Reliability assessment with density scanned adaptive Kriging

Teixeira, Rui; Nogal , Maria; O'Connor, Alan; Martinez-Pastor, Beatriz

DOI
10.1016/j.ress.2020.106908

Publication date
2020

Document Version
Final published version

Published in
Reliability Engineering and System Safety

Reliability Engineering and System Safety 199 (2020) 106908

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Reliability assessment with density scanned adaptive Kriging

Rui Teixeira[*,a], Maria Nogal[c], Alan O'Connor[a], Beatriz Martinez-Pastor[b]

[a] Department of Civil, Structural, and Environmental Engineering, Trinity College Dublin, Ireland
[b] School of Civil Engineering, University College Dublin, Ireland
[c] Department of Materials, Mechanics, Management & Design, Technical University Delft, The Netherlands

ARTICLE INFO

Keywords:
Reliability
Adaptive Kriging
Active learning
Metamodeling
Density scanning

ABSTRACT

Reliability assessment with adaptive Kriging has gained notoriety due to the Kriging capability of accurately replacing the performance function while performing as a self-improving function for learning procedures.

Recent works on adaptive Kriging pursued to improve the efficiency of the active learning through the application of distinct learning functions, sampling methods, or frameworks to assess the learning space. Within this context, the present work exploits three innovative applications of density scanning to improve the efficiency of the adaptive Kriging. Density scanning has significant synergies with adaptive Kriging implementation. For most learning criteria, candidate points occur in dense clusters. This is due to the fact that the most efficient learning strategies pursue to improve predictions near the failure region, or when the prediction uncertainty is large.

Identifying dense clusters of points, and fomenting exploitation of these, parallelizing computations, and limiting the generation of dense clusters in the design of experiments are examples of learning frameworks that can be achieved with density scanning. Three reference examples are researched in the present work, a complex function, a series system, and a relatively high dimension engineering problem. For all the cases, the application of density scanning is identified to improve the active learning efficiency.

1. Introduction

Uncertainty is prevalent in real engineering systems. To a large extent every single variable in an engineering system is prone to uncertainty. Therefore, in order to enable robust engineering designs, the different sources of uncertainty need to be comprehensively analysed during the life-cycle assessment of an engineering system. In this context, reliability analysis, is the approach that provides the tools and techniques to characterize the probabilistic behaviour of a structure or system with the ultimate goal of characterizing its susceptibility to failure.

Due to its relevance, significant efforts have been made in order to improve the reliability analysis procedures. The growing trends in data availability, in modeling complexity, and availability of resources for analysis have increased the requirement for efficient reliability characterization procedures. Adaptive Kriging (AK) procedures are a reference example that have gained significant relevance in the context of efficient reliability analysis.

The idea of using Kriging models for reliability analysis of complex limit-state functions has been proven successful over the past few years. Kriging models are exact interpolators based on the idealization of the numerical model response as a realization of a Gaussian stochastic process [1]. Their capability to enclose uncertainty is of relevance for classification problems, such as the problem of reliability. They allow a significant reduction of the effort required for reliability analysis without compromising its accuracy.

The objective of the present paper is then to further exploit the application of Kriging models for reliability calculations. The analysis presented is focused on the application of an alternative complementary active learning framework that improves the reliability calculations with Kriging models.

It is noted that the most successful Kriging metamodeling applications to reliability analysis use an active learning function that provides unsupervised improvement in the surrogate approximation of the performance function. It consists in creating a surrogate that is progressively improved with heuristic enrichment of its design of experiments. The term enrichment is commonly used to characterize the procedure of active learning and sequential selection of new points points to improve the metamodeling approximation. Methods that use this approach are frequently classified as AK procedures.

The AK denomination has its roots in the work of [1], where the authors defined an innovative procedure for reliability analysis using Kriging models and Monte Carlo Sampling (MCS), the AKMCS. It uses the so-called U function, which has gained notoriety in the application

* Corresponding author.

https://doi.org/10.1016/j.ress.2020.106908
Received 7 June 2019; Received in revised form 17 January 2020; Accepted 27 February 2020
Available online 29 February 2020
0951-8320/

Nomenclature

β	Polynomial trend function
$\boldsymbol{X}$	Sample of support points
$\boldsymbol{Y}$	Performance evaluation at support points
g_{eval}	Number of performance function evaluations
e_r	Relative error of prediction
EGRA	Efficient Global Reliability Analysis
EFF	Expected Feasibility Function
IS	Importance Sampling
MCS	Monte Carlo Sampling
SS	Subset Sampling
LIF	Least Improvement Function
REIF	Reliability Expected Improvement function
ISKRA	Improved sequential reliability analysis
AK	Adaptive Kriging
g	Performance function
P_f	Probability of failure
CoV	Coefficient of variation
N_{MCS}	Size of Monte Carlo sample
G	Metamodel of g
DoE	Design of Experiments
f	Metamodel polynomial approximation
p	Degree of metamodel polynomial approximation
Z	Metamodel zero mean Gaussian component
d	Size of dimensional space
$\mathbf{C}$	Covariance matrix
σ^2	Metamodel constant process variance
μ_G and σ_G^2	Metamodel mean and variance prediction
R	Correlation function
θ	Metamodel hyperparameter
S	Search function
X_{n+1}	Selected candidate to enrich $\boldsymbol{X}$
$E[\cdot]$	Expected value
U	U learning function
Φ and ϕ	Standard cumulative and densities functions
ζ	Density scanning distance measure
MinPts	Minimum number of points in a density scanned group
D and D_{c_k}	Threshold quantifiers of ζ
KB	Kriging believer
c	Size of candidate points
x	Generic variable in the space
L	Subset of candidates
L_c	Subset of cross-candidates from L to build L_+
L_+	Cross-classification subset
C	Density group classification
l	Number of L subsets
P_m	Probability of misclassification
η	Expectation of error from misclassification
ds	Density scanning
RD	Response-distance function
H	H learning function
n_{iter}	Number of iterations

of AK due to its simplicity and efficiency. Despite the initial denomination of AKMCS, to a broader extent almost every AK procedure applies an active learning procedure and MCS to evaluate the probability of failure (P_f).

It is noted that application of Kriging models in reliability analysis dates previously to the work of [2]. In this pioneer work, limitations and improvements for further application of these models in the context of reliability analysis were discussed.

Later, Bichon et al. [3] introduced the approach of Efficient Global Reliability Analysis (EGRA), built on the Efficient Global Optimization (EGO) of Jones et al. [4], and highlighting the need for a progressive improvement in the reliability calculations with Kriging models. The authors propose the Expected Feasibility Function (EFF) for the active learning procedure.

In [5] the AKMCS was combined with importance sampling (IS), AKIS, for limit-state functions where the failure is confined to a particular region of the space, such that the most probable failure point (MPP) can be accurately defined. Fauriat and Gayton [6] also adapted the AKMCS, but for system reliability analysis. Huang et al. [7] applied subset sampling (SS) with an AK, AK-SS, improving even further the efficiency of the algorithm for very small probabilities of failure. Tong et al. [8] applied a similar AK approach that combined IS and SS, creating an hybrid algorithm called AK-SSIS. Significant research efforts have been identified in the introduction of new search functions and hybrid procedures. Lv et al. [9] uses line sampling combined with AK and introduces the learning function H, which uses information entropy to enrich the design of experiments used to create the metamodel. Sun et al. [10] also proposed a new learning function, the Least Improvement Function (LIF), that directly evaluates the probability of making a wrong classification through an uncertainty function (UF), and that considers the influence of the neighbour candidates. The LIF selects points that are expected to diminish the UF. Zhang et al. [11] proposed another search function, the Reliability Expected Improvement Function (REIF), that relates to the expected improvement (EI) of [4]. Gaspar et al. [12] uses a trust region for efficient single point design reliability assessments. A convergence criterion based on the stability of the probability of failure is also introduced in order to compromise the cost of the computational effort and the accuracy in the AK implementations. Wen et al. [13] introduced the concept of improved sequential reliability analysis (ISKRA), which uses parallel computations and adaptive samples to improve the AK implementation. A k-means algorithm and the Kriging believer approach of [14] are applied to enable multiple point enrichment. Recently, Lelièvre et al. [15] also improved the AKMCS using multiple point enrichment and parallel computations with k-means clustering. This new approach also allowed the reduction of the number of iterations demanded for the method. However, in both cases, this was only possible through sacrificing the number of performance function evaluations. Teixeira et al. [16] introduces the concept of biased randomisation in AK in order to weight the learning with existing *a priori* knowledge. Cui and Ghosn [17] proposed the SSKK method, which combines AK approach with k-means clustering and SS. Recently, Dong et al. [18] further elaborated on Kriging implementations by using it to solve the effort-demanding and relevant problem of time-variant reliability analysis. The added complexity of time-variant reliability analysis is indicative of the importance of AK, and in particular, of efficient metamodeling approaches. Other examples of the application of the Kriging models for practical structural engineering problems can be found in [19–25].

One of the initial limitations of the AK procedures was the definition of the convergence criterion that halts the learning function. In general, the learning criteria pursue to provide an accurate surrogate of the performance function, or to establish a confident P_f prediction. Schöbi et al. [26] highlighted the limitation of pursuing an accurate characterization of the surrogate AK surface, instead of focusing on the probability of failure. The results is that recent discussions regarding the learning approach have then been extended also to the stopping criteria, *e.g.*, [19,27].

In the present paper, innovative AK frameworks are researched and discussed. These result from the joint application of AK and density scanning. Candidate points in AK procedures that apply MCS, and that surrogate complex performance functions, are likely to form groups of large density close to the failure region, or where the uncertainty in the

metamodel prediction is large. In regions of large density the enrichment is expected to contribute to the classification of close points in the space. As a result, density scanning and identification of dense clusters in the space of candidate points has large synergy with the AK methodologies and with the requirement for having a notion of improvement that relates to the problem of reliability. A new approach to significantly improve the implementation of density scanning is proposed. It successfully tackles the large computational time and power required by it, enabling its feasibility for AK.

