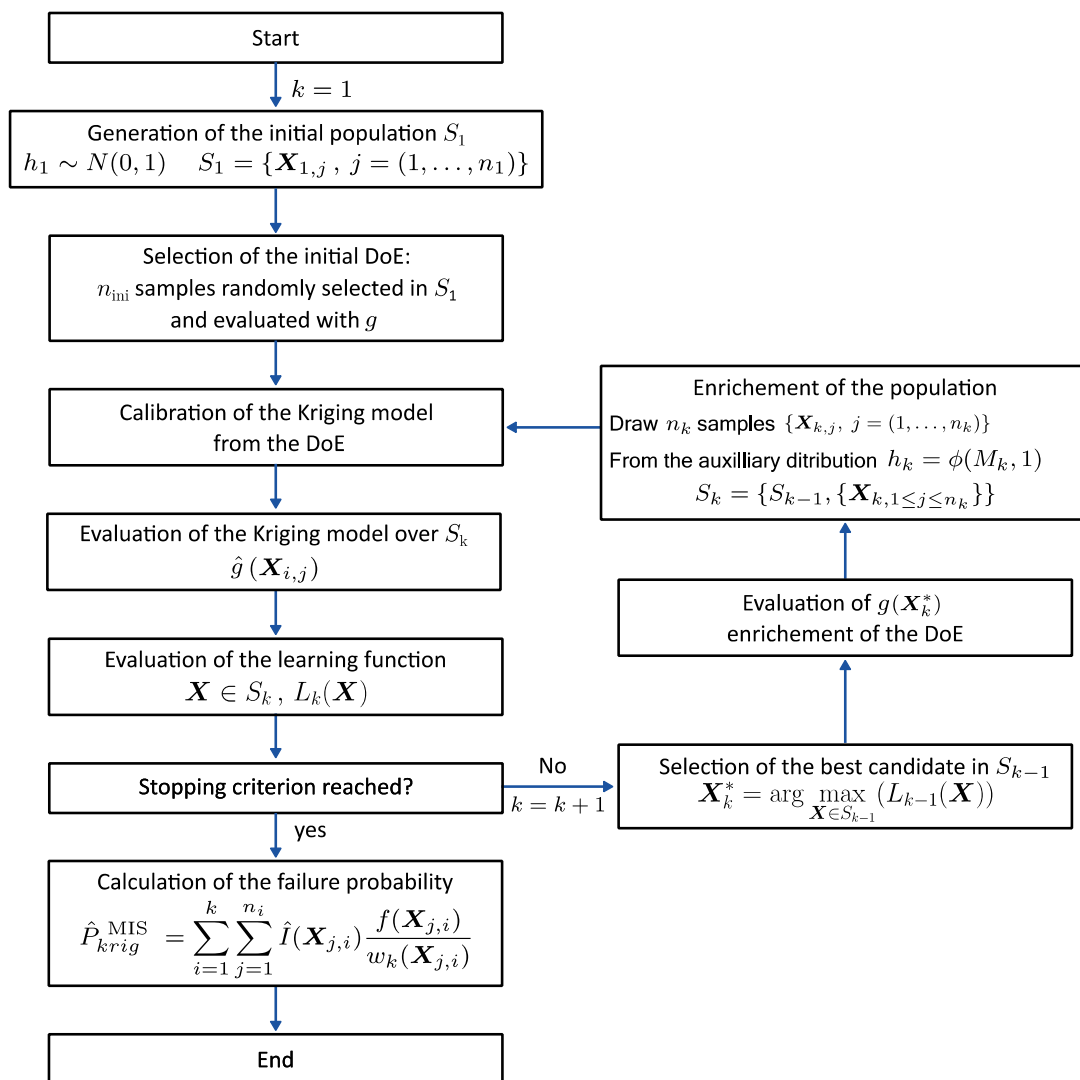

Fig. 2 Flowchart of the AK-AMIS procedure

As noticed by Wang et al. (2019) estimators directly based on the Kriging model as the one proposed above usually do not directly correlate with the quantity they are intended to measure. This is the reason why they are associated with very conservative thresholds. Even though this has not been the scope of this work, more sophisticated stopping criteria such as the ESC (Wang and Shafieezadeh 2019; Yi et al. 2020) may improve significantly the robustness of the proposed method.

3 Applications

For the validation of the proposed approach the results will be compared with related methods from the literature namely AK-MCS (Echard et al. 2011) or AK-IS (Echard et al. 2013) when applicable, meta-AK-IS^2 (Cadini et al. 2014) and MetaAL-OIS (Razaaly and Congedo 2018). The first three are chosen because their efficiency have been thoroughly proven, meta-AK-IS^2 being one of the most popular method combining Kriging surrogate with adaptive importance sampling. The last is selected because it is more recent, exhibits challenging performances on the proposed examples and it is closely related to the proposed approach. In addition to the main comparative study, additional performance comparison are provided in Appendix 2 considering two other reference methods form the literature: AK-ALIS (Liu et al. 2020) and AKOIS (Zhang et al. 2020).

The performances are then compared over four test cases, three are challenging analytical examples which are standard in the literature of comparable methods. They are also the three most challenging of the five examples used for the performance validation of MetaAL-OIS, allowing a fair comparison between the two methods. These analytical examples are also included in the generalized benchmark for active learning from Moustapha et al. (2021), which aims at providing a broad overview of the performances of adaptive-reliability methods and facilitating performance comparison in the literature. The last example is an engineering problem in the context of radioactive waste storage and associated with extremely rare failure event.

The number of new samples drawn per iteration n was set equal to 10^4. The stopping criterion is as discussed in Sect. 2.4 with $E_{stop} = 5\%$, and $CoV_{stop} = 0.4\%$. It is assumed that no primary knowledge is available regarding the behavior of the performance functions and the deterministic part of the Kriging model is defined as a constant function. The covariance function chosen is anisotropic squared exponential. Since the population grows iteratively, its size can become an issue for the most challenging cases (e.g., requiring several hundred of iterations) and can significantly slow down the evaluation of the Kriging model. Therefore, the population is capped to 10^6 samples per dimensions in the following application, once the cap is reached only the DoE keeps being refined.

In all the following applications, the initial DoE n_{ini} is defined using a Latin hypercube design of size equal to the dimension of the problem plus one. As the following section aims at demonstrating, the proposed approach is very efficient at exploring the input space and does not require an extensive initial training set. A sparse initial training set is therefore chosen to illustrate that advantage.

The results presented are the number of calls to the performance function (number of iterations + number of samples in the initial DoE), the estimated failure probability, the corresponding relative error and the coefficient of variation of the estimator. The results are given for the AK-AMIS method in terms of average and extrema over 100 applications of the algorithm, with the relative error considering the Monte Carlo estimation as a reference.

3.1 A four failure modes function

A well-known 2-D series system with four failure modes specifically designed to challenge methods based on location a single "design point". The performance function is the following (Razaaly and Congedo 2018):

$$g(X) = \min\big(g_1(X), g_2(X), g_3(X), g_4(X)\big),$$

$$\begin{cases} g_1(X) = k_1 + \frac{(x_1-x_2)^2}{10} - \frac{1}{\sqrt{2}}(x_1 - x_2), \\ g_2(X) = k_1 + \frac{(x_1-x_2)^2}{10} + \frac{1}{\sqrt{2}}(x_1 - x_2), \\ g_3(X) = \frac{k_2}{\sqrt{2}} + x_1 - x_2, \\ g_4(X) = \frac{k_2}{\sqrt{2}} - x_1 + x_2, \end{cases} \tag{25}$$

with x_1 and x_2 independent standard Gaussian random variables and k_1, k_2 parameters affecting the failure probability. Two cases are considered to observe how the algorithm performances scale with the failure probability; the first ($k_1 = 3$, $k_2 = 7$) is associated with reasonable failure probability and the second ($k_1 = 5.5$, $k_2 = 11$) induces extremely low failure probability). The first case has been extensively studied in the literature and the comparison with reference methods will be conducted on this case. The performances of the chosen reference methods have not been published for the second case and is therefore only used as an illustration of performance evolution of the proposed method.

