Rare Event Estimation Using Polynomial-Chaos Kriging

R. Schöbi, B. Sudret, S. Marelli

▶ **To cite this version:**

R. Schöbi, B. Sudret, S. Marelli. Rare Event Estimation Using Polynomial-Chaos Kriging. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 2016, 3 (2), 10.1061/AJRUA6.0000870 . hal-01429018

HAL Id: hal-01429018
https://hal.science/hal-01429018v1

Submitted on 10 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ETH zürich

Rare event estimation using Polynomial-Chaos-Kriging

R. Schöbi, B. Sudret, S. Marelli

Chair of Risk, Safety and Uncertainty Quantification
Stefano-Franscini-Platz 5
CH-8093 Zürich

Risk, Safety & Uncertainty Quantification

Data Sheet

Journal:	ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering
Report Ref.:	RSUQ-2016-001
Arxiv Ref.:	-
DOI:	10.1061/AJRUA6.0000870 (D4016002)
Date submitted:	March 31, 2015
Date accepted:	January 4, 2016

Rare event estimation using Polynomial-Chaos-Kriging

R. Schöbi[1], B. Sudret[1], and S. Marelli[1]

[1] *Chair of Risk, Safety and Uncertainty Quantification*

ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland

March 31st, 2015

Abstract

Structural reliability analysis aims at computing the probability of failure of systems whose performance may be assessed by using complex computational models (*e.g.* expensive-to-run finite element models). A direct use of Monte Carlo simulation is not feasible in practice, unless a surrogate model (such as Kriging, a.k.a Gaussian process modeling) is used. Such meta-models are often used in conjunction with adaptive experimental designs (*i.e.* design enrichment strategies), which allows one to iteratively increase the accuracy of the surrogate for the estimation of the failure probability while keeping low the overall number of runs of the costly original model.

In this paper we develop a new structural reliability method based on the recently developed Polynomial-Chaos Kriging (PC-Kriging) approach coupled with an active learning algorithm known as AK-MCS. We formulate the problem in such a way that the computation of both small probabilities of failure and extreme quantiles is unified. We discuss different convergence criteria for both types of analyses, and show in particular that the original AK-MCS stopping criterion may be over-conservative. We finally elaborate a multi-point enrichment algorithm which allows us to add several points in each iteration, thus fully exploiting high-performance computing architectures.

The proposed method is illustrated on three examples, namely a two-dimensional case which allows us to underline the advantages of our approach compared to standard AK-MCS. Then the quantiles of the 8-dimensional borehole function are estimated. Finally the reliability of a truss structure (10 random variables) is addressed. In all case, accurate results are obtained with about 100 runs of the original model.

KeyWords Adaptive experimental design, design enrichment, Kriging, meta-modeling, PC-Kriging, rare events

1 Introduction

Advanced computer simulations of complex physical systems (*e.g.* finite element models) are common in modern engineering to assess and optimize their performance. At the same time, awareness is growing on the concepts of structural reliability and robust design, hence making the quantification and propagation of uncertainties a key issue (Sudret, 2007; De Rocquigny et al., 2008).

Due to the high cost of repeatedly evaluating complex computational models, analyses with classical sampling techniques such as Monte Carlo (MC) are often intractable. In this context, meta-modeling techniques allow one to develop fast-to-evaluate surrogate models from a limited collection of evaluations of the original computational model, which is called the experimental design. Commonly used techniques include Kriging (Santner et al., 2003; Rasmussen and Williams, 2006), Polynomial Chaos Expansions (Ghanem and Spanos, 2003), support vector machines (Gunn, 1998) and neural networks (Hurtado and Alvarez, 2001).

Two cases are generally distinguished in the context of rare event estimation, namely the estimation of failure probabilities (so-called structural reliability analyses) and that of extreme quantiles. Both quantities describe important statistics of the response of the computational model. In the context of engineering and high reliability targets, failure probabilities are generally low (in the order of 10^{-3} to 10^{-6}) and the quantiles extreme (*e.g.* 99% quantile), therefore structural reliability falls into the broad class of rare events estimation.