Density scanning is applied in two main innovative approaches, for sequential serial and parallel enrichment. The serial approach allows the prioritization of regions of the space of variables that are expected to enclose more information about the reliability problem, under the assumption that enriched points will contribute to classify their neighbors in the Design of Experiments (DoE), and exploiting synergies of the AK functionality (in terms of probability of failure P_f approximation), such as discussed in [10,27]. The parallel approximation uses the capability of the density algorithm to define the inter-relation between points of space with limited information and in an unsupervised way, hence mitigating the enrichment with points that have less efficient heuristics in the learning context and that may enclose redundant information; such as the prevalent k-means selection of points that are non-optimum in the context of the heuristics studied. Improvements of parallel and serial computing are presented, with the ultimate goal of minimizing the number of evaluation functions demanded for efficient reliability analysis. Density scanning is also researched to mitigate the selection of points that are close in the space.

To achieve the proposed goal, Section 2 discusses the theoretical background behind the reliability calculations with Kriging models, Section 3 discusses the application of density scanning in different frameworks for reliability analysis with AK, Section 4 presents examples of application and their discussion, which are then used to compile in Section 5 the main conclusions of the work developed.

2. Reliability analysis with Kriging models

Reliability analysis involves a problem that classifies a performance function $g(x)$ in the x point as failure or non-failure accordingly to,

$$\begin{cases} I_F(x) = 0,\ g(x) \geq 0 \\ I_F(x) = 1,\ g(x) < 0 \end{cases}$$

with I_F being the performance function binary evaluation of failure ($I_F(x) = 1$) and non-failure($I_F(x) = 0$). The probability of failure associated to $g(x)$ may be calculated using different techniques. A convenient approach to assess P_f is the MCS, which consists in sampling x accordingly to the probability density function ($f(x)$) that characterizes its occurrence. Estimation of the probability of failure in this case is given by the ratio of failure of the total number of assessed x points (N_{MCS}).

$$P_f \approx \hat{P}_f = \frac{1}{N_{MCS}} \sum_{i=1}^{N_{MCS}} I_F(x_i) \tag{1}$$

And for MCS, the coefficient of variation (CoV) of this probability, for $P_f > 0$, is given by

$$CoV_{P_f} = \sqrt{\frac{1 - P_f}{N_{MCS} P_f}} \tag{2}$$

As the number of samples N_{MCS} increases the approximation given by $\hat{P}_f$ is increasingly more accurate, $\hat{P}_{f_{N_{MCS}\to\infty}} \to P_f$. It is straightforward to understand that the larger the N_{MCS}, the larger the associated analysis cost will be. Commonly, for complex problems where $g(x)$ is costly to evaluate, it is not feasible to use MCS.

The idea of using Kriging models is then to surrogate the performance function $g(x)$, using a metamodel $G(x)$.

2.1. Kriging metamodeling

A metamodel is a generic description that includes different types of models. These can be usually described as a model of a model. The direct benefit of their application is the possibility of limiting the number of times the original, costly to evaluate, model needs to be evaluated in the analysis.

The idea of using Kriging models for reliability analysis is then to create a surrogate of the failure evaluation procedure as discussed. Furthermore, definition of a Kriging model $G(x)$, an approximation of $g(x)$, allows using the capability of the Kriging to perform as a self-improving function, for which a measure of improvement can be defined.

The main basis to use a Kriging model is to approximate a true state function $g(x)$ that depends on $x \subset \mathrm{IR}^d$, in a d dimensional space, with an approximate mathematical model $G(x)$ that considers uncertainty in the approximatio. Assuming that the true function $g(x)$ can be defined $\forall x$ the process of defining $G(x)$ demands a sample of k support points or observations to be defined and that are usually designated as DoE; $DoE = [\boldsymbol{X}, \boldsymbol{Y} = = g(X)]$ being $\boldsymbol{X} = [x_1, x_2, \cdots x_k]$ a vector of realisations of x and $\boldsymbol{Y}$ the respective true evaluations of $g(x)$ at $\boldsymbol{X}$.

Using a Kriging surrogate model, the true response function $g(x)$ can then be approximated as

$$G(x) = f(\boldsymbol{\beta}; x) + Z(x) \tag{3}$$

$$f(\boldsymbol{\beta}; x) = \beta_1 f_1(x) + \ldots + \beta_p f_p(x) \tag{4}$$

where $f(\boldsymbol{\beta}; x)$ is a function determined by a regression model with p ($p \in \mathbb{N}^+$) basis trend functions $f_p(x)$ and p regression coefficients β to be defined by the known sample $\boldsymbol{X}$; while $Z(x)$ is a Gaussian stochastic process with zero mean that relates to a covariance matrix

$$\mathbf{C}(x_i, x_j) = \sigma^2 R(x_i, x_j; \theta) \qquad with\ i,\ j = 1, 2, 3, \ldots, k \tag{5}$$

the covariance matrix $\mathbf{C}$ relates generic $\boldsymbol{X}$ points using; σ^2 which is a scale parameter called constant process variance and a correlation function $R(x_i, x_j; \boldsymbol{\theta})$.

For the structural analysis, $\mathbf{C}$ is assumed to be stationary and to take the so-called, and widely applied in computer experiments [28], *separable* form in Eq. (6). Other types of correlation can be applied [29,30].

$$R(x_i, x_j; \theta) = \prod_{i=1}^{d} R(h_i; \theta_i),\ \ \theta \in \mathbb{R}^d \tag{6}$$

The correlation function depends then on $h = [h_1, \cdots, h_d]$, a set of incremental values of $x - x_i$ type and $\boldsymbol{\theta}$ hyperparameters.

For a given sample of support points the problem of prediction can then be solved through a generalised least squares formulation, where the estimators for $\boldsymbol{\beta}$ and σ^2 depend on $\boldsymbol{\theta}$ and are given by the following equations.

$$\boldsymbol{\beta} = (\mathbf{F}^T \mathbf{C}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{C}^{-1} \boldsymbol{Y} \tag{7}$$

$$\sigma^2 = \frac{1}{k} (\boldsymbol{Y} - \mathbf{F}\boldsymbol{\beta})^T \mathbf{C}^{-1} (\boldsymbol{Y} - \mathbf{F}\boldsymbol{\beta}) \tag{8}$$

$\mathbf{F}$ is the k x p regression matrix in which the rows are trend functions $f_p(x)$ evaluated at the k support points. A Maximum Likelihood formulation can be used to optimize the likelihood of the observations $\boldsymbol{Y}$. The likelihood function can then be maximized by minimizing the opposite natural logarithm,

$$\theta = \arg\min(-\log L(\boldsymbol{Y}|\boldsymbol{\beta}, \sigma^2, \theta)) \tag{9}$$

$L(\boldsymbol{Y}|\beta, \sigma^2, \boldsymbol{\theta})$ is the likelihood function based on the assumption that $\boldsymbol{Y}$ points follow a Gaussian distribution.

A prediction for the true realisation $g(u)$ in a point u in the space is then given based on the Kriging expected value μ_G and variance σ_G^2:

$$\mu_G(u) = f(u)^T\ \boldsymbol{\beta} + \boldsymbol{c_v}(u)^T \mathbf{C}^{-1} (\boldsymbol{Y} - \mathbf{F}\boldsymbol{\beta}) \tag{10}$$

$$\sigma_G^2(u) = \sigma^2(1 + D(u)^T(\mathbf{F}^T\mathbf{C}^{-1}\mathbf{F})^{-1}D(u) - \boldsymbol{c_v}(u)^T\mathbf{C}^{-1}\boldsymbol{c_v}(u)) \quad (11)$$

$$D(u) \equiv \mathbf{F}^T\mathbf{C}^{-1}\boldsymbol{c_v}(u) - f(u); \quad (12)$$

where $\boldsymbol{c_v}(u) = c_v(u, x_i),\ i = 1, 2, \cdots, k$ is the correlation vector that relates the realisation to be evaluated with the known points and $f(u)$ is the vector of trend functions evaluated at u. $D(u)$ is introduced for the sake of brevity.

If computational time is a constrain, it is not efficient to randomly select the DoE. In fact, in order to fully exploit the Kriging model, an heuristic measure of improvement should be established. Therefore, when metamodeling with Kriging for reliability analysis it is common to use a criterion to select new points in x; this criterion is frequently denominated infill criterion and is a feature of major interest to improve computational efficiency. As highlighted previously, the active learning procedure to select new points in a set of c candidate points is commonly denominated as enrichment.

A Gaussian correlation function and a constant trend function are applied in the further implementations.

2.2. Infill criterion

It was seen that different infill criteria have been defined since the establishment of Kriging models in reliability analysis [1,3,9–11]. Picheny et al. [31] highlighted the importance of having a notion of improvement when researching the noisy Kriging.

The common approach is then to define a search function $S(x)$ that selects the X_{n+1} to be included in the DoE from a set of c candidates.

2.2.1. Expected feasibility function

In the context of defining $S(x)$, the EGRA algorithm [3] uses the EFF, built on contour estimation of [32]. The EFF is one of the most established enrichment techniques in active learning methods for reliability that use Kriging. It has been applied in different frameworks.

Its function is defined such that a certain contour a (for reliability $a = 0$) is searched in the design space using the expectation $E[\cdot]$ defined as,

$$E[F(x)] = \int_{a-\epsilon}^{a+\epsilon} (\epsilon(x) - |a - G(x)|) f_G \, dG \quad (13)$$

with $G(x)$ prediction being given by the Kriging $\mu_{G(x)}$ and $\sigma_{G(x)}$, respectively mean and standard deviation predictions, and $\epsilon(x) = 2\sigma_{G(x)}$ defining the neighbourhood for the integration and search. This integral can be expressed as

$$\begin{aligned} E[F(x)] = \ & \mu_{G(x)}\left[2\Phi\left(\frac{-\mu_{G(x)}}{\sigma_{G(x)}}\right) - \Phi\left(\frac{-\epsilon-\mu_{G(x)}}{\sigma_{G(x)}}\right) - \Phi\left(\frac{\epsilon-\mu_{G(x)}}{\sigma_{G(x)}}\right)\right] \\ & - \sigma_{G(x)}\left[2\phi\left(\frac{-\mu_{G(x)}}{\sigma_{G(x)}}\right) - \phi\left(\frac{-\epsilon-\mu_{G(x)}}{\sigma_{G(x)}}\right) - \phi\left(\frac{\epsilon-\mu_{G(x)}}{\sigma_{G(x)}}\right)\right] \\ & + \epsilon\left[\Phi\left(\frac{\epsilon-\mu_{G(x)}}{\sigma_{G(x)}}\right) - \Phi\left(\frac{-\epsilon-\mu_{G(x)}}{\sigma_{G(x)}}\right)\right] \end{aligned} \quad (14)$$

with $\Phi(\cdot)$ and $\phi(\cdot)$ being respectively the standard cumulative and densities functions. By construction it is possible to infer that the EFF will be large when $\mu_{G(x)}$ is close to a or $\sigma_{G(x)}$ is large.