Figure 3 illustrates how the surrogate approximate the limit-state on a case of multiple failure regions on both cases. It can be observed how the algorithm efficiently spends more iterations refining areas based on how significant they are for the overall failure probability. This behavior leads to observable inaccuracies in areas of low interest but
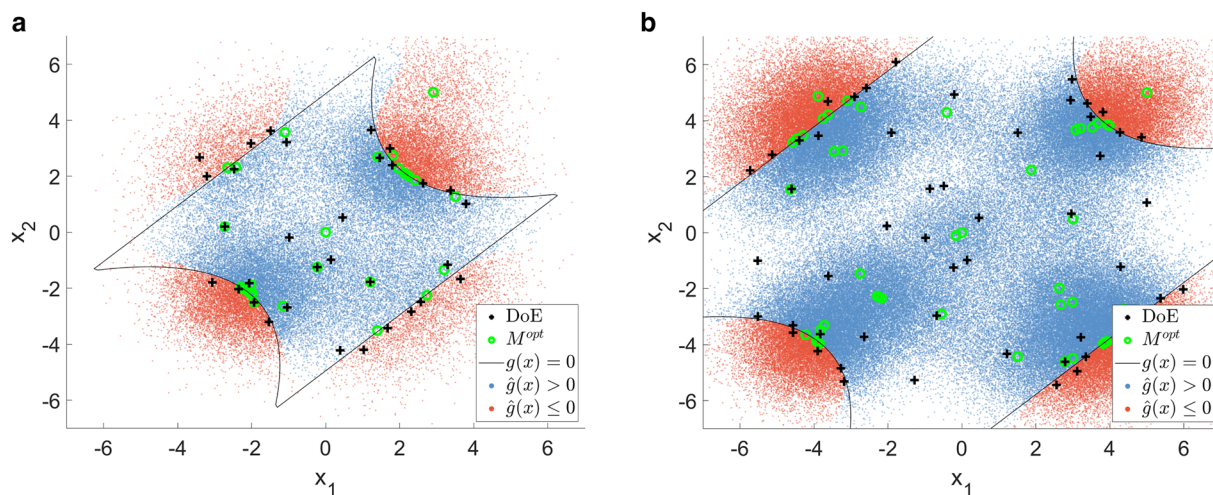
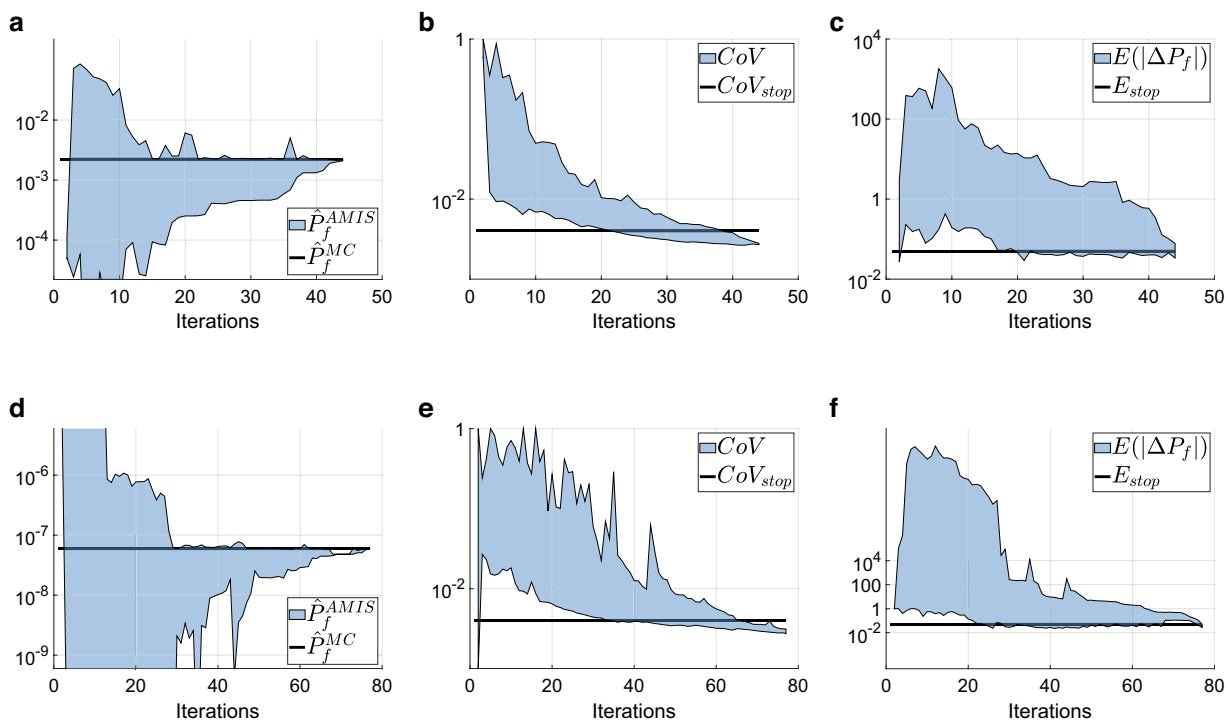**Fig. 3** Scatter plot of the limit-state, DoE and population on **a** case 1 and **b** case 2 of application 1

Fig. 4 Convergence figure for application 1 case 1 and 2 of **a**, **d** the failure probability, **b**, **e** the CoV stopping parameter, and **c**, **f** the error stopping parameter. **d** Histogram of the number of calls to the performance function

represents a good compromise allowing the algorithm to quickly terminate while reaching the desired precision level.

Figure 4 illustrates the convergence of the algorithm with respect to the failure probability, the error-based criterion and the variance-base criterion. The modeling error seem to be the limiting criterion on this example, with the variance-based one being satisfied much earlier. These results can be related to the complicated shape of the limit-state and the fact that the failure probability is not very low. In addition, we can observe that both sampling strategies efficiently and quickly converge toward the correct failure probability. This figure also illustrates the necessity of using both of the proposed stopping criteria with the one considering only the covariance of the estimator leading to premature termination.

Table 1 Reliability results for application 1

Method	MCS	AK-MCS[a]	Meta AK-IS2[a]	MetaAL-OIS[a]	AK-AMIS	
					Averaged	Amplitude
Case 1: $k_1 = 3, k_2 = 7$						
N_{calls}	2×10^6	96	138	69	36.7	$[26, 49]$
$\widehat{P}_f$	2.22×10^{-3}	2.23×10^{-3}	2.22×10^{-3}	2.21×10^{-3}	2.15×10^{-3}	$[1.85, 2.54] \times 10^{-3}$
CoV(%)	1.51				3.12	$[2.63, 3.96]$
$\Delta P_f(\%)$	–		1.7	0.10	4.06	$[1.48 \times 10^{-1}, 17.5]$
Case 2: $k_1 = 5.5, k_2 = 11$						
N_{calls}	2×10^{10}				52	$[38, 80]$
$\widehat{P}_f$	5.95×10^{-8}				5.83×10^{-8}	$[4.79, 6.24] \times 10^{-8}$
CoV(%)	2.91				3.69×10^{-1}	$[2.76, 3.99] \times 10^{-1}$
$\Delta P_f(\%)$	–				2.45	$[2.05 \times 10^{-2}, 19.5]$

[a]Results from Table 4 in Razaaly and Congedo (2018)