Rare events estimation has been in turn often related to the concept of experimental design enrichment (a.k.a. adaptive experimental design). The core idea is to start calibrating a meta-model with a limited set of computational runs and then to add new model evaluations iteratively to increase the prediction accuracy of the rare event statistics of interest. Kriging (a.k.a. Gaussian process modeling) has been introduced by Kaymaz (2005) and Schueremans and Van Gemert (2005b) to the field of structural reliability. The use of adaptive experimental design algorithms has been investigated by Bichon et al. (2008), Echard et al. (2011), and Ran-

jan et al. (2012) who combine an adaptive Kriging meta-model with Monte Carlo simulation, and Echard et al. (2013) and Cadini et al. (2014), who modify it for the case of importance sampling. Under the name "stepwise uncertainty reduction" Arnaud et al. (2010), Bect et al. (2012) and Jala (2013) discuss an adaptive Kriging approach in a Bayesian context. Bourinet et al. (2011) use instead support vector machines in the context of subset simulations. Dani et al. (2008) and Srinivas et al. (2012) use upper confidence bounds. Dai et al. (2014) discuss the use of wavelet-density-based adaptive importance sampling. Finally, the use of adaptive Kriging to build up a quasi-optimal importance sampling density (thus leading to unbiased estimates of the failure probability) has been recently proposed in Dubourg et al. (2013) and Dubourg and Sudret (2014).

In this paper we combine Polynomial-Chaos-Kriging (PC-Kriging) (Schöbi and Sudret, 2014; Schöbi et al., 2015; Kersaudy et al., 2015) with design enrichment to solve reliability analysis and extreme quantile problems. We focus on the best choice of additional samples in the experimental design as well as on the option of adding multiple samples in each iteration of the algorithm. Adding multiple samples in each iteration would allow one to take advantage of distributed computing facilities.

This paper is organized as follows: after the introduction of a proper formalism for structural reliability analysis, the principles of Polynomial-Chaos-Kriging (Schöbi et al., 2015) are given. An existing adaptive experimental design algorithm is then presented and extended to PC-Kriging and the concept of distributed computing. It is shown how the same algorithm can be used to estimate both failure probabilities and extreme quantiles with only slight modification. The application of the resulting framework is finally demonstrated on a low-dimensional analytical function and on two realistic engineering applications.

2 Structural reliability analysis

2.1 Computational model

In this paper, a *computational model* $\mathcal{M}$ is defined as a mapping of the M-dimensional input vector of parameters $\boldsymbol{x}$ to the one-dimensional output scalar y, *i.e.* $\mathcal{M} : \boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} \subset \mathbb{R}^M \to y = \mathcal{M}(\boldsymbol{x}) \in \mathbb{R}$. The uncertainties in the input parameters are represented by a random vector $\boldsymbol{X}$

with joint probability density function (PDF) $f_{\boldsymbol{X}}$. The components of $\boldsymbol{X} = \{X_1, \ldots, X_M\}$ are assumed independent throughout this paper for the sake of simplicity. Moreover, when the components are dependent, it is always possible to recast the problem in terms of independent variables using the Nataf or Rosenblatt transform (Lemaire, 2009; Lebrun and Dutfoy, 2009a,b). The model response then becomes a random variable Y obtained by propagating the input uncertainty in $\boldsymbol{X}$ through the computational model:

$$Y = \mathcal{M}(\boldsymbol{X}). \tag{1}$$

It is further assumed that the computational model $\mathcal{M}$ is a *black-box*, *i.e.* for each realization $\boldsymbol{x}$ of the input random vector, only the corresponding response $y = \mathcal{M}(\boldsymbol{x})$ is accessible. In other words we assume that the inner structure and mechanisms of the computational model are not observable. This is typically the case for *e.g.* a finite element model in which, generally speaking, the governing equations cannot be solved analytically.

2.2 Estimation of a failure probability

In structural reliability analysis, the failure probability P_f is defined as the probability that a model response Y is smaller than a given threshold value y_0:

$$P_f = \mathbb{P}\left(Y \leq y_0\right) = \mathbb{P}\left(\mathcal{M}(\boldsymbol{X}) \leq y_0\right). \tag{2}$$