2.2.2. U function

An alternatively widely applied efficient infill criterion for reliability analysis is the U function introduced in [1]. This function, specified for reliability analysis, selects new points in the DoE using a ratio that characterizes the probability of having misclassified the candidate points. It is defined as

$$U(x) = \frac{|\mu_{G(x)}|}{\sigma_{G(x)}} \quad (15)$$

a ratio of the mean and standard deviation. This function is directly related to the probability of misclassifying a points in the x space, considering the limit-state condition is set for $g(x) = 0$.

The selected candidate is the point that minimizes the U function. Strong candidates for enrichment are expected to occur in two circumstances: either $\mu_{G(x)}$ is close to 0, or $\sigma_{G(x)}$ is large. One of the notorious characteristics of the U function is its relative simplicity when considering that it has proven to be highly efficient.

Other examples of $S(x)$ can be identified in the literature, such as the LIF [10], the REIF [11], and the entropy-based search function H [9]. In the present work the interest is to discuss a framework that uses density scanning for reliability calculations applied with the U function. The U function is directly related to the probability of misclassification, and expectation of inaccurate predictions, and hence is of interest for density scanning. Nonetheless, the density based framework presented may be applied with any of the alternative learning functions.

3. Application of density scanning of candidates

Candidates in an AK procedure present some spatial correlation between them. Sun et al. [10], Wen et al. [13] identified such behaviour when tackling the problem of partitioning the space of candidate points in clusters, and when pursuing to classify points in x that would also enable the classification of the nearest neighbours in the space of variables.

It is common for candidate points to happen in dense clusters of some size. Even considering that the enrichment procedure is highly dependent on $S(x)$, as $S(x)$ is frequently based on $G(x)$, strong candidates for enrichment are likely to occur close to each other. Hence, in general, candidate points (for the enrichment functions identified) are expected to cluster more densely near $G(x) = 0$ or where there is larger uncertainty on $\sigma_{G(x)}$. As a result, grouping points in the x space as function of the point densities is highly related to functionality of the AK problem.

The k-means clustering algorithm has been applied previously to successfully cluster candidates in the space [13]. One of the limitations of the k-means clustering algorithm is the requirement to define *a priori* the number of clusters that are to be formed in the space, which results in, despite significantly reducing the number of iterations of the AK procedure, a significant increase of the number of $g(x)$ evaluations required in the AK procedure [15].

In the present paper, application of density-based spatial scanning and clustering, here denominated simply as density scanning (ds), is researched. The fundamental idea behind density scanning is to merge the c candidates function of their spatial densities. Notwithstanding, density scanning can be exploited in diverse forms.

The algorithm of [33] (dbscan) is applied for scanning, in an adapted form for AK classification. Despite being relatively new, the density scanning and clustering technique has captivated significant relevance since its introduction in computational experiments. Points in densities are characterized using a measure of distance ζ, and a threshold for the creation of clusters (Minimum number of points in a cluster - MinPts). Fig. 1 presents an example of the density-based clustering procedure.

The radius given by ζ defines the neighbourhood of a reference point (circles). If in its radius there are more than MinPts, a point is classified as a core point in the cluster; on the other hand, if there are not sufficient MinPts in the neighbourhood but there are adjacent points that are core points of a cluster, then the point is reachable by this cluster, and is part of its density. Finally, a point is an outlier if not reachable by the core points of any adjacent cluster. Selection of MinPts commonly adopts a rule of thumb that is to consider the dimension of $\mathbb{R} + 1$, or $d + 1$. It is noted that ζ is the parameter of most relevance for the density algorithm, it will define the neighbourhood to merge densities. [33] proposes an heuristic to define the minimum value of ζ. In the case of the AK methodologies, it is proposed for ζ to be selected as the minimum value that minimizes the number of outliers and does not compromise the definition of separate clusters. In this way ζ can be

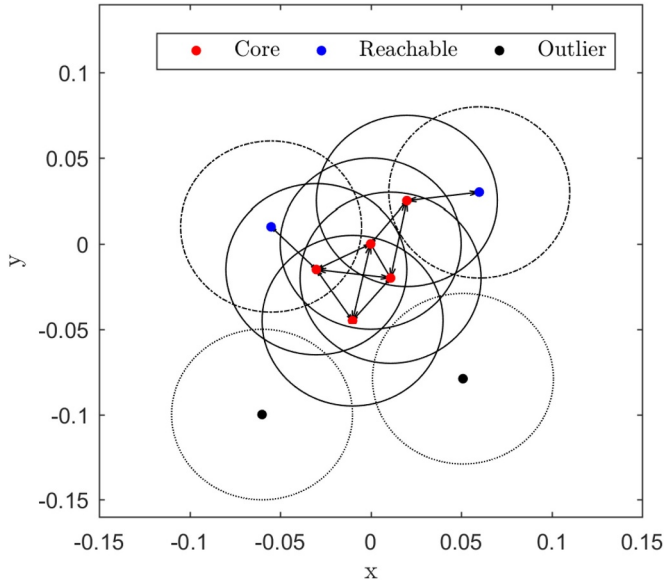

Fig. 1. Example of dbscan clustering. ζ is the radius of the circles that characterize the neighborhood of the classified points.

calculated in an adaptive scheme using,

$$\zeta = \min[\max(\boldsymbol{D});\ D_{c_k}]; \qquad \boldsymbol{D} = [D_{i=1}, \cdots, D_{i=c}] \tag{16}$$

where $\boldsymbol{D}$ is a vector of the minimum adjacent point distance in the c candidate points,

$$D_i = \min(\sqrt{\sum_{j=1}^{d} (X_i^j - X_{h\neq i}^j)^2}), \quad i = h = [1, \cdots, c], \quad j = [1, \cdots, d] \tag{17}$$

and D_{c_k} is the estimated minimum distance between the centres of clusters evaluated using representative samples of the most dense regions during the build-up of the algorithm. One of the advantages of using $\boldsymbol{D}$ and D_{c_k} to establish ζ is that these measures are intrinsically calculated in the density scanning, and as a result, no new measure of distance needs to be built to define $\boldsymbol{D}$. $\boldsymbol{D}$ dominates the learning procedure in almost every single search. Consideration of D_{c_k} is only required for the cases where there are isolated outliers occurring in the x space that are far from the nearest core point. If outliers are identified, these should be enclosed in the group of the closest point in the c candidates classification. Alternatively than classifying outliers using the nearest reachable, outliers can be also treated as a class to be analysed. It is noted that this may increase the number of performance function evaluations needed in the case of reliability analysis. Therefore, as a class, it may be of interest to consider outliers as a low-density class, where an outlier is added to the DoE when it is a more adequate candidate (in the light of the learning function) than candidates from classified clusters. In the present work, indexing outliers using the nearest reachable or core, results in accurate cluster separation for reliability analysis. It is emphasized that, for robust exploration, no potential candidates (attending to the search function applied, and to guarantee its adequate performance) should be left out of the learning process. Further details on the density clustering approach are presented in the following section.

On a first instance using density scanning and clustering allows the prioritization of areas of the DoE where misclassification is expected to enclose significant uncertainty in the definition of P_f, having significant contribution to the convergence of it, and where the enrichment is expected to suppress a large number of candidate points. Also, the same technique can be used to prioritize points in the DoE enrichment, and in particular, to avoid selection of points that are so close to each other that have little contribution to improve $G(x)$'s accuracy. Section 3.3 presents the frameworks implemented to exploit the synergies between density scanning and the active learning that uses AK.

Despite being able to efficiently manage large datasets [34], the main disadvantage of density clustering in relation to the aforementioned k-means is its computational demand. It shares a common feature to all algorithms that depend on euclidean distances, that is, it suffers from the curse of dimensionality. In the following section a subset-cross classification framework is proposed in order to accelerate density scanning and clustering in AK applications.

3.1. Cross-subset classification framework

In order to increase the efficiency of the density scanning, and considering that *a priori* information exists about the general problem of reliability analysis with AK, a subset approach to clustering is proposed. It consists in using a subset division of the main candidates to perform density scanning and then, cross-classification using a fictitious sample built from the subset division. It is of interest to scan when large candidate samples occur.

This approach is possible for reliability classification because candidate points are expected to occur in large quantities close to the region of $G(x) = 0$ and large σ_G. Additionally, these regions are expected to be the most relevant for the AK prediction accuracy.

Let L_v be v subsets of the c candidates, with c being the total number of candidates. The density classification for each L_v will originate l dense clusters characterized by their C_i classes, with $i = 1. .l$.

Then an additional subset L_+ can be defined by selecting L_{c_i} points from each C_i class of the L_v subsets. L_+ is additional cross-classification sample that can be classified in j classes and used to associate the subset classes of the selected L_{c_i} points, classifying the c candidates according to a cross classification scheme with

$$C_i[L_{c_i}] = C_j^*[L_{c_i}] \tag{18}$$

C_j^* being the indexed class of the L_+ sample. The remaining points of the C_i class are classified accordingly to the indexed C_j^* classification of its associated L_{c_i} referenced for cross classification in L_+. In the occurrence when more one class is identified in $C_j^*[L_{c_i}]$, then $C_i[L_{c_i}]$ should be classified according to the most prevalent (repeats more often) $C_j^*[L_{c_i}]$ class. This may only occur if the subset samples are taken in the boundaries of a cluster, which is highly unlikely. Also, it may occur that a very small cluster is identified in less dense regions, however for it to be previously associated with a class C_i, it means that some bridge to this cluster was already established previously in L_v and that these points are in fact part of a close cluster.

Fig. 2 presents the scheme of the approach to cross-classify the entire c candidates using smaller subsets, and hence, enabling classification of large candidate samples.