Fig. 5 Histogram of **a**, resp. **c** the number of calls to the performance function and **b**, resp. **d** the absolute relative error over the 100 applications of the algorithm for application 1 case 1, resp. 2
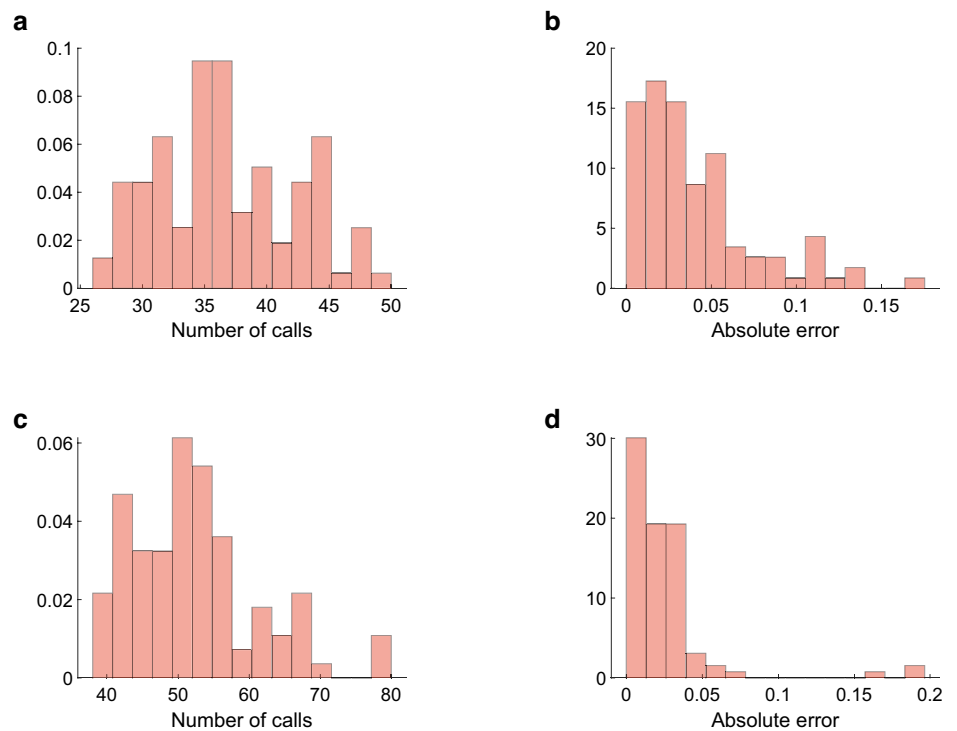
The results of AK-AMIS are given in Table 1 and compared with MCS, AK-MCS, MetaAK-IS2 and MetaAL-OIS. The variation of performances over the one hundred applications of the algorithm are further illustrated on Fig. 5. On that example, "design point" based methods such as AK-IS would fail at providing an accurate estimation of the failure probability, AK-MCS is therefore used as an alternative for reference. The proposed method converges significantly quicker than the selected references, with a reduction of about forty percent of the number of call compared to MetaAL-OIS. However, the precision of the estimation is also lower and interestingly reaches on average the error threshold (4%). The error amplitude for both cases in Table 1 and histogram on Fig. 5 show that some outliers exist in the performances with respectively ten and three applications terminating with an error above 10%. While the results illustrate how efficiently the algorithm converges, these outliers highlight that there is room for improvement with respect to the stopping criteria. The performance evolution between case one and two highlights how efficiently the method scales with decreasing failure probabilities with the number of calls being
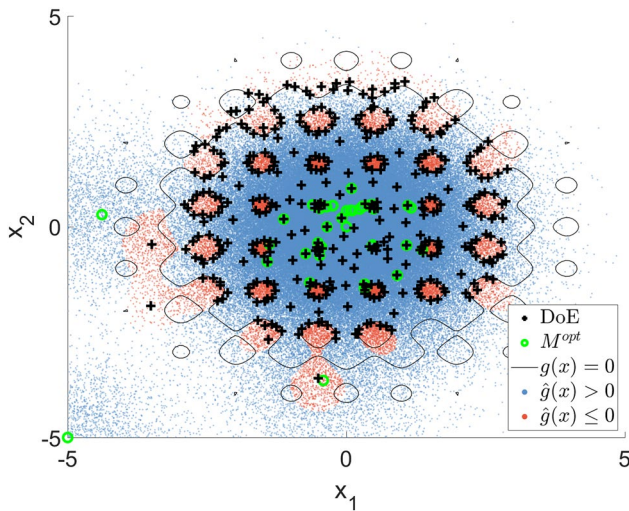
Fig. 6 Scatter plot of the limit-state, DoE and population on application 2

failure domains. The shape of its limit-state makes this example particularly challenging. It is associated with a quite high failure probability ($P_f \sim 7 \times 10^{-2}$). Considering the high probability of failure this case is not the kind of application AK-AMIS is designed for, allowing us to observe the limits of its applicability. The performance function reads as follows (Razaaly and Congedo 2018):

$$g(X) = 10 - \sum_{i=1}^{2} \left(x_i^2 - 5\cos(2\pi x_i) \right). \tag{26}$$

x_1 and x_2 are independent standard Gaussian random variables. Figure 6 illustrates that the algorithm successfully discovers and refine the calibration around the complex limit state. It is also worth noticing that the mean of the Gaussian components stayed quite close to the origin, suggesting that there is not a great added value from the population refinement scheme.

The limiting criterion in this application is the modeling error as illustrated in Fig. 7 illustrating a second time that relying solely on the covariance-based criterion would have led to a premature termination of the algorithm. The modeling error successfully converges in all applications of the method but the amplitude of variation is quite large and noisy even after a large number of iterations, which is probably linked to the complexity of the limit state geometry.
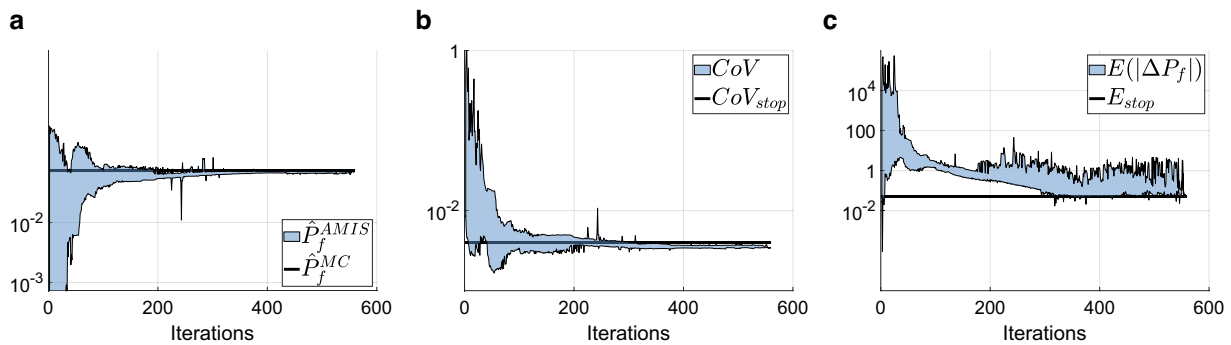
increased by a factor of less than two for a decrease in failure probability from an order 10^{-3} to 10^{-8}.