The failure probability may be recast as the following integral:

$$P_f = \int_{\mathcal{D}_f} f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \quad \text{s.t.} \quad \mathcal{D}_f = \{\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} : \ \mathcal{M}(\boldsymbol{x}) \leq y_0\}, \tag{3}$$

where $f_{\boldsymbol{X}}$ is the joint probability density function of the input vector $\boldsymbol{X}$ and $\mathcal{D}_f$ is the so-called *failure domain*. Note that failure is defined as in Eq. (2) and (3) to ensure consistency with the following sections. In other literature, failure is often defined as a probability of exceedence, *e.g.* $\mathbb{P}\left(\mathcal{M}(\boldsymbol{X}) \geq y_{\text{adm}}\right)$ which is also cast as $P_f = \mathbb{P}\left(g(\boldsymbol{X}) \leq 0\right)$ where $g(\boldsymbol{X}) = y_{\text{adm}} - \mathcal{M}(\boldsymbol{X})$. The integral in Eq. (3) has an implicit integration domain $\mathcal{D}_f$ and cannot be solved analytically in the general case. A Monte Carlo estimation of the failure probability can be formulated to circumvent this limitation. Assume a reasonably large sample of $\boldsymbol{X}$ denoted by

$\mathcal{S} = \{\boldsymbol{x}_i, ..., \boldsymbol{x}_n\}$. The failure probability can be estimated by:

$$\widehat{P}_f = \frac{n_f}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\mathcal{M}(\boldsymbol{x}_i)\leq y_0}, \tag{4}$$

where n_f is the number of failure samples $\boldsymbol{x}_i \in \mathcal{D}_f$, $n = |\mathcal{S}|$ is the total number of samples and $\mathbb{I}_{\mathcal{M}(\boldsymbol{x})\leq y_0}$ is an indicator function for failure such that $\mathbb{I} = 1$ for $\mathcal{M}(\boldsymbol{x}) \leq y_0$ and $\mathbb{I} = 0$ otherwise.

The accuracy of this estimate can be measured directly with the theoretical coefficient of variation for Monte Carlo samplings:

$$\mathrm{CoV}\left[\widehat{P}_f\right] = \sqrt{\frac{1-\widehat{P}_f}{n\cdot\widehat{P}_f}}. \tag{5}$$

The main limitation of pure Monte Carlo approaches, namely the loss of efficiency for small failure probabilities, is apparent in Eq. (5). The required number of samples to obtain a target coefficient of variation in Eq. (5) rapidly increases with decreasing failure probability. For instance, to achieve a target coefficient of variation $\mathrm{CoV}\left[\widehat{P}_f\right] = 0.1$ given a failure probability $\widehat{P}_f = 10^{-k}$, $n \approx \left(\widehat{P}_f \cdot \mathrm{CoV}\left[\widehat{P}_f\right]^2\right)^{-1} = 10^{k+2}$ samples are needed.

Note that the indicator function $\mathbb{I}$ in Eq. (4) transforms the integration problem in Eq. (3) into a classification problem: identifying the failure ($\mathcal{M}(\boldsymbol{x}) \leq y_0$) and safe ($\mathcal{M}(\boldsymbol{x}) > y_0$) domains. Generally, it is sufficient to approximate the boundary between those domains, the *limit state surface* ($\mathcal{M}(\boldsymbol{x}) = y_0$), to accurately estimate the associated failure probability.

2.3 Estimation of a quantile

Quantile estimation consists in determining a quantile q_α so that the probability of $\mathcal{M}(\boldsymbol{X})$ being smaller than q_α equals a preset value α, *i.e.*:

$$\mathbb{P}\left(\mathcal{M}(\boldsymbol{X}) \leq q_\alpha\right) = \alpha, \tag{6}$$

where $\alpha = (0, 1)$. This constitutes an inverse problem compared to classical failure probability estimation shown in Eq. (2).

The quantile estimation problem can also be solved by Monte Carlo simulation. Consider again a reasonably large sample set $\mathcal{S} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ and the corresponding response values

$y = \{y_1, ..., y_n\} = \{\mathcal{M}(\boldsymbol{x}_1), \ldots, \mathcal{M}(\boldsymbol{x}_n)\}$. Assume the response values are ranked in ascending order and denote them by $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$. Then the estimator of the quantile q_α reads:

$$\widehat{q}_\alpha = y_{\lfloor n\alpha \rfloor}, \tag{7}$$

where $\lfloor n\alpha \rfloor$ is the largest integer smaller than $n\alpha$.