In Fig. 2 it is possible to see in the left side that the algorithm is started with a candidate population. If relatively small, clustering can be directly applied. If large, this population is divided in v subset samples L_v. These L_v samples are classified individually, originating C_i clusters for each subset. The information about C_i is kept to be compared in the top node that merges C_i and C_i^*. A representative sample (L_{c_i}) is then selected from each of the L_v classified C_i. This sample will be

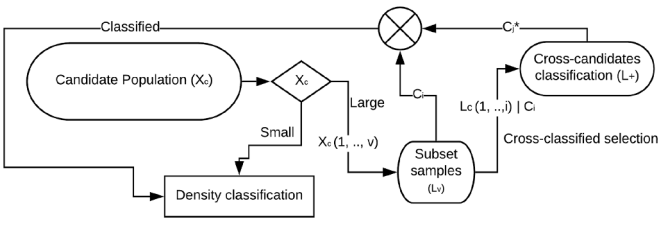

Fig. 2. Example of methodology used for fast-classification of densities according to dbscan and using L_v samples to build a cross classification scheme.

used to build a cross-classification subset (L_+), which is then classified in C_j^* classes. By merging the information of C_j^*, with the existing knowledge about the classification of L_{c_i}, the identified C_i classes are re-classified and a single homogeneous classification is produced for X_c.

In order to define both L_v and L_+ a minimum representative subset sample needs to be selected. In Fig. 2 it can be seen that two sizes v and L_{c_i} need to be selected to divide the c candidates and to subset the C_i classes.

Sheskin [35] provides a comprehensive description on how to select a minimum representative sample size in order to subset populations of candidates. It is noted that above a minimum sample size for the subsets the increase in their representativeness of the population is negligible [36]. In the present implementation, subsets are classified using a very representative subset defined with the nearest integer to 10^4 resulting from the candidate population division, and L_+ is defined so that its size does not surpass an increase of 100% of this value. This allows for a fast computation, without compromising the accuracy of the scanning.

An example of application of cross-classification is presented in Fig. 3.

It is possible to infer that in the original candidate sample the density clustering is correctly classified through the division of the c candidates in three smaller subsets. While L_1 and L_2 are classified according to L_+, the same does not occur for L_3. In L_3 a small cluster is formed in the top of the green cluster with outliers on the bottom (outliers are commonly marked as class 0, and in the case of the subset samples are transferred to L_+ where they are clustered). By using subset samples of C_i the Cross L sample is built and classified. This small cluster (left top purple cluster) is re-classified in L_+ as a part of the big green cluster. The same occurs with the outliers (black circles) in L_3. It is noted that all the points in the procedure should be classified as part of a cluster in the cross-classification (even if not reachable by a core), as no new classifications are performed. This classification is transferred backwards to the L_v sample, such as in the case of L_3. One of the limitations of this approach that uses reachable points to classify outliers in L_+ is the possibility of indexing a point that is isolated from all the existing clusters. However, this is of relevance for efficient exploration of the space (a potential isolated candidate may unlock new search regions in x). Cross-clustering is of interest when X_c is relatively large (e.g. comparative results and limit of direct scanning of Fig. 4.)

Only potential candidate points should be considered in the scanning procedure, *i.e.*, points that have some likelihood of being selected. When using the U criteria these are the points with $U < 2$. Performing calculation with points which are not "true" candidates and do not influence the choice of neighbours has no interest.

In terms of computational effort, the classification of subsets is considerably faster and requires less computational power to be performed. Fig. 4 compares both the scanning times of cross-classification with direct scanning of c candidates.

While for a low number of c, in the vicinity of 10^4 c, the direct scanning is faster, with the increase of c the time required for the direct scanning increases much faster for direct classification than cross-classification. It is interesting to note that the dimension of the space d has limited impact in the times demanded for scanning, however, increasing d increases the computational (power) demand of the analysis. The fact that it was not possible to perform scanning by $c \geq 5 \times 10^4$ for a Dual Xeon CPU with 64 GB of RAM is indicative of the significant memory needs that are required for direct clustering.

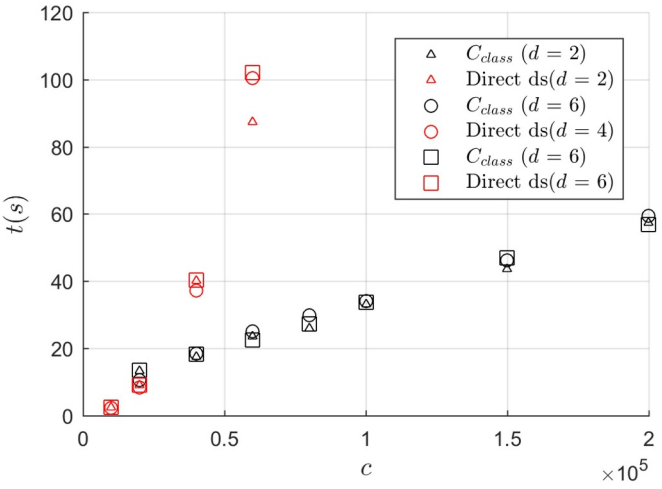

Fig. 4. Example of scanning time using cross-sampling and direct scanning for different number of candidates c in the creation of two clusters originated by standard normal distributions with origins at (-5, -5) and (5,5). Direct scanning for c values larger or equal to 1×10^5 was not possible due to computational memory limitations.

Further mitigation of the time demanded for scanning can be attained through optimization of the sample sizes. It is important to highlight that c is significantly smaller than the MCS sample used for AK learning and that the time for density scanning may be neglected in comparison to the time that is demanded to evaluate a complex engineering structure or system (*e.g.*, with Finite-Element-Methods). The same approach may be applied with other alternative algorithms widely applied in AK procedures, such as the k-means, enabling the application of larger sample sizes in the AK calculations.

3.2. Stopping criterion

The efficiency of the Kriging implementation is highly dependent on the stopping criterion applied to evaluate if convergence was attained in the active learning procedure.

Different learning functions use different stopping criteria, and different criteria may are applied for the same learning function. The EFF procedure commonly applies a minimum threshold for the improvement with the EFF, commonly of 0.001. The AKMCS with the U function assumes an upper threshold in the confidence of the meta-model classification, which is commonly given by a U of value 2. Gaspar et al. [19] showed this criterion to be conservative for reliability

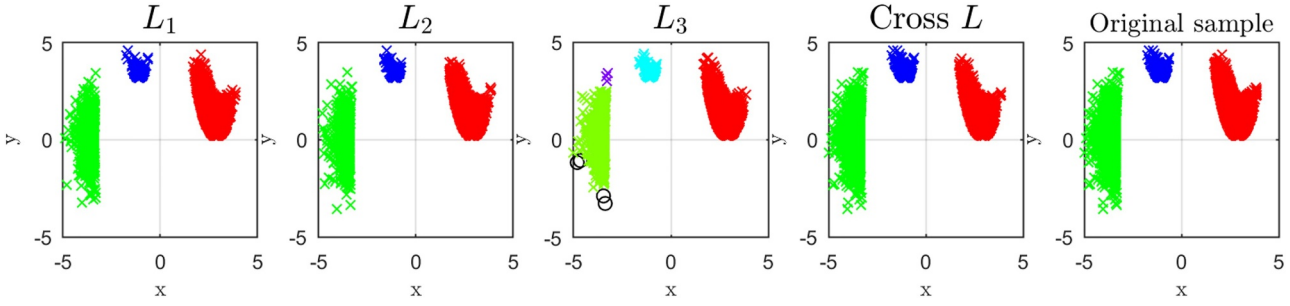

Fig. 3. Example of using three samples to build a cross classification scheme. The different colours represent different clusters in each subset. Three samples are used, a cross validation sample is built which is then used to classify the entire original sample (size of 30000 candidate points). Cross L refers to the L_+ sample.

calculations. Lelièvre et al. [15] recently proposed a less conservative U misclassification stopping criterion that uses a joint probability of misclassification.

Sun et al. [10] uses a criterion of expectation enclosed in the amount of probability left to address. Jian et al. [27] further discusses this criterion by constructing two measures to assess the accuracy of $G(x)$ in relation to $g(x)$, of particular interest for reliability analysis. For the expectation of probability enclosed in misclassification the expected error in the $G(x)$ prediction can be given by

$$E[P_m] = \int_x \Phi(-U(x)) f(x) dx \tag{19}$$

with P_m being the Probability of misclassification, $f(x)$ the joint density function of the x, and Φ refers again to the standard normal distribution function. In this context, two type of errors may occur in the classification, 1: Misclassified $g(x) < 0$; or 2: Misclassified $g(x) \geq 0$. The first corresponds to the cases where the unknown $g(x) < 0$ but $G(x)$ prediction is positive, and the second for the case of $g(x) \geq 0$, but $G(x)$ is negative. For each,

$$P_m^1 = \int_x \Phi(-U(x) | G(x) < 0) f(x) dx \tag{20}$$

$$P_m^2 = \int_x \Phi(-U(x) | G(x) \geq 0) f(x) dx \tag{21}$$

constitute indicators of convergence of $G(x)$ to $g(x)$ that can be approximated for AK with MCS as

$$P_m^{1,2} = \frac{1}{N_{MCS}} \sum_{i=1}^{N_{MCS}} \Phi(-U(x_i) | G(x)) \tag{22}$$

conditional on whether $G(x) < 0$ or $G(x) \geq 0$ respectively for 1 and 2. x_i refers to the x part of the MCS. The two measures will evaluate the expectation of the assessed $\hat{P}_f$ to increase or decrease. An interval of expectation of error can then be applied to evaluate the convergence of the results

$$\frac{P_m^1 + P_m^2}{\hat{P}_f} < \eta \tag{23}$$

with η being an expectation of error enclosed in the remaining c candidates and that can be defined using, *e.g.* a ratio of P_f or using $CoV_{\hat{P}_f}$, and that can be established for the expected positive and negative deviations of $\hat{P}_f$. The division of 1 and 2, can be applied to identify the expected sensitivity in the P_f prediction. Jian et al. [27] discusses and provides an upper limit for the error in $\hat{P}_f$ resulting from the application of this criterion.