3.2 The modified Rastrigin function

A standard 2-D analytical function with highly non-linear behavior known for involving non-convex and isolated

Fig. 7 Convergence figure for application 2 of **a** the failure probability, **b** the CoV stopping parameter, and **c** the error stopping parameter. **d** Histogram of the number of calls to the performance function

Table 2 Reliability results for application 2

Method	MCS	AK-MCS[a]	Meta AK-IS[2a]	MetaAL-OIS[a]	AK-AMIS Averaged	AK-AMIS Amplitude
N_{calls}	2×10^6	391	480	151	390	$[297, 562]$
$\hat{P}_f$	7.30×10^{-2}	7.43×10^{-2}	7.35×10^{-2}	7.31×10^{-2}	7.00×10^{-2}	$[6.79, 7.90] \times 10^{-2}$
CoV(%)	2.51×10^{-1}				3.49×10^{-1}	$[3.13, 3.79] \times 10^{-1}$
ΔP_f(%)	–	2.39	2.39	0.10	4.44	$[1.11, 8.14]$

[a]Results from Table 5 in Razaaly and Congedo (2018)

Fig. 8 Histogram of **a** the number of calls to the performance function and **b** the absolute relative error over the 100 applications of the algorithm for application 2
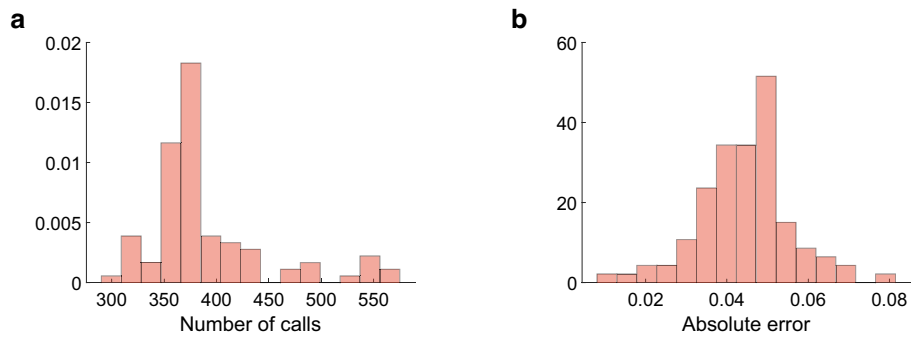
Table 3 Input distributions and parameters of example 3

Parameter	Distribution	μ	σ
c_1	Normal	1	0.1
c_2		0.1	0.01
m		1	0.05
r		0.5	0.05
t_1		1	0.02
F_1		1	0.02

Convergence of the other two parameters is overall steady but quite slow.

The results, given in Table 2 and illustrated in Fig. 8, show that the number of calls before convergence varies significantly between 297 and 562 with an average of 390. The error stays at satisfactory levels with a maximum of 8.14% and an average of 4.44. Compared with the reference methods the performances are mixed with slower convergence and higher error than MetaAL-OIS, but comparable performances as AK-MCS. This result is not particularly surprising since this application was designed to challenge the proposed method on cases it is not designed for. On this application there is very little benefit from the population refinement scheme since the limit-state is already well covered by a reasonable Monte Carlo population. Under those

circumstances the proposed method simply behaves as AK-MCS and exhibit relatively similar performances.

3.3 Dynamic response of a non-linear oscillator

The last standard function is a dimension six problem with a relatively smooth limit-state and reasonable failure probability. The performance function corresponds to non-linear undamped single degree of freedom system an reads as (Razaaly and Congedo 2018):

$$g(c_1, c_2, m, r, t_1, F_1) = 3r - \left| \frac{2F_1}{m\omega_0^2} \sin\left(\frac{\omega_0 t_1}{2} \right) \right| \qquad (27)$$

with $\omega_0 = \sqrt{\frac{c_1 + c_2}{m}}$. The six input parameters are independent Gaussian random variables with parameters as detailed in Table 3. The random variables are not standard Gaussian; however, the problem is solved in the standard space. This example is not extremely challenging but allows us to compare the performances on a slightly higher dimension case.

The convergence plots in Fig. 9 illustrate how fast the algorithm converges to the failure probability in this example with a very narrow amplitude over all applications after only twenty iterations. On this example the limiting stopping criterion is the one related to the covariance of the estimator
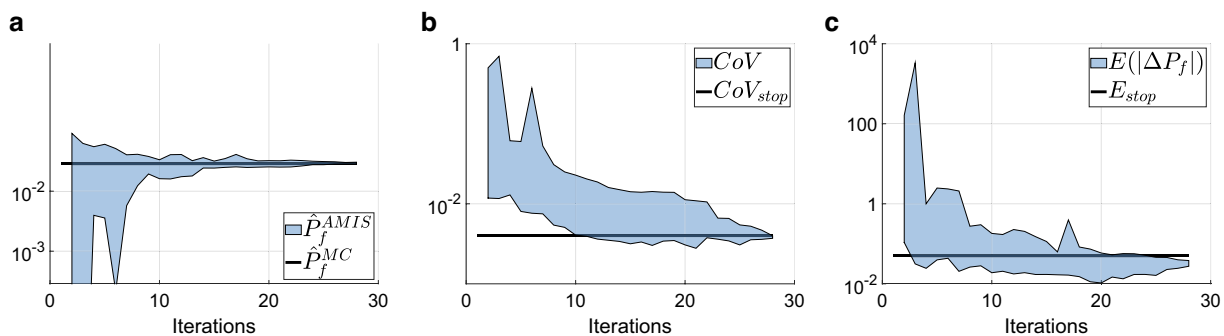

Fig. 9 Convergence figure for application 3 of **a** the failure probability, **b** the CoV stopping parameter, and **c** the error stopping parameter. **d** Histogram of the number of calls to the performance function

Table 4 Reliability results for application 3

Method	MCS	AK-MCS[a]	MetaAL-OIS[a]	AK-AMIS	
				Averaged	Amplitude
N_{calls}	2×10^7	58	70	24.7	$[18, 35]$
$\hat{P}_f$	2.83×10^{-2}	2.83×10^{-2}	2.85×10^{-2}	2.87×10^{-2}	$[2.50, 3.17] \times 10^{-2}$
CoV(%)	1.30×10^{-1}			3.67×10^{-1}	$[2.78, 4.00] \times 10^{-1}$
$\Delta P_f(\%)$	–	0.1	0.6	2.69	$[3.09 \times 10^{-2}, 12.4]$

[a]Results from Table 7 in Razaaly and Congedo (2018)

Fig. 10 Histogram of **a** the number of calls to the performance function and **b** the absolute relative error over the 100 applications of the algorithm for application 3
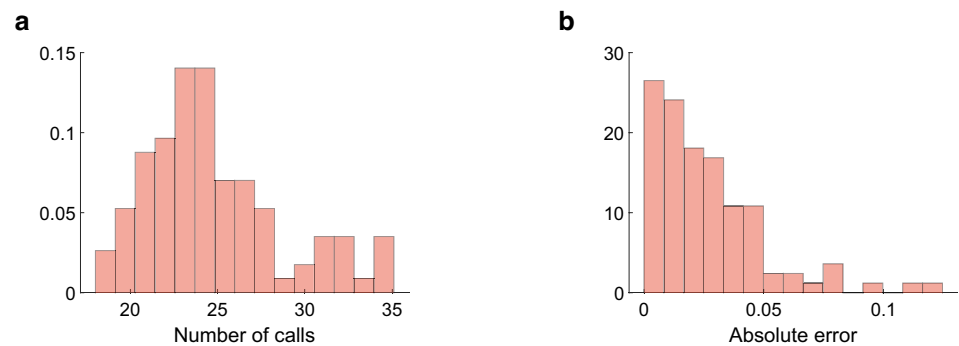
meaning that relying only on the error-based on would have led to termination of the algorithm before convergence of the failure probability. This result is the opposite of what was observed in the two last example highlighting the necessity of both criteria. The results in Table 4 and histograms in Fig. 10 highlight very good performances compared to the reference methods. It converges on average twice as fast as AK-MCS and MetaAL-OIS while providing very satisfactory accuracy with an average error of 2.8%. The histograms of performance metrics show that these results are quite consistent, event tough a few outliers exist with 2 of the 100 simulations reaching an error above 10%.

3.4 An industrial case with extremely rare failure event

The last example is an industrial example based on a finite-element model in dimension five and associated with an extremely rare failure event. The system is a low-alloy steel overpack for radioactive waste storage designed for the French deep geological repository project in collaboration with the French national radioactive waste management agency (Andra). The main characteristics of the system are discussed below, for more details the reader can refer to Persoons et al. (2021).