Analogously to failure probability estimation, the quantile estimation problem can also be cast as a classification problem:

$$\text{find} \quad \widehat{q}_\alpha \quad \text{s.t.} \quad \alpha \approx \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\mathcal{M}(\boldsymbol{x}_i) \leq \widehat{q}_\alpha}. \tag{8}$$

2.4 Limit state surface

Both the estimation of failure probabilities (Eq. (4)) and of quantiles (Eq. (8)) may be treated as classification problems dividing the input domain into failure and safe regions. We therefore introduce a *limit state parameter* a which determines whether a response y belongs to the failure ($y \leq a$) or to the safe domain ($y > a$). For the estimation of failure probabilities $a = y_0$, whereas for the estimation of quantiles $a = \widehat{q}_\alpha$. Due to those similarities we will denote failure probability and quantiles generically as statistics of interest in the remaining of this paper.

3 Meta-modeling

3.1 Kriging

When the computational model $\mathcal{M}$ is an expensive-to-evaluate function, Eq. (4) and Eq. (8) may become intractable. The behavior of the computational model $\mathcal{M}$ can then be approximated by a meta-model, such as Kriging (a.k.a. Gaussian process modeling). Kriging interprets the computational model as a realization of an underlying Gaussian process (Santner et al., 2003):

$$\mathcal{M}(\boldsymbol{x}) \approx \mathcal{M}^{(\mathrm{K})}(\boldsymbol{x}) = \boldsymbol{\beta}^{\mathsf{T}} \cdot \boldsymbol{f}(\boldsymbol{x}) + \sigma^2 Z(\boldsymbol{x}, \omega), \tag{9}$$

where $\boldsymbol{\beta}^{\mathsf{T}} \cdot \boldsymbol{f}(\boldsymbol{x})$ is the mean value of the Gaussian process (a.k.a. *trend*), σ^2 is the process variance, $Z(\boldsymbol{x}, \omega)$ is a zero-mean, unit-variance stationary Gaussian process and $\omega \in \Omega$ denotes an elementary event in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The Gaussian process is characterized by an autocorrelation function $R = R(|\boldsymbol{x} - \boldsymbol{x}'|; \boldsymbol{\theta})$ and its hyper-parameters $\boldsymbol{\theta}$. In this paper, we consider two types of Kriging models: ordinary Kriging and universal Kriging. Ordinary Kriging assumes a constant trend ($\boldsymbol{f}(\boldsymbol{x}) = 1$) whereas universal Kriging assumes a trend dependent on the input variables, *e.g.* $\boldsymbol{f}(\boldsymbol{x}) = \{1, x_1, \ldots, x_M\}$ for a linear trend.

Consider a set of N samples of the input vector $\boldsymbol{\mathcal{X}} = \{\boldsymbol{\chi}^{(1)}, \ldots, \boldsymbol{\chi}^{(N)}\}$, called *experimental design*, and the corresponding response of the exact computational model $\mathcal{Y} = \left\{\mathcal{Y}^{(1)} = \mathcal{M}(\boldsymbol{\chi}^{(1)}) \right.$ $\left. \ldots, \mathcal{Y}^{(N)} = \mathcal{M}(\boldsymbol{\chi}^{(N)})\right\}$. The estimation of the Kriging parameters $\{\boldsymbol{\beta}, \sigma^2\}$ is computed by generalized least-squares solution (Santner et al., 2003):

$$\boldsymbol{\beta}(\boldsymbol{\theta}) = \left(\mathbf{F}^{\mathsf{T}} \mathbf{R}^{-1} \mathbf{F}\right)^{-1} \mathbf{F}\, \mathbf{R}^{-1} \mathcal{Y}, \tag{10}$$

$$\sigma^2(\boldsymbol{\theta}) = \frac{1}{N} \left(\mathcal{Y} - \mathbf{F}\, \boldsymbol{\beta}\right)^{\mathsf{T}} \mathbf{R}^{-1} \left(\mathcal{Y} - \mathbf{F}\, \boldsymbol{\beta}\right), \tag{11}$$

where $\mathbf{R}_{ij} = R(|\boldsymbol{\chi}^{(i)} - \boldsymbol{\chi}^{(j)}|; \boldsymbol{\theta})$ is the correlation matrix of the experimental design points and $\mathbf{F}_{ij} = f_j(\boldsymbol{\chi}^{(i)})$ is the information matrix. In case the hyper-parameters $\boldsymbol{\theta}$ are unknown, their optimal values can be estimated through maximum likelihood estimation.