It is noted that this stopping criterion has a relevant synergy with density scanning. Dense regions of x enclose more expectation, and enriching the DoE should weight on the direct influence of the candidate to the problem of reliability, more specifically, the accurate assessment of P_f. Recent works of [11,37] research on a functional approach to AK procedures by applying convergence criterion that directly relate to the P_f estimation.

Zhang et al. [38] discusses exploitation and exploration when considering η as a stopping criterion and proposes a complementary criterion to evaluate exploration, and to enhance robustness of the learning procedure. In the present example, as all the points were enclosed in the learning process and the scanning allowed the improvement of the convergence to P_f. A reference η of 0.01 (using half of this value for each P_m) as the ratio of expectation of error in P_f was identified to produce robust results in this regard.

This referred synergy of density scanning with the error in P_f is of major relevance for the present application and is exploited in the following section. Three main approaches that apply density scanning and clustering in AK are researched.

3.3. Density scanned active learning framework

It was previously highlighted that the combination of density scanning with AK can generate different frameworks due to the flexibility of the algorithms. It can be combined in sequential enrichment with parallel and non-parallel selection strategies.

Three main applications of density scanning are discussed in the present work. The dsAK, AK with a density scanning and individual selection of points; the dsAKP, AK with density scanning and sequential parallel enrichment; and finally, the ds^{2}AK, that uses either of the previous methodologies but that groups candidates with the DoE. While the first two approaches are straightforward to understand, the last consists in clustering candidates with points already existing in the DoE in order to limit the enrichment with points that are within ζ of an already existing point in the DoE.

The approaches implemented in the present paper go as follows (S for serial, P for parallel and ds^2 for the selection with DoE scanning are used to distinguish these):

1 - Initiate the DoE and the selection space x (Latin Hypercube Sampling (LHS), and MCS, respectively);
2 - Fit $G(x)$;
3 - Predict P_f and P_m;
4 - Evaluate the stopping condition(s). If not fulfilled continue, or else, stop the active learning;
5 - Compute the measure $S(x)$ (in the present implementation th U-function is applied) and select a subset of c candidates;
6 - Classify the c candidates in accordance to density scanning;
7 - Calculate $P_m[C_i]$ as the $E[P_m]$ at each dense cluster;
8 (S - dsAK) - select the most suitable candidate in accordance to $S(x)$ within the dense cluster that encloses the largest $P_m[C_i]$;
9. (P - dsAKP) - select the most suitable(s) candidates in accordance to $S(x)$ from each of the C_i clusters;
10. (ds^2) - select the most suitable candidate (in application of 8S or 8P) not clustered with the current DoE within a ζ value;
9 - Enrich the DoE and repeat procedure, return to 2;

Step 8 distinguishes the three different frameworks. For example, if ds^2AKP is applied, then parallel computations are enabled with exclusion of candidates clustered in density with the current DoE. At the first iterations, P_f is likely to be 0, no clusters are formed, and enrichment should use $S(x)$.

By having density clustering in active learning the enrichment is encouraged in the dense "pools" that enclose major contribution to P_f, under the assumption that classifying a point in these "pools" of points will contribute improve the classification of the neighbourhood. The presented approach increases the capacity of the active learning to converge with less iterations in complex limit-state functions, and to limit the additional number of evaluations demanded to use parallel computing.

4. Examples of application

Three examples that use complex $g(x)$ functions are discussed as a reference of implementation for the framework(s) proposed. It is noted that for all the cases considered only "true candidates" in the light of the stopping criterion are considered in the active learning framework that uses density scanning. Without downsizing the sample to the so-called "true candidates", density clustering has no interest (enclosing all points forms a single dense cluster). The U function is applied as $S(x)$ for the representative example of application of a scanning framework, and therefore, for it, "true candidates" are the points with $U < 2$.

Since its introduction, the U-function has become very relevant and widely applied in the field of AK. Nonetheless, it is important to highlight that density scanning is expected to have synergy with other $S(x)$ that depend on $\mu_G(x)$ and $\sigma_{G(x)}$. In such cases, dense clusters of "true

candidates" are expected due to similarities of $\mu_G(x)$ and $\sigma_{G(x)}$ in neighbour points in x, e.g., as in the case of the U-function. In order to accelerate clustering other measures can be selected to estimate the candidate sample, e.g., selecting percentage of more likely candidates. Unless otherwise indicated, the AK initial N_{MCS} sample size was equal to the g_{eval} of the MCS evaluation.

4.1. Bivariate non-dimensional performance function

The first example analysed is a non-linear bi-variate performance function introduced in [3], in the form applied in [11]. Both x_1 and x_2 variables follow a standard Gaussian distribution.

$$g(x) = q - \frac{1}{20}(x_1^2 + 4)(x_2 - 1) + sin\left(\frac{5}{2}x_1\right) \quad (24)$$

The $g(x)$ for this function is presented in Fig. 5, along with the formation of clusters within three iterations of the dsAK procedure. It is possible to infer that, due to the non-linearity of $g(x)$, isolated "pools" of candidates occur in the space. The scanning framework is responsible for identifying these and selecting the ones to be exploited. If the parallel approach is implemented, a point from each "pool" is considered to enrich the DoE. The scanning framework discussed is able to divide the space such that the active learning may consider to perform parallel computations without external inputs (such as the number of clusters), and only if dense "pools" are formed. The partition of the space is automated in every iteration in order to reduce the number of non-optimum points taken at each step. It is noted that the fact that every point is part of a cluster may be of relevance within the U function to foment robust exploration of the space.

Benchmarked results from [11] are presented and compared with the AKMCS approach that uses density scanning in Table 1.

It is possible to infer that application of density scanning contributes to decrease the number of evaluations demanded for the learning approach without significant trade-off in the P_f prediction accuracy. The larger error in the P_f prediction is mostly due to the stopping criterion, which allows an earlier halt of the learning with less g_{eval}, and taking advantage of the synergy with the scanning framework that prioritizes dense clusters, providing complementary knowledge of the AK procedure and fomenting the convergence to P_f.

The parallel approach allows the reduction of the average number of iterations, from 19.4 to 10.2 when compared with the serial ds implementation. One particular relevance of density scanning is related to the capability of using parallel computing without compromising the number of g_{eval}. Since all the candidates are evaluated, and cluster separation is unsupervised (does not demand inputs and partition occurs only when strictly necessary), the gain in computational time is very significant regardless of the stopping criteria applied. This is the example of the U function, which may still use the $U > 2$ criteria in a dsAKP framework, with direct gain in the number of iterations required, and with limited compromise in the number of g_{eval}. Further results for the AK with scanning procedure using distinct implementation strategies presented in Table 2, and provide further insight into this capability of the dsAKP.

Table 1
Comparative results for distinct AK implementations benchmarked from [11]. g_{eval} refers to the number of $g(x)$ evaluations. The average results presented in this example used 20 active learning implementations with initial DoE size of 10 LHS points. An η of 0.01 was considered in ds approach. ISKRA is the Improved Sequential Kriging Reliability Approach presented in [13], with KB as the Kriging Believer implementation.

$q = 2.0$								
		AKMCS			ISKRA		ds	
	MCS	U	REIF	Serial	KB	k-means	AK	AKP
$P_f\,(10^{-5})$	1.375	1.374	1.379	1.368	1.369	1.376	1.350	1.389
g_{eval}	4×10^7	41.8	45.1	40.0	51.0	73.4	29.4	30.6
$e_r(\%)$	–	0.06	0.28	0.52	0.46	0.08	1.9	1.1

In Table 2 it is possible to infer that the serial approach is able to converge faster to the reference P_f, generating a smaller error for a similar number of iterations, see the example of Fig. 6 where results are presented, considering the same initial DoE, for the case of applying scanning and selection of candidates in the "pool" with largest expectation P_m.

The improvement in convergence for similar number of iterations occurs because large clusters of candidates are likely to be prioritized in the search, as a result of the prioritization of regions that have larger values of $E[P_m]$. After 11 iterations the scanning technique with prioritization of dense regions of candidates achieves a relative error in P_f similar to the error obtained with the non-scanned approach only after 15 iterations. It is noted that if the criterion of $U > 2$ is considered, both methods will approach a similar number of iterations, however, most of the iterations of II will pursuit to classify what would be "outliers" in the context of density scanning (*i.e.,* no relevant dense clusters are formed). It is noted, however, that a value of $\eta = 0.05$, despite providing in average and repetitively a reasonable approximation, does not provide entirely robust results in terms of exploration. Therefore, $\eta = 0.01$ provides an efficient trade-off of computational

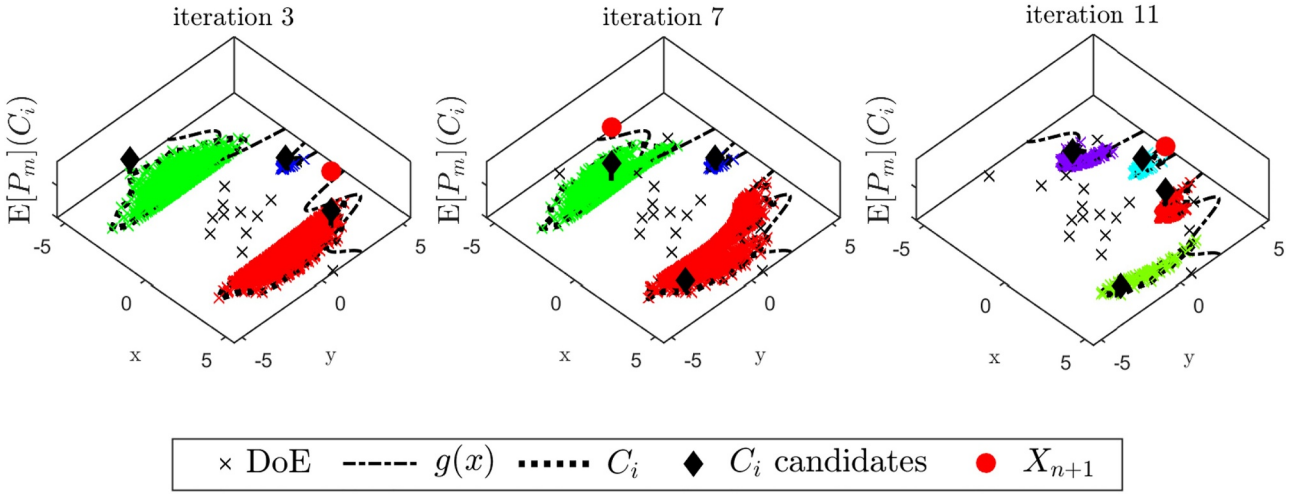

Fig. 5. Dense "pool" formation and selection of candidates in bi-linear function example, with application of *ds*AK. The z axis is augmented for the sake of representation, e.g. in iteration =3 most of the $E[P_m]$ are enclosed in major clusters.