The overpack is the external part sealing the waste and isolating it from the environment during the first few centuries of storage. The part is near cylindrical and is stored in horizontal tunnels supported by a cylindrical steel lining

as illustrated in Fig. 11. The overpack is subjected to uncertain and time-dependent corrosion and uncertain and time-dependent mechanical loading. Water can accumulate in the tunnel by diffusion through the rock environment, creating a time-dependent water level H_w in the tunnel and affecting the corrosion process. The mechanical loading is composed of a constant and deterministic external fluid pressure F_f and, after buckling of the liner, a contact pressure with second order pressure profile characterized by a time-dependent contact angle α.

Five uncertain parameters control the system, four describe the corrosion process and one additional for the mechanical loading:

- C_a^i: initial stage (first century) atmospheric corrosion rate (µ*m*/year).
- C_a^f: long-term stabilized atmospheric corrosion rate (µ*m*/year).
- C_w^f: long term stabilized corrosion rate in water (µ*m*/year).
- C_{lin}: corrosion rate at the outer surface of the liner (µ*m*/year). This parameter is related to the buckling date of the liner and the start of the external mechanical contact pressure applied to the overpack.
- P_c: nominal contact pressure at the center of the pressure profile (*MPa*).

Each parameter is associated with independent lognormal distributions calibrated from expert's judgement elicitation.
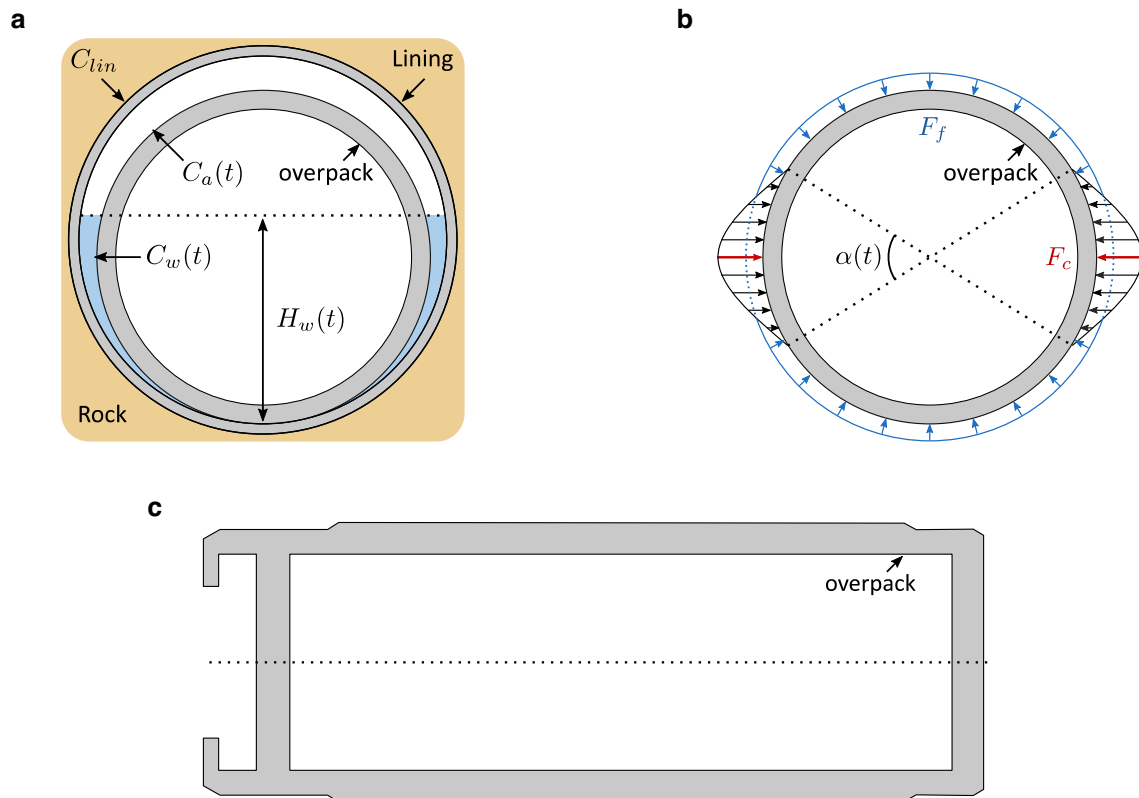
Fig. 11 Schematic illustration of **a** a cross section of the corrosion process, and **b** a cross-section of the mechanical loading, and **c** the longitudinal geometry of the overpack

Table 5 Input distributions and parameters of industrial case study

Parameter	Distribution	Mean	Standard deviation
C_a^i	Lognormal	$\log\left(\sqrt{30}\right)$	$\frac{1}{3}\log\left(\sqrt{30}*\frac{3}{2}\right)$
C_a^f		0	$\frac{1}{3}\log(15)$
C_w^f		$\log(10)$	$\frac{1}{3}\log(3)$
C_{lin}		$\log\left(\frac{1}{\sqrt{10}}\right)$	$\frac{1}{3}\log\left(\frac{15}{\sqrt{10}}\right)$
F_c		$\log(8)$	$\frac{1}{3}\log(2)$

The details about the distributions and their parameters are detailed in Table 5.

The overpack is considered as failing if it loses its sealing before 500 years either by plastic fracture or corrosion. The mechanical behavior of the system is studied through a finite element model with elastic–plastic behavior associated with a Gurson's plastic degradation model (calibrated experimentally). The corrosion is simply considered as a time-dependent and non-homogeneous reduction of the thickness of the overpack. The continuous variation of the
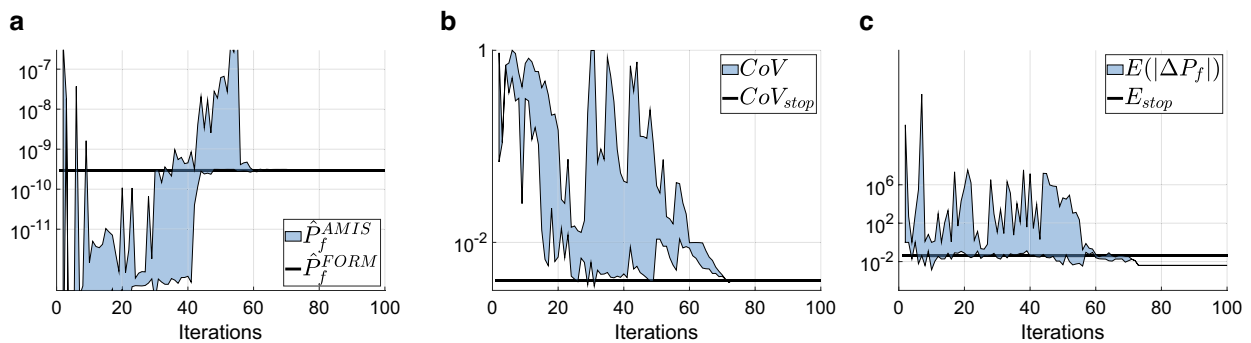

Fig. 12 Convergence figure for the application 4 of **a** the failure probability, **b** the CoV, and **c** the error stopping parameter

geometry though time is handled by iteratively solving static problems (one per century) corresponding to the state of the system at the given date.

Figure 12 illustrates the convergence of the method on this case. Due to the high computation time of the model, the method has only been applied 5 times which is why the convergence plots are much noisier than the previous examples. We can observe that the failure probability estimate clearly converged before termination of the algorithm. On that case the variance-based criterion is the limiting one, which is explained by the extremely low failure probability.

On the original study of the system AK-MCS was unable to estimate the failure probability because no-reasonably sized Monte Carlo population contained failing samples. The computation time associated with the model makes it impossible to apply classical simulation methods. The results are therefore only compared with FORM, see Table 6. It is worth mentioning that, in order to save computation time, the FORM analysis was initiated with a known failing sample, previously identified with the AK-AMIS method. The number of evaluations of the performance function therefore doesn't account for the extra few iterations generally necessary to locate one when no information is available about the system.