The prediction of the response value y of an arbitrary input sample point $\boldsymbol{x}$ is a Gaussian random variable characterized by (Santner et al., 2003):

$$\mu_{\widehat{Y}}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{\beta} + \boldsymbol{r}(\boldsymbol{x})^{\mathsf{T}} \mathbf{R}^{-1} \left(\mathcal{Y} - \mathbf{F}\boldsymbol{\beta}\right), \tag{12}$$

$$\sigma^2_{\widehat{Y}}(\boldsymbol{x}) = \sigma^2_y \left(1 - \langle \boldsymbol{f}(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{r}(\boldsymbol{x})^{\mathsf{T}} \rangle \begin{bmatrix} \mathbf{0} & \mathbf{F}^{\mathsf{T}} \\ \mathbf{F} & \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{f}(\boldsymbol{x}) \\ \boldsymbol{r}(\boldsymbol{x}) \end{bmatrix} \right), \tag{13}$$

where $\mu_{\widehat{Y}}(\boldsymbol{x})$ and $\sigma^2_{\widehat{Y}}(\boldsymbol{x})$ are the prediction mean value and variance, and $r_i(\boldsymbol{x}) = R(|\boldsymbol{x} - \boldsymbol{\chi}^{(i)}|; \boldsymbol{\theta})$ is the correlation between the new sample $\boldsymbol{x}$ and the experimental design point $\boldsymbol{\chi}^{(i)} \in \boldsymbol{\mathcal{X}}$.

3.2 PC-Kriging

3.2.1 Framework

Polynomial-Chaos-Kriging (PC-Kriging) consists of a universal Kriging model whose trend is modeled by a set of orthogonal polynomials with respect to the input variables (Schöbi et al., 2015):

$$\mathcal{M}(\boldsymbol{x}) \approx \mathcal{M}^{(\mathsf{PCK})}(\boldsymbol{x}) = \sum_{\boldsymbol{\tau} \in \mathcal{T}} t_{\boldsymbol{\tau}}\, \psi_{\boldsymbol{\tau}}(\boldsymbol{x}) + \sigma^2\, Z(\boldsymbol{x}, \omega), \tag{14}$$

where $\sum_{\boldsymbol{\tau} \in \mathcal{T}} t_{\boldsymbol{\tau}} \psi_{\boldsymbol{\tau}}(\boldsymbol{x})$ is a trend defined by a set of $|\mathcal{T}|$ multivariate orthonormal polynomials $\psi_{\boldsymbol{\tau}}(\boldsymbol{x})$ indexed by the multi-index $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_M\}$ and $t_{\boldsymbol{\tau}}$ are the corresponding coefficients. Due to the assumed independence of the input variables, the multivariate polynomials can be composed by tensor product of univariate orthogonal polynomials:

$$\psi_{\boldsymbol{\tau}}(\boldsymbol{x}) = \prod_{i=1}^{M} \psi_{\tau_i}^{(i)}(x_i), \tag{15}$$

where $\psi_{\tau_i}^{(i)}(x_i)$ is a univariate polynomial of degree τ_i in the i-th variable. For each X_i the orthonormality in coherency with the input distributions is given by:

$$\langle \psi_j^{(i)}, \psi_k^{(i)} \rangle = \int_{\mathcal{D}_i} \psi_j^{(i)}(x)\, \psi_k^{(i)}(x)\, f_{X_i}(x) dx = \delta_{jk}, \tag{16}$$

where $\psi_j^{(i)}$, $\psi_k^{(i)}$ are two univariate polynomials in the i-th variable, $\mathcal{D}_i$ is the support of the distribution of X_i, f_{X_i} is the marginal probability density function, and $\delta_{jk} = 1$ for $j = k$ and $\delta_{jk} = 0$ otherwise. A summary of common univariate orthonormal polynomials can be found in Xiu and Karniadakis (2002).

3.2.2 Optimal PC-Kriging

There are several ways to determine the most efficient set of polynomials $\mathcal{T}$ and combine it with universal Kriging. Amongst them is the so-called *Optimal-PC-Kriging* (OPC-Kriging) approach (Schöbi and Sudret, 2014; Schöbi et al., 2015), which is presented briefly in this paper.