Table 2
Average results for the bivariate non-dimensional performance function, based on values for 50 active learning implementations using the same initial DoE size of 10 LHS points. g_{eval} refers to the number of $g(x)$ evaluations.

$q = 1.6$

Algorithm (stop criterion)	$\hat{P}_f$ ($\times 10^{-4}$)	CoV $\hat{P}_f$	e_r(%)	n_{iter}	CoV n_{iter}	g_{eval}	CoV g_{eval}
MCS	1.714	*	–	–	–	4×10^6	–
AKMCS-EFF	1.747	0.04	1.9	27.27	0.05	37.27	0.04
AKMCS-U	1.736	0.04	1.3	26.36	0.08	36.36	0.06
AKMCS-U ($\eta < 0.05$)	1.667	0.11	2.7	18.1	0.14	28.1	0.14
AKMCS-U ($\eta < 0.01$)	1.698	0.09	0.9	20.04	0.12	30.04	0.12
dsAK(U > 2)	1.717	0.05	0.1	26.22	0.12	36.22	0.09
dsAK($\eta < 0.05$)	1.660	0.11	3.3	13.76	0.16	23.76	0.09
dsAK($\eta < 0.01$)	1.691	0.06	1.4	17.36	0.11	27.36	0.07
ds^{2}AK-U($\eta < 0.01$, $\zeta = 0.1$)	1.750	0.04	2.0	16.73	0.06	26.73	0.04
dsAKP(U > 2)	1.736	0.04	1.3	13.68	0.15	36.84	0.06
dsAKP($\eta < 0.05$)	1.686	0.08	1.7	7.4	0.16	27.87	0.08
dsAKP($\eta < 0.01$)	1.723	0.05	0.6	9.0	0.13	29.3	0.08
ds^{2}AKP ($\eta < 0.01$, $\zeta = 0.1$)	1.725	0.05	0.6	8.87	0.13	30.87	0.09

time and average accuracy and will be applied in the remaining examples. It is also noted that η robustness for 0.05 may depend on additional considerations, such as the MCS sample size. The criteria of [27] or [38] may be applied to complement the search procedure in this regard.

In the case of parallel computation of clusters, the total number of evaluations is similar to the AK using both U and η criteria. Two clusters may be formed very close to each other which may provide similar information about the enrichment, such as in Fig. 5 - iteration 7 case. Nonetheless, as highlighted the scanning is able to produce efficient unsupervised partitions in the space of candidates maintaining a reasonable amount of g_{eval}. Also, using the $U > 2$ criterion permits the reduction of the number of iterations by almost 60%, using a similar number of g_{eval}. The dense 'pools" that are formed are usually left to be addressed until the U function starts their exploitation. Therefore, dsAKP allows to accelerate both, exploitation of these, as well as, exploration of the candidate space (more than one region is addressed in a single step).

The ds^2 further improves the computations. The U function may tend to compute close points (in particular due to $\mu_{G(x)}$) in the DoE when enriching x. If these are very close, such as in the vicinity of $\zeta = 0.1$, they may provide redundant information in relation to $g(x)$

4.2. Example 2: series system

A series system is researched in the present example. This function was previously researched for reliability calculations in *e.g.*, [1,27,37]. In the presented example the (m, k)-dependent function applied in [11] is researched. Its performance function is described by the following system of equations.

$$g(x) = \min \begin{cases} g_1(x) = k + 0.1(x_1 - x_2)^2 - \frac{x_1 + x_2}{\sqrt{2}} \\ g_2(x) = k + 0.1(x_1 - x_2)^2 + \frac{x_1 + x_2}{\sqrt{2}} \\ g_3(x) = (x_1 - x_2) + \frac{m}{\sqrt{2}} \\ g_4(x) = (x_2 - x_1) + \frac{m}{\sqrt{2}} \end{cases} \tag{25}$$

Results for the reliability calculations of density scanned AK of the series system in comparison to benchmarked results from [37] are presented in Table 3.

Results show that the usage of η and its direct relation to P_f improves the computation in relation to the upper limit of U. This is notorious in the application of LIF and both density scanning approaches in relation to the remaining AKMCS. The efficiency of the density scan is similar to the FPS framework [37]. Both approaches use a measure of sensitivity

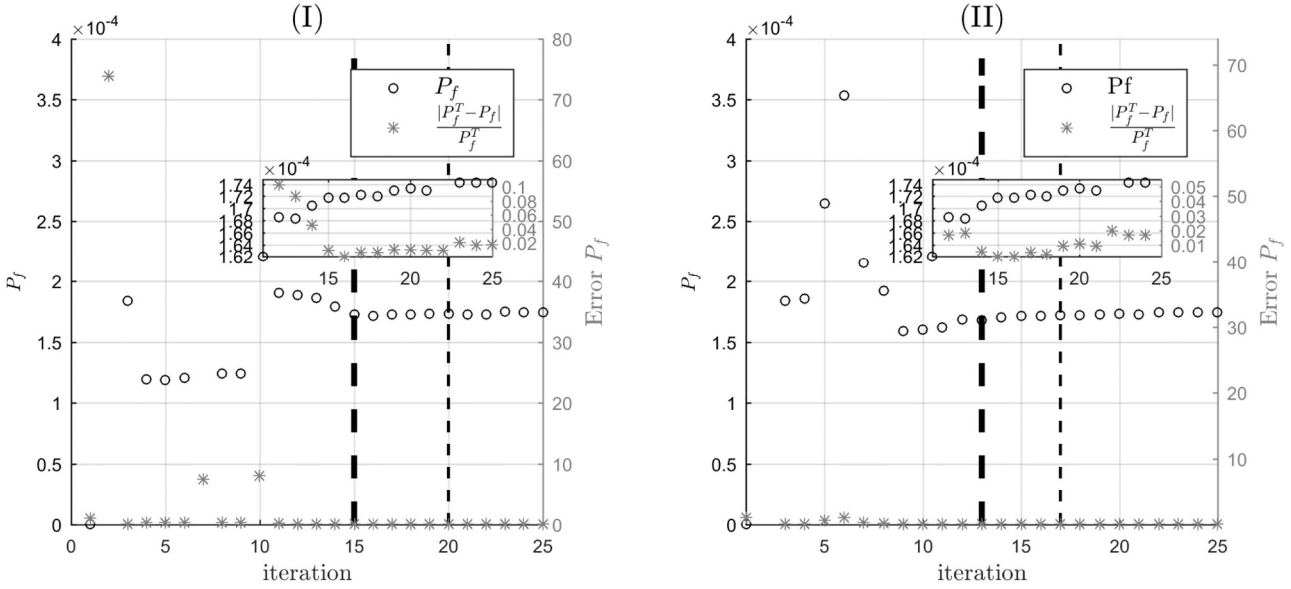

Fig. 6. Example of convergence for the AK-MCS (I) and using density clusters to prioritize the search (II). The black trimmed vertical lines represents the stopping condition for the $\eta < 0.05$ (thicker line) and $\eta < 0.01$. Gray line should be read on gray right axis.

Table 3
Comparative results for distinct AK implementations benchmarked from [37]. RD refers to the introduced response-distance function from the reference work. The sign * indicates that the AK estimation uses an initial sample of candidates of 10^5, which is expected to be also adequate considering the CoV of P_f. g_{eval} refers to the number of $g(x)$ evaluations. In the cases where parallel computing is applied, the number of iterations is presented in brackets.The average results presented in this example used 20 active learning implementations with initial DoE size of 10 LHS points.

$k = 3; m = 6$

		AKMCS				FPS framework							ds	
	MCS	U	EFF	H	LIF	U	EFF	H	LIF	RD	AK	AKP	AK*	AKP*
$P_f\,(10^{-3})$	4.454	4.435	4.475	4.456	4.471	4.423	4.456	4.411	4.497	4.478	4.444	4.412	4.433	4.408
g_{eval}	10^6	106.1	114.0	97.5	64.8	64.7	64.3	69.7	56.5	56.3	55.85	75.7 (48.2)	56.77	60.85 (32.7)
$e_r(\%)$	–	0.43	0.47	0.05	0.38	0.70	0.05	0.97	0.97	0.54	0.2	0.94	0.47	1.2

to prioritize the search (here the amount of P_m), and this may be behind the similarity of their results.

For the AK with dsAKP* the number of iterations required to perform the reliability evaluation was of 32.7, while the AKP approach demanded 48.2 iterations and more g_{eval}. This shows that the size of the initial sample of MCS points may have a large influence on the scanned searching procedure. [13] identified previously this characteristic when discussing an adaptive approach. In the case of the smaller initial sample of candidates there is a smaller probability of generating a circular density of candidate points in this particular limit-state function (see function shape in Fig. 7), which significantly improves the efficiency of the learning procedure. A sample size that results in slightly less than 5% of CoV in P_f was identified to produce efficient learning. Nonetheless, as P_f is unknown *a priori* in a generic implementation, adaptive sample sizes may contribute to significantly increase the efficiency of the AK learning.

Table 4 presents results for the active learning AK analysis of the series system considering different density scanning frameworks.

For the initial sample size considered, formation of dense clusters in the present example is less prevalent after the initial enrichment. This occurs due to the occasional creation of a single circular dense cluster that mitigates the partition of the candidates. Only after exploitation of this large circular cluster, separate clusters are formed again. This results in a lower improvement in the reduction of the number of iterations when compared with the individual enrichment. However, this small reduction of the number of iterations is accompanied by only a slight increase in g_{eval} and, again, may be improved by adaptive sampling procedures. Nonetheless, the scanned approach has the advantage of being able to proceed with parallel computations in an unsupervised implementation, which is of relevance when no knowledge exists about the form of $G(x)$.