Table 6 Reliability results for application 4

Method	FORM	AK-AMIS	
		Averaged	Amplitude
N_{calls}	> 46	68.8	$[55, 78]$
$\widehat{P}_f$	2.96×10^{-10}	3.13×10^{-10}	$[3.09, 3.16] \times 10^{-2}$

We can observe that AK-AMIS provides a failure probability estimation compatible the FORM approximation for a reasonable computation cost (about 50% more evaluations than the search for the design point from FORM). The FORM estimation is smaller and very close ($\sim 6\%$ difference) with those of AK-AMIS, which suggest that the limit-state is concave and quite linear around a unique design point.

From previous study of the system, we know that the failure is mainly driven by the corrosion process, and from observing the results in the 2D space C_a^i, C_w^f we can see a rough estimation of the limit-state in the standard space (Fig. 13a) and in the physical space (Fig. 13b). The system exhibits two failure regions associated with a case of extremely fast early stage atmospheric corrosion and extremely fast late-stage corrosion under water. Both can be precisely located by solving an optimisation problem and their respective local failure probabilities can be estimated using FORM as 2.96×10^{-10} and 4.12×10^{-13}. The figure illustrates that AK-AMIS managed to locate both failure modes by efficiently exploring the input space.

4 Conclusions and discussion

This article aims at describing a new adaptive Kriging method combined with adaptive importance sampling, reducing both the modeling and sampling errors dynamically. The solution developed with the AK-AMIS method is to adaptively build a near optimal auxiliary Gaussian mixture distribution simultaneously as enriching the DoE. To do so, AK-AMIS relies on a new scheme combining Adaptive Kriging with adaptive multiple importance sampling in which a sampling distribution is defined at each iteration
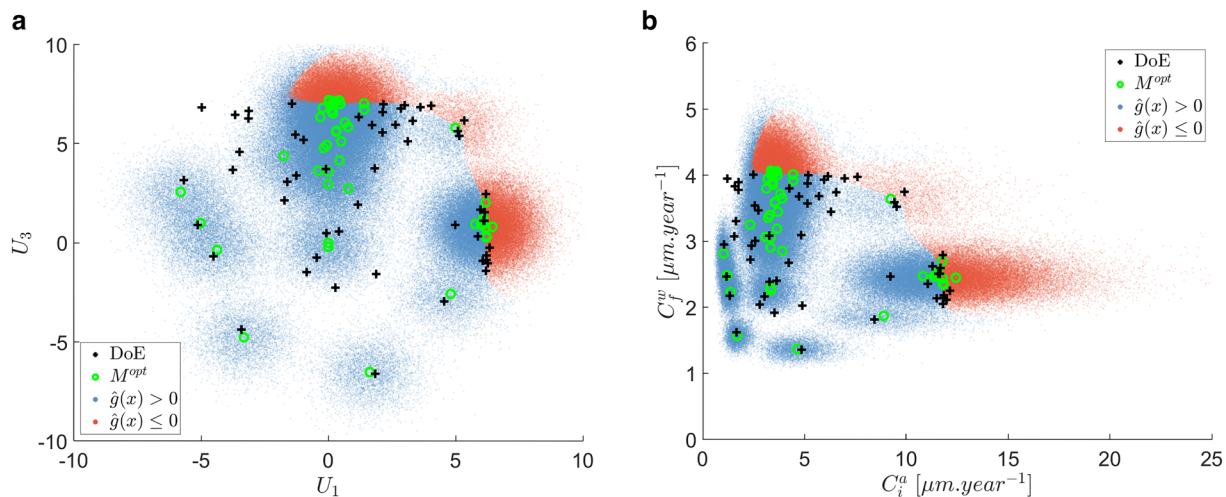

Fig. 13 Scatter plot of the limit-state, DoE and population on application 4 for the auxiliary distribution strategy **a** in the and **b** in the physical space

and new samples are drawn from it before being added to the mixture population.

The proposed strategy exhibits very good performances especially on cases of low failure probability where it significantly outperforms comparable references methods, while still providing accurate and reasonably fast estimation on challenging cases it is not designed for. Two appealing results exhibited are first that even with a very sparse initial training dataset, the method is able to efficiently explore the input space and converge toward the correct failure probability. Second, the approach scales very efficiently with decreasing failure probabilities. Overall, the optimization-based sampling strategy performed better on the illustrated examples on rare failure events; efficiently exploring the input space and providing a comprehensible compromise between accuracy and computation time. The method is finally tested on an actual engineering case with extremely rare failure event, where it managed to efficiently explore the input space an locate two failure modes.

The proposed method could still benefit from improvements, the first main limitation being the cases of very high dimension, which is a well-known limitation of the Kriging models. For these cases, since the covariance function is based on Euclidean distance, the posterior standard deviation tends to be very high and roughly equal for all samples of the population. This result prevents the learning function from providing useful weights for the samples and in turn prevents the algorithm from efficiently converging toward the limit-state. Some methods exist in the literature to tackle this limitation the Kriging models such as additive Gaussian process regression models and are planned to be explored in future development. Second, the stopping criteria proposed are quite basic and, even though their combined usage provide satisfactory performances, they are not reliably correlated with the quantity they are intended to measure. More sophisticated error-based criteria have been explored and might provide improved robustness of the result.

Appendix 1: Discussion of the *p* hyperparameter

As mentioned in Sect. 2.2 the choice of the p hyperparameter is at least partly arbitrary since it does not directly relate to an interpretable performance metric. In order to choose an adequate value one can only rely on empirical evidences. A good basis for such evidence is to study the sensitivity of the performances with respect to the parameter. The results of such a study are presented in this Appendix where the method is applied on three analytical examples (examples one and three from Sects. 3.1 and 3.2 and example four from Appendix 2) with varying values of p.

Thirty values of p are tested in the interval $p \in [1, 6]$. For each, the method is applied ten times to observe the variation of performances (the same ten random number regenerator seeds are used for each value of p). The results are illustrated in Figs. 14 and 15 presenting respectively the mean and amplitude of number of calls to the performance function and relative error.

It can be observed from Fig. 14 that, in terms of computation time, there is a tendency for very high values of p to perform poorly, especially for the first two examples (a) and (b). For p between one and four no strong tendency is observable for the last two examples (b) and (c). For the first example (a) there seem to be an optimal value between $p = 3$ and $p = 4$.

In Fig. 15 it can be observed that for the first example (a) small values of p induce a higher variance of performance. In contrast, on the second example (b), the same tendency is observed for high values of p. No strong tendency can be observed from example three (c). From Fig. 15a the optimal value seem to be between $p = 3$ and $p = 5$, while for Fig. 15b the optimal value seem to be between $p = 1$ and $p = 3$

From these results, the choice of $p = 3$ as proposed in Sect. 2.2 appears as a reasonable compromise providing satisfying performances and computation time on all tested examples.
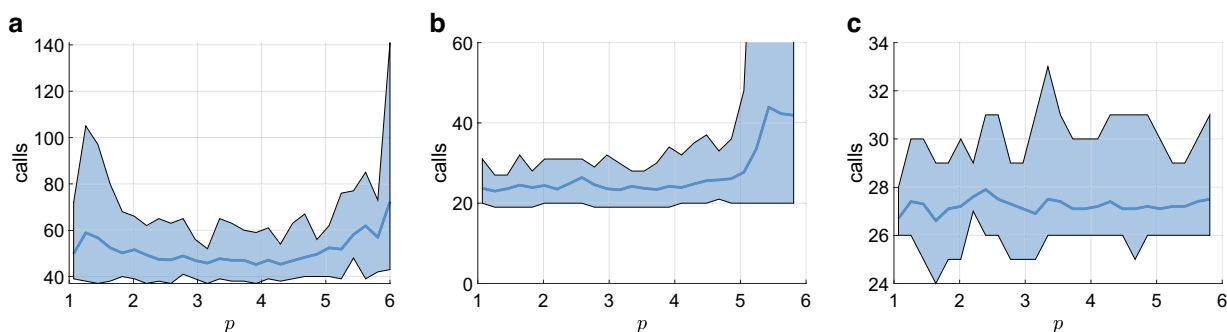

Fig. 14 Average and extrema of the number of calls to the performance function as a function of the p hyperparameter for **a** example 1 (case 2), **b** example 3, and **c** example 4 (see) Appendix 2

Appendix 2: Additional performance comparison

In addition to the four examples presented in Sect. 3 the performances of the proposed approach are compared to two other relevant methods form the literature, i.e., AK-ALIS (Liu et al. 2020) and AKOIS (Zhang et al. 2020). The study is limited to the analytical example function presented in the respective articles and the reference results are reported as published. The performances of the proposed AK-AMIS method are studied as described in Sect. 3 and a hundred application have been performed on each application.