The information required for the algorithm comprises the input marginals, the autocorrelation function R, an experimental design $\boldsymbol{\mathcal{X}}$ and the corresponding response values $\mathcal{Y}$. The best

sparse set of multivariate polynomials (summarized in the index set $\mathcal{T}$) is obtained by least-angle regression (LARS) (Efron et al., 2004; Blatman and Sudret, 2011). LARS results in a ranked set of P polynomials. From this ordered set, P Kriging models are calibrated with an increasing number of polynomials each. The first trend is the polynomial chosen first in the LARS algorithm, while the trend of the P-th meta-model includes all P polynomials. The various meta-models obtained (with a polynomial trend having $1, 2, \ldots, P$ terms respectively) are ranked according to their respective *leave-one-out* (LOO) cross-validation error and the meta-model with the minimal LOO error is chosen as the OPC-Kriging meta-model.

3.2.3 Characteristics

PC-Kriging combines a global with a local approximation of the behavior of the computational model. The global behavior is modeled by the PCE trend whereas the local behavior is modeled by the Gaussian process. This allows one to meta-model a computational model more accurately than PCE or other Kriging meta-models taken separately, see details in Schöbi and Sudret (2014) and Schöbi et al. (2015).

3.3 Model accuracy

Universal Kriging-based meta-models generally approximate the behavior of the original model $\mathcal{M}$ most accurately close to the points in the experimental design. These points, however, are not always optimal for the estimation of the statistics of interest of Y. As an example, in order to compute a small failure probability $P_f = \mathbb{P}(Y \leq a)$ w.r.t. a given threshold a, it is of interest to concentrate the experimental design in the region close to the limit state surface ($y \approx a$), which is usually unknown a-priori.

Consider the case of a given initial experimental design consisting of N_0 samples. By adding samples in a *guided way*, the accuracy of the estimation of the statistics of interest can be enhanced more efficiently than adding samples randomly. This is the main idea behind adaptive experimental design algorithms that are now presented.

4 Adaptive experimental design algorithm

4.1 Main algorithm

We adopt a slightly modified version of the *Adaptive-Kriging-Monte-Carlo-Simulation* (AK-MCS) algorithm introduced in Echard et al. (2011) for enriching the experimental design. An overview of the algorithm is shown in the flowchart in Figure 1. The white boxes mark the information required to start the algorithm, whereas the blue boxes represent the computational tasks. The main steps can be summarized as follows:

1. An initial experimental design $\boldsymbol{\mathcal{X}}$ is generated by Latin-hypercube sampling and the corresponding exact response $\mathcal{Y}$ is computed.

2. A meta-model $\widehat{\mathcal{M}}$ is calibrated based on $\{\boldsymbol{\mathcal{X}}, \mathcal{Y}\}$. In this paper, PC-Kriging meta-models are used, whereas ordinary Kriging meta-models are used in Echard et al. (2011).

3. A large set of candidate samples $\mathcal{S} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ is generated from $\boldsymbol{X}$ and the response values are predicted by the meta-model $\widehat{\mathcal{M}}$.

4. The limit state surface is estimated based on the current meta-model.

5. All candidate samples $\boldsymbol{x} \in \mathcal{S}$ are ranked according to an enrichment criterion.

6. Then the *selection step* determines the sample(s) $\boldsymbol{\chi}^*$ to be added to the experimental design of the meta-model, so that $\boldsymbol{\mathcal{X}} \leftarrow \{\boldsymbol{\mathcal{X}}, \boldsymbol{\chi}^*\}$.

7. The corresponding response of the exact computational model is evaluated and added to $\mathcal{Y} \leftarrow \{\mathcal{Y}, \mathcal{Y}^*\}$. The iterative algorithm then goes back to step 2 where the meta-model is now calibrated with the enriched experimental design.

8. The iterations are terminated through a convergence measure (stopping criterion) with respect to the statistics of interest in order to stop the design enrichment algorithm reliably.