The AKds efficiency is also affected by the appearance of dense candidates in a circular formation, and the improvement from its application is less pronounced than in the previous example. On the other hand, application of ds^2 in the present case foments exploration and improves the accuracy of the P_f prediction for the same number of enriched points, see Fig. 7. The U function is likely to sample points that are very close to each other and that may give limited improvement of the $g(x)$ prediction. Fig. 7 shows that for the same number of iterations exploration is fomented by the usage of a density penalty in the DoE enrichment which comes at virtually no cost.

It is possible to infer that, in iteration 30, case I has already identified the four branches of the function while in case II the AK is still exploiting the remaining branches.

4.3. Examples 3: cantilever tube

Finally, a relatively high dimensional problem considering the cantilever beam introduced in [39] is researched, Fig. 8. The cantilever beam problem has been widely studied in research literature, e.g. [13,37].

The performance function of the cantilever tube is given by,

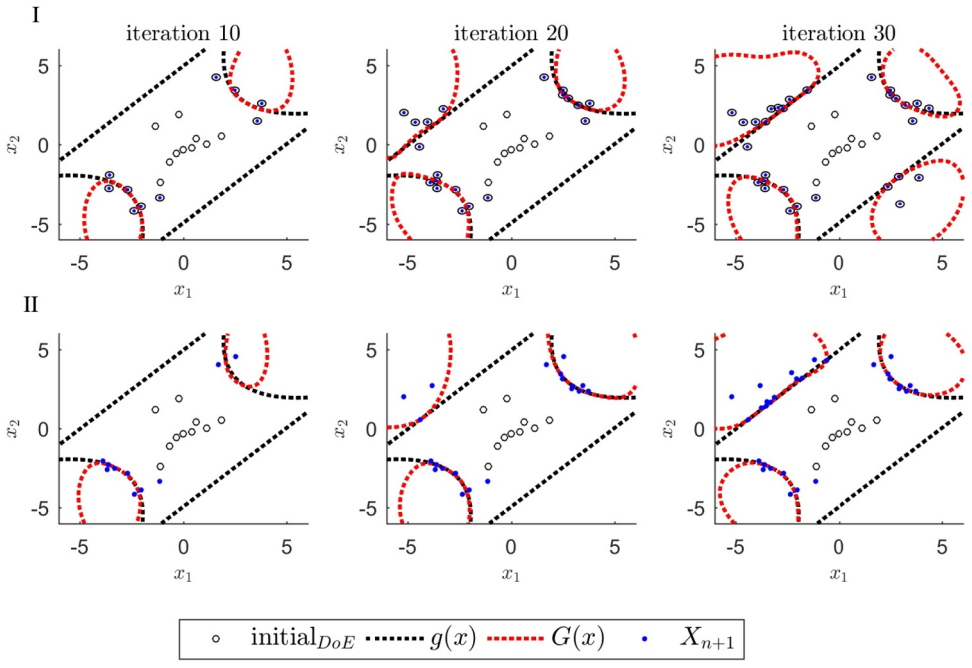

Fig. 7. Example of AK selection procedure for the (I) ds^2AK considering $\zeta = 0.2$; and (II) AKMCS-U.

Table 4
Average results for the series system performance function, based on values for 30 active learning implementations using the same initial DoE size of 10 LHS points. Results of dsAK consider a value of $\eta = 0.01$ to evaluate the convergence the learning procedure. g_{eval} refers to the number of $g(x)$ evaluations.

$k = 4; m = 7$

Algorithm (stop criterion)	$\hat{P}_f$ ($\times 10^{-4}$)	CoV $\hat{P}_f$	e_r(%)	n_{iter}	CoV n_{iter}	g_{eval}	CoV g_{eval}
MCS	5.040	–	–	–	–	2×10^6	–
AKMCS-U	5.036	0.05	0.1	55.20	0.09	65.20	0.06
AKMCS-EFF	5.042	0.04	0.1	67.7	0.08	77.7	0.07
dsAK-U	4.902	0.05	2.8	45.8	0.15	55.8	0.12
ds^{2}AK-U(ζ=0.1)	4.981	0.05	1.2	45.70	0.13	55.70	0.11
ds^{2}AK-U(ζ=0.2)	5.021	0.04	0.4	43.65	0.12	53.65	0.10
dsAKP-U	4.976	0.04	1.3	29.93	0.27	59.50	0.10
ds^{2}AKP-U(ζ=0.2)	5.021	0.04	0.4	32.33	0.21	62.53	0.10

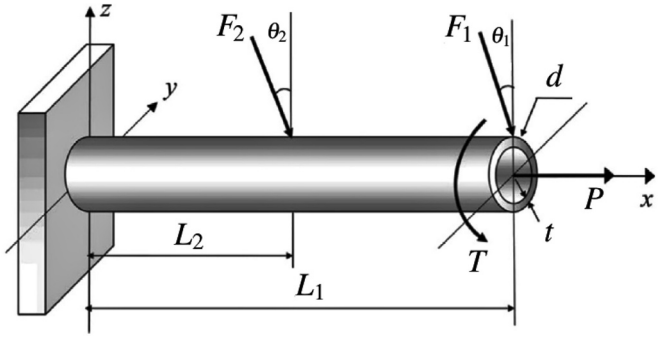

Fig. 8. Cantilever tube, adapted from [39].

$$g(x) = S_y - \sqrt{\sigma_x^2 + 3\tau_{zx}^2} \tag{26}$$

with S_y being the yield strength of the material and σ_x the normal stress, calculated using the following equation,

$$\sigma_x = \frac{P + F_1 \sin\theta_1 + F_2 sin\theta_2}{A} + \frac{Mr}{I} \tag{27}$$

that accounts for the normal stress due to axial forces (first term in the right-hand side) and normal stress due to the applied bending moment M, given by,

$$M = F_1 L_1 \cos\theta_1 + F_2 L_2 \cos\theta_2 \tag{28}$$

Area (A), Second moment of Area (I), and radius(r) and are obtained from the geometric properties of the cantilever.

$$A = \frac{\pi}{4}[D^2 - (D-2t)^2], \quad I = \frac{\pi}{64}[D^4 - (D-2t)^4], \quad r = \frac{D}{2} \tag{29}$$

Finally, the contribution of the torsional stress τ_{xz} is given by,

$$\tau_{xz} = \frac{TD}{2J}, \quad J = 2I \tag{30}$$

In the present example 11 random variables are considered. These are listed in Table 5.

The results for the density scanning implementation are presented in Table 6.

The e_r is larger in average for all the AK procedures studied, which may be related to the complexity of the example (large d). The performance of the AKMCS-U is similar to the AKMCS-EFF. The dsAK with $\eta = 0.01$ requires approximately less 40% of g_{eval} for a similar performance. The parallel computation gain is smaller in the present example, which indicates that there is no prevalence of formation of separate regions of candidates. Nonetheless, the small reduction in the average number of iterations comes at a negligible increase of g_{eval}. It is noted that the application of ds^{2}AK requires an uniform comparison of the size for all the variables, which can be achieved by a transformation of the input variables.

Due to the cross-clustering technique, computational time and power availability was never a limitation in any of the cases, even for a high dimensional problem as the one presented. Moreover, the additional time to perform density scanning is expected to be negligible when compared with the time required to obtain a solution of a complex engineering model, such as Finite Element Models. Moreover, in the particular case of large number of variables, reduction of the n_{iter} and g_{eval} is relevant to maintain the computational demand of the Kriging prediction practicable.

The dsAK demanded in average one third of the computational time required for the AKMCS-U to halt the search procedure, when evaluated in a dual Intel Xeon(R) computer. The cost of predictions of $G(x)$ increases significantly with the increase of the number of support points. It is noted, that in both cases the cost to performing the learning is assumed to be negligible in comparison to the cost of decreasing the size of g_{eval}.

It is of interest, to further improving computational efficiency, the cross classification procedure can be even further optimized, through the application of indexing or R-trees [40,41]. Improvements of the scanning algorithm could remove its large dependence on the distance, and future works should research on potential applications of alternatives in order to improve the implementation proposed (e.g. even faster cluster or larger limits for direct clustering). Adaptive MCS candidates, such as implemented in [42], are also of interest for future research in *ds*. Adaptive candidate samples may allow to reduce implementation times and improve cluster identification.

Further works that use density scanning in AK should also tackle the definition of the initial sample size. As highlighted previously, adaptive samples are of interest. If the initial sample generated to perform the active learning is too large, in comparison to the one required for an accurate P_f prediction, the effort required for the active learning and density scanning may increase with no relevant benefit to the accuracy of the P_f prediction. Nonetheless, this is a feature common to most procedures that use AK. To conclude, it is also important to emphasize that full exploitation of the benefits of the added complexity of the density framework may depend on the complexity of the performance function. Metamodeling would benefit from a sense of hierarchy in their implementation, which should be also researched in future works.

Table 5
Random variables of the Cantilever tube problem.

Variable	Parameter 1	Parameter 2	Distribution
$x_1[t(mm)]$	5.0(μ)	0.1(σ)	Gaussian
$x_2[D(mm)]$	42.0(μ)	0.5(σ)	Gaussian
$x_3[L_1(mm)]$	119.75(*lower bound*)	120.25(*upper bound*)	Uniform
$x_4[L_2(mm)]$	59.75(*lower bound*)	60.25(*upper bound*)	Uniform
$x_5[F_1(kN)]$	3.0(μ)	0.3(σ)	Gaussian
$x_6[F_2(kN)]$	3.0(μ)	0.3(σ)	Gaussian
$x_7[P(kN)]$	12.0(μ)	1.2(σ)	Gaussian
$x_8[T(Nm)]$	90.0(μ)	9.0(σ)	Gumbel
$x_9[S_y(MPa)]$	220.0(μ)	22.0(σ)	Gaussian
$x_{10}[\theta_1(\circ)]$	5.0(μ)	0.5(σ)	Gaussian
$x_{11}[\theta_2(\circ)]$	10.0(μ)	1.0(σ)	Gaussian

Table 6
Average results for the bivariate non-dimensional performance function, based on values for 20 active learning implementations using the same initial DoE size of 12 LHS points ($d + 1$ points). g_{eval} refers to the number of $g(x)$ evaluations. AKMCS-U and EFF use the stopping criterion of U $>$ 2 and EFF $<$ 0.001.