Overall the proposed AK-AMIS method performs very well on these four additional examples and reaches satisfying precision on all of them. AK-AMIS converges around 25% quicker than AK-ALIS while consistently staying below 2.4% error (Table 7). On examples six and seven AK-AMIS successfully converges but requires about 25% more evaluations of the performance function than AK-OIS (Tables 8 and 9). On example 8 (Table 10) AK-AMIS converges around 50% faster than AK-OIS while inducing a slightly higher error of about 2%.

Case 5: An analytical function from Liu et al. (2020),

Case 6: A bivariate example (Zhang et al. 2020),

Case 7: A polynomial function with multiple MPPs (Zhang et al. 2020),

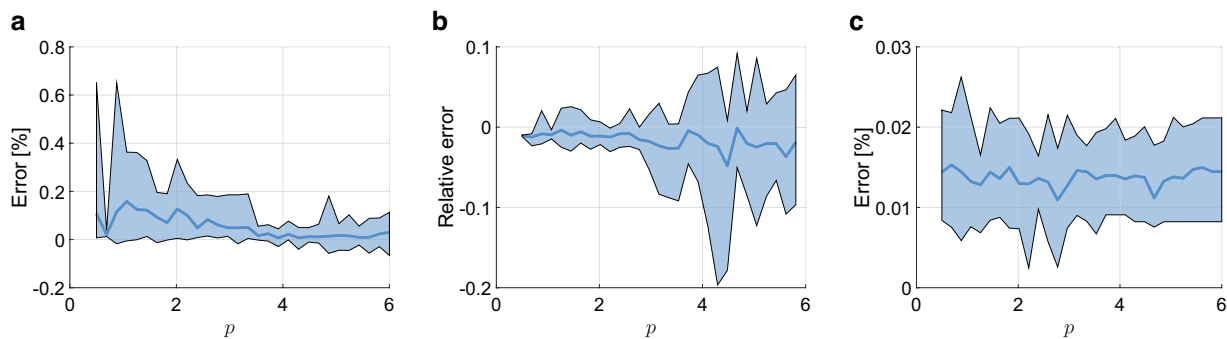Case 8: A polynomial function with multiple MPPs (Zhang et al. 2020).

Fig. 15 Average and extrema of the failure probability error as a function of the p hyperparameter for **a** example 1 (case 2), **b** example 3, and **c** example 4 (see) Appendix 2

Table 7 Results for application 5

$$g(X) = 1 - \frac{(x_1-1)^2}{30} - \frac{(x_2-1)^3}{36}$$

Method	AK-ALIS + U	AK-ALIS + H	AK-AMIS	
			Averaged	Amplitude
N_{calls}	36	38	26.9	[24, 33]
$\widehat{P}_f$	1.67×10^{-5}	1.70×10^{-5}	1.64×10^{-5}	$[1.62, 1.66] \times 10^{-5}$
CoV(%)	2.3	2.5	3.94×10^{-1}	$[3.84, 4.00] \times 10^{-1}$
$\Delta P_f(\%)$	–	–	1.57	$[7.19 \times 10^{-1}, 2.4]$

Table 8 Results for application 6

$$g(X) = \frac{1}{2}(x_1 - 2)^2 - \frac{3}{2}(x_2 - 5)^3 - 3$$

Method	AKOIS + U	AKOIS + EFF	AK-AMIS	
			Averaged	Amplitude
N_{calls}	19	20	27.6	[25, 31]
$\widehat{P}_f$	2.87×10^{-5}	2.87×10^{-5}	2.87×10^{-5}	$[2.85, 2.91] \times 10^{-5}$
CoV(%)	0.76	0.94	3.95×10^{-1}	$[3.9, 4.00] \times 10^{-1}$
$\Delta P_f(\%)$	0.06	0.13	0.37	$[4.87 \times 10^{-4}, 1.4]$

Table 9 Results for application 7

$$g(X) = 5 + 0.5(x_1 + 2)^2 - 1.5(x_1 + 2)^2 - x_2$$

Method	AKOIS + U	AKOIS + EFF	AK-AMIS	
			Averaged	Amplitude
N_{calls}	18	18	25	$[22, 32]$
$\widehat{P}_f$	6.85×10^{-4}	6.87×10^{-4}	6.85×10^{-4}	$[6.75, 6.92] \times 10^{-4}$
CoV(%)	–	–	4.90×10^{-1}	$[4.78, 4.99] \times 10^{-1}$
ΔP_f(%)	0.79	1.23	4.17×10^{-1}	$[1.13 \times 10^{-2}, 1.5]$

Table 10 Results for application 8A

$$g(X) = \min\big(g_1(X), g_2(X), g_3(X), g_4(X)\big), \begin{cases} g_1(X) = 5 + \frac{(x_1 - x_2)^2}{10} - \frac{1}{\sqrt{2}}(x_1 - x_2) \\ g_2(X) = 5 + \frac{(x_1 - x_2)^2}{10} + \frac{1}{\sqrt{2}}(x_1 - x_2) \\ g_3(X) = \frac{8}{\sqrt{2}} + x_1 - x_2 \\ g_4(X) = \frac{8}{\sqrt{2}} - x_1 + x_2 \end{cases}$$

Method	AKOIS + U	AKOIS + EFF	AK-AMIS	
			Averaged	Amplitude
N_{calls}	53	62	30.7	$[25, 61]$
$\widehat{P}_f$	6.45×10^{-5}	6.41×10^{-5}	6.28×10^{-5}	$[6.00, 6.45] \times 10^{-5}$
CoV(%)	3.71	3.12	3.78×10^{-1}	$[2.60, 4.00] \times 10^{-1}$
ΔP_f(%)	–	–	2.13	$[1.80 \times 10^{-1}, 6.99]$

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Replication of Results The results related to applications 1–3 have been generated using the codes included as Supplementary Material. The results and model of application 4 are proprietary in nature, their availability is handled by Andra. Requests should be addressed to Dr. Bumbieler (frederic.bumbieler@ andra.fr).

References

Au S-K, Beck JL (2001) Estimation of small failure probabilities in high dimensions by subset simulation. Probab Eng Mech 16(4):263–277. https://doi.org/10.1016/S0266-8920(01)00019-4

Balesdent M, Morio J, Marzat J (2013) Kriging-based adaptive importance sampling algorithms for rare event estimation. Struct Saf 44(September):1–10. https://doi.org/10.1016/j.strusafe.2013.04.001

Bichon BJ, Eldred MS, Swiler LP, Mahadevan S, McFarland JM (2008) Efficient global reliability analysis for nonlinear implicit performance functions. AIAA J 46(10):2459–2468. https://doi.org/10.2514/1.34321

Cadini F, Santos F, Zio E (2014) An improved adaptive Kriging-based importance technique for sampling multiple failure regions of low probability. Reliab Eng Syst Saf 131(November):109–117. https://doi.org/10.1016/j.ress.2014.06.023

Cornuet J-M, Marin J-M, Mira A, Robert CP (2012) Adaptive multiple importance sampling. Scand J Stat 39(4):798–812. https://doi.org/10.1111/j.1467-9469.2011.00756.x

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297. https://doi.org/10.1007/BF00994018

Crestaux T, Le Maître O, Martinez J-M (2009) Polynomial chaos expansion for sensitivity analysis. Reliab Eng Syst Saf 94(7):1161–1172. https://doi.org/10.1016/j.ress.2008.10.008