Algorithm	$\hat{P}_f$ ($\times 10^{-4}$)	CoV $\hat{P}_f$	e_r(%)	n_{iter}	CoV n_{iter}	g_{eval}	CoV g_{eval}
MCS	1.421	–	–	–	–	4×10^6	–
AKMCS-U	1.469	0.04	2.1	56.95	0.04	68.95	0.03
AKMCS-EFF	1.468	0.05	3.3	58.33	0.11	70.33	0.09
dsAK	1.488	0.05	4.6	29.43	0.17	41.43	0.12
dsAKP	1.477	0.04	3.8	25.13	0.18	43.18	0.15
ds^{2}AK ($\zeta = 0.2$)	1.490	0.04	4.9	31.86	0.15	43.86	0.11

5. Conclusion

The present paper researched on the application of an unsupervised technique to partition the space of candidates in Adaptive Kriging procedures. It consists in using density scanning to group dense agglomerations of candidatesin the space. There are multiple advantages resulting from using this type of partition. It can be used in a sequential selection of dense regions to exploit in the active learning with advantage of, in average, accelerating the convergence of the Kriging prediction due to prioritization of dense regions of candidates. Moreover, it allows the parallelization of the evaluations of the performance function in an unsupervised procedure where the active learning is able to select the number of parallel computations function of the formation of dense "pools" of points. The advantage of this approach, apart from its automated character, is that it comes at little or no expense of the number of performance function evaluations, since no more than one point is selected for enrichment in a single dense region. Finally, density scanning is also applied to mitigate the selection of very close points in the design of experiments.

Density scanning of the space has direct synergies with the Adaptive Kriging procedures. A significant percentage of the learning criteria used in these procedures are dependent on the Kriging metamodel, and tend to create dense regions of candidates. A cross-clustering technique is successfully applied in order to mitigate the expense of performing density scanning, which was identified initially to be prohibitive in terms of computational cost. It consists in using subsets of cross-classification in order to classify the set of candidates. The stopping criterion that uses an expectation of failure misclassification is applied to decide on the convergence of the algorithm. This criterion was identified to produce efficient results. Nonetheless, if a more conservative stopping criterion, such as the one that uses the minimum probability of misclassification is applied, density scanning is still highly valuable to reduce the number of iterations without the expense of increasing the performance function evaluations. This unsupervised capability of reducing the number of iterations at no expense of performance function evaluations has high interest for any adaptive Kriging implementation, regardless of the learning function or stopping criterion.

Three examples with different characteristics were researched. Density scanning was identified to be of particular interest for highly non-linear performance functions where dense agglomerations of points may be formed in different regions of the candidate space. Its application to high-dimensional problems (in the present case with 11 variables) is feasible at limited additional cost, in particular when compared with the cost of evaluating a Kriging metamodel with an enriched design of experiments.

Future research on the presented framework should consider the usage of adaptive sample sizes, which are expected to significantly improve the efficiency of applying density scanning.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ress.2020.106908

References

[1] Echard B, Gayton N, Lemaire M. Ak-mcs: an active learning reliability method combining kriging and monte carlo simulation. Struct Safety 2011;33(2):145–54.
[2] Kaymaz I. Application of kriging method to structural reliability problems. Struct Safety 2005;27(2):133–51.
[3] Bichon B, Eldred M, Swiler L, Mahadevan S, McFarland J. Efficient global reliability analysis for nonlinear implicit performance functions. AIAA J 2008;46(10):2459–68.
[4] Jones D, Schonlau M, Welch W. Efficient global optimization of expensive black-box functions. J Global Optim 1998;13(4):455–92.
[5] 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 Safety 2013;111:232–40.
[6] Fauriat W, Gayton N. Ak-sys: an adaptation of the ak-mcs method for system reliability. Reliab Eng Syst Safety 2014;123:137–44.
[7] Huang X, Chen J, Zhu H. Assessing small failure probabilities by ak–ss: an active learning method combining Kriging and subset simulation. Struct Safety 2016;59:86–95.
[8] Tong C, Sun Z, Zhao Q, Wang Q, Wang S. A hybrid algorithm for reliability analysis combining Kriging and subset simulation importance sampling. J Mech Sci Technol 2015;29(8):3183–93.
[9] Lv Z, Lu Z, Wang P. A new learning function for Kriging and its applications to solve reliability problems in engineering. Comput Math Appl 2015;70(5):1182–97.
[10] Sun Z, Wang J, Li R, Tong C. LIF: A new Kriging based learning function and its application to structural reliability analysis. Reliab Eng Syst Safety 2017;157:152–65.
[11] Zhang X, Wang L, Sørensen J. REIF: a novel active-learning function towards adaptive Kriging surrogate models for structural reliability analysis. Reliab Eng Syst Safety 2019.
[12] Gaspar B, Teixeira A, Guedes Soares C. Adaptive surrogate model with active refinement combining Kriging and a trust region method. Reliab Eng Syst Safety 2017;165:277–91.
[13] Wen Z, Pei H, Liu H, Yue Z. A sequential Kriging reliability analysis method with characteristics of adaptive sampling regions and parallelizability. Reliab Eng Syst Safety 2016;153:170–9.
[14] Ginsbourger D, Le Riche R, Carraro L. Kriging is well-suited to parallelize optimization. Computational intelligence in expensive optimization problems. Springer; 2010. p. 131–62.
[15] Lelièvre N, Beaurepaire P, Mattrand C, Gayton N. Ak-mcsi: a Kriging-based method to deal with small failure probabilities and time-consuming models. Struct Safety 2018;73:1–11.
[16] Teixeira R, OConnor A, Nogal M. Adaptive Kriging with biased randomisation for reliability analysis of complex limit state functions. 17th international probabilistic workshop (IPW2019). 2019.
[17] Cui F, Ghosn M. Implementation of machine learning techniques into the subset simulation method. Struct Safety 2019;79:12–25.
[18] Dong Y, Teixeira A, Soares CG. Application of adaptive surrogate models in time-variant fatigue reliability assessment of welded joints with surface cracks. Reliab Eng Syst Safety 2020;195:106730.
[19] Gaspar B, Teixeira A, Guedes Soares C. Assessment of the efficiency of kriging surrogate models for structural reliability analysis. Probab Eng Mech 2014;37:24–34.
[20] Echard B, Gayton N, Bignonnet A. A reliability analysis method for fatigue design. IntJ Fatigue 2014;59:292–300.
[21] Zhang L, Lu Z, Wang P. Efficient structural reliability analysis method based on advanced kriging model. Appl Math Modell 2015;39(2):781–93.
[22] Teixeira R, O'Connor A, Nogal M, Nichols J, Spring M. Structural probabilistic assessment of offshore wind turbine operation fatigue based on kriging interpolation. Proceedings of the 27th European safety and reliability conference (ESREL2017), Portorož, Slovenia, June. 2017. p. 18–22.
[23] Peijuan Z, Ming W, Zhouhong Z, Liqi W. A new active learning method based on the learning function u of the ak-mcs reliability analysis method. Engineering

Structures 2017;148:185–94.
[24] Teixeira R, Nogal M, OConnor A, Nichols J, Dumas A. Stress-cycle fatigue design with kriging applied to offshore wind turbines. Int J Fatigue 2019;125:454–67.
[25] Teixeira R, O'Connor A, Nogal M. Fatigue reliability using a multiple surface approach. 13th international conference on applications of statistics and probability in civil engineering(ICASP13). Seoul, South Korea, May 26-30; 2019.
[26] Schöbi R, Sudret B, Marelli S. Rare event estimation using polynomial-chaos Kriging. ASCE-ASME J Risk UncertainEng Syst, Part A 2016;3(2):D4016002.
[27] Jian W, Zhili S, Qiang Y, Rui L. Two accuracy measures of the Kriging model for structural reliability analysis. Reliab Eng Syst Safety 2017;167:494–505.
[28] Roustant O, Ginsbourger D, Deville Y. Dicekriging, diceoptim: two r packages for the analysis of computer experiments by Kriging-based metamodeling and optimization. J Stat Softw 2012.
[29] Rasmussen C. Gaussian processes in machine learning. Summer school on machine learning. Springer; 2003. p. 63–71.
[30] Marelli S, Sudret B. Uqlab: a framework for uncertainty in matlab. Proceedings 2nd int. conf. on vulnerability, risk analysis and management (ICVRAM). 2014. p. 2554–63.
[31] Picheny V, Wagner T, Ginsbourger D. A benchmark of Kriging-based infill criteria for noisy optimization. Struct Multidiscip Optim 2013;48(3):607–26.
[32] Ranjan P, Bingham D, Michailidis G. Sequential experiment design for contour estimation from complex computer codes. Technometrics 2008;50(4):527–41.
[33] Ester M, Kriegel H, Sander J, Xu X, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd. 96. 1996. p. 226–31.
[34] Fahad A, Alshatri N, Tari Z, Alamri A, Khalil I, Zomaya A, et al. A survey of clustering algorithms for big data: taxonomy and empirical analysis. IEEE TransEmerging TopicsComput 2014;2(3):267–79.
[35] Sheskin D. Handbook of parametric and nonparametric statistical procedures. CRC Press; 2003.
[36] Krejcie R, Morgan D. Determining sample size for research activities. EducPsycholMeasur 1970;30(3):607–10.
[37] Jiang C, Qiu H, Yang Z, Chen L, Gao L, Li P. A general failure-pursuing sampling framework for surrogate-based reliability analysis. Reliab Eng Syst Safety 2019;183:47–59.
[38] Zhang J, Xiao M, Gao L. An active learning reliability method combining Kriging constructed with exploration and exploitation of failure region and subset simulation. Reliab Eng Syst Safety 2019.
[39] Guo J, Du X. Reliability sensitivity analysis with random and interval variables. Int J Numer MethodsEng 2009;78(13):1585–617.
[40] Chen M, Gao X, Li H. Parallel dbscan with priority r-tree. 2010 2nd IEEE international conference on information management and engineering. IEEE; 2010. p. 508–11.
[41] Kumar K, Reddy A. A fast dbscan clustering algorithm by accelerating neighbor searching using groups method. Pattern Recognit 2016;58:39–48.
[42] Wang Z, Shafieezadeh A. REAK: Reliability analysis through error rate-based adaptive Kriging. Reliab Eng Syst Safety 2019;182:33–45.