Ditlevsen O, Olesen R, Mohr G (1986) Solution of a class of load combination problems by directional simulation. Struct Saf 4(2):95–109. https://doi.org/10.1016/0167-4730(86)90025-1

Dubourg V, Sudret B, Deheeger F (2013) Metamodel-based importance sampling for structural reliability analysis. Probab Eng Mech 33(July):47–57. https://doi.org/10.1016/j.probengmech.2013.02.002

Echard B, Gayton N, Lemaire M (2011) AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation. Struct Saf 33(2):145–154. https://doi.org/10.1016/j.strusafe.2011.01.002

Echard B, Gayton N, Lemaire M, Relun N (2013) A combined importance sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models. Reliab Eng Syst Saf 111(March):232–240. https://doi.org/10.1016/j.ress.2012.10.008

Elvira V, Martino L, Luengo D, Bugallo MF (2015) Efficient multiple importance sampling estimators. IEEE Signal Process Lett 22(10):1757–1761. https://doi.org/10.1109/LSP.2015.2432078

Elvira V, Martino L, Luengo D, Bugallo MF (2019) Generalized multiple importance sampling. Stat Sci 34(1):129–155. https://doi.org/10.1214/18-STS668

Hasofer AM, Lind NC (1974) Exact and invariant second-moment code format. J Eng Mech Div 100(1):111–121

Hurtado JE, Alvarez DA (2001) Neural-network-based reliability analysis: a comparative study. Comput Methods Appl Mech Eng 191(1–2):113–132. https://doi.org/10.1016/S0045-7825(01)00248-1

Lee I, Noh Y, Yoo D (2012) A novel second-order reliability method (SORM) using noncentral or generalized chi-squared distributions. J Mech Des 134(10):100912. https://doi.org/10.1115/1.4007391

Lemaire M (2013) Structural-reliability. Wiley, Hoboken

Liu F, Wei P, Zhou C, Yue Z (2020) Reliability and reliability sensitivity analysis of structure by combining adaptive linked importance sampling and Kriging reliability method. Chin J Aeronaut 33(4):1218–1227. https://doi.org/10.1016/j.cja.2019.12.032

Matheron G (1973) The intrinsic random functions and their applications. Appl Probab Trust 5(3):439–46831

Melchers RE (1990) Radial importance sampling for structural reliability. J Eng Mech 116(1):189–203. https://doi.org/10.1061/(ASCE)0733-9399(1990)116:1(189)

Morio J (2011) Non-parametric adaptive importance sampling for the probability estimation of a launcher impact position. Reliab Eng Syst Saf 96(1):178–183. https://doi.org/10.1016/j.ress.2010.08.006

Moustapha M, Marelli S, Sudret B (2021) A generalized framework for active learning reliability: survey and benchmark, June. ArXiv:2106.01713 [Stat], http://arxiv.org/abs/2106.01713

Nie J, Ellingwood BR (2000) Directional methods for structural reliability analysis. Struct Saf 22(3):233–249. https://doi.org/10.1016/S0167-4730(00)00014-X

Owen A, Zhou Y (2000) Safe and effective importance sampling. J Am Stat Assoc 95(449):135–143

Persoons A, Serveaux J, Beaurepaire P, Labergere C, Chateauneuf A, Saanouni K, Bumbieler F (2021) "Reliability analysis of the ductile fracture of overpacks for high-level radioactive waste in repository conditions. ASCE–ASME J Risk Uncertain Eng Syst A 7(2):04021010. https://doi.org/10.1061/AJRUA6.0001121

Pradlwarter HJ, Pellissetti MF, Schenk CA, Schueller GI, Kreis A, Fransen S, Calvi A, Klein M (2005) Realistic and efficient reliability estimation for aerospace structures. Comput Methods Appl Mech Eng 194:1597–1617

Rackwitz R (2001) Reliability analysis—a review and some perspectives. Struct Saf 23(4):365–395. https://doi.org/10.1016/S0167-4730(02)00009-7

Razaaly N, Congedo PM (2018) Novel algorithm using active metamodel learning and importance sampling: application to multiple failure regions of low probability. J Comput Phys 368(September):92–114. https://doi.org/10.1016/j.jcp.2018.04.047

Sbert M, Havran V, Szirmay-Kalos L (2016) Variance analysis of multi-sample and one-sample multiple importance sampling. Comput Graph Forum 35(7):451–460. https://doi.org/10.1111/cgf.13042

Schuëller GI, Pradlwarter HJ, Koutsourelakis PS (2004) A critical appraisal of reliability estimation procedures for high dimensions. Probab Eng Mech 19(4):463–474. https://doi.org/10.1016/j.probengmech.2004.05.004

Song K, Zhang Y, Yu X, Song B (2019) A new sequential surrogate method for reliability analysis and its applications in engineering. IEEE Access 7:60555–60571. https://doi.org/10.1109/ACCESS.2019.2915350

Sudret B (2008) Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Syst Saf 93(7):964–979. https://doi.org/10.1016/j.ress.2007.04.002

Sudret B (2012) Meta-models for structural reliability and uncertainty quantification. arXiv. http://arxiv.org/abs/1203.2062.

Teixeira R, Nogal M, O'Connor A (2021) Adaptive approaches in metamodel-based reliability analysis: a review. Struct Saf 89(March):102019. https://doi.org/10.1016/j.strusafe.2020.102019

Veach E, Guibas LJ (1995) Optimally combining sampling techniques for Monte Carlo rendering. In: Proceedings of the 22nd annual conference on computer graphics and interactive techniques—SIGGRAPH '95. ACM Press, pp 419–428. https://doi.org/10.1145/218380.218498

Wang Z, Shafieezadeh A (2019) ESC: an efficient error-based stopping criterion for Kriging-based reliability analysis methods. Struct Multidisc Optim 59(5):1621–1637. https://doi.org/10.1007/s00158-018-2150-9

Wei P, Tang C, Yang Y (2019) Structural reliability and reliability sensitivity analysis of extremely rare failure events by combining sampling and surrogate model methods. Proc Inst Mech Eng O 233(6):943–957. https://doi.org/10.1177/1748006X19844666

Xiong B, Tan H (2018) A robust and efficient structural reliability method combining radial-based importance sampling and Kriging. Sci China Technol Sci 61(5):724–734. https://doi.org/10.1007/s11431-016-9068-1

Yang X, Liu Y, Fang X, Mi C (2018) Estimation of low failure probability based on active learning Kriging model with a concentric ring approaching strategy. Struct Multidisc Optim 58(3):1175–1186. https://doi.org/10.1007/s00158-018-1960-0

Yi J, Zhou Q, Cheng Y, Liu J (2020) Efficient adaptive Kriging-based reliability analysis combining new learning function and error-based stopping criterion. Struct Multidisc Optim 62(5):2517–2536. https://doi.org/10.1007/s00158-020-02622-3

Yun W, Lu Z, Jiang X (2018) An efficient reliability analysis method combining adaptive Kriging and modified importance sampling for small failure probability. Struct Multidisc Optim 58(4):1383–1393. https://doi.org/10.1007/s00158-018-1975-6

Zhang J, Taflanidis AA (2018) Adaptive Kriging stochastic sampling and density approximation and its application to rare-event estimation. ASCE–ASME J Risk Uncertain Eng Syst A 4(3):04018021. https://doi.org/10.1061/AJRUA6.0000969

Zhang X, Wang L, Sørensen JD (2020) AKOIS: an adaptive Kriging oriented importance sampling method for structural system reliability analysis. Struct Saf 82(January):101876. https://doi.org/10.1016/j.strusafe.2019.101876

Zhao H, Yue Z, Liu Y, Gao Z, Zhang Y (2015) An efficient reliability method combining adaptive importance sampling and Kriging metamodel. Appl Math Model 39(7):1853–1866. https://doi.org/10.1016/j.apm.2014.10.015