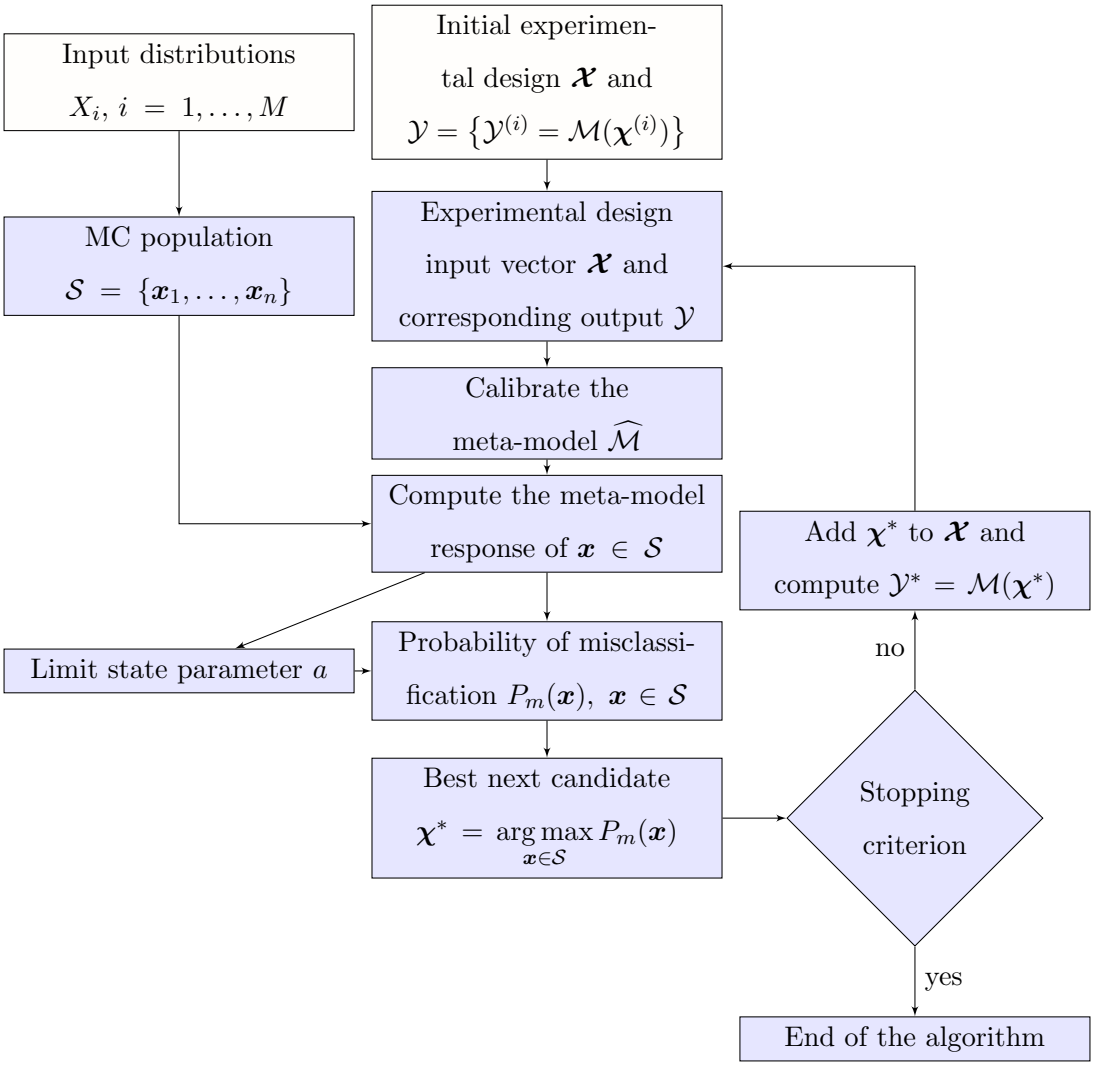


Figure 1: Flowchart of the adaptive algorithm for rare event estimation

4.2 Initial experimental design

In order to have a well-behaved initial experimental design, a space-filling sampling method is required. In this paper the Latin-hypercube sampling algorithm is used. In particular, the unit hyper-cube $[0,1]^M$ is sampled and the samples are mapped to the variable space by the inverse cumulative distribution function of the marginals. Other sampling strategies include quasi-random sequences, such as Sobol' sequences.

4.3 Candidate selection

4.3.1 Single sample selection

A *learning function* (LF) estimates the expected value of information gained with respect to the statistics of interest when point $\boldsymbol{x}$ is added to the experimental design $\boldsymbol{\mathcal{X}}$ of the meta-model. A variety of learning functions has been proposed in the literature including the *Expected Feasibility Function* (EFF) (Bichon et al., 2008), the *Expected Improvement* (EI) functions (Jones et al., 1998; Ginsbourger et al., 2013), the *Gaussian process upper confidence bounds* (GP-UCB) (Srinivas et al., 2012) and the *U-function* (Echard et al., 2011). In the following, we discuss the U-function more in depth.

The U-function is based on the concept of misclassification and the very nature of the Gaussian process meta-model. To each sample point $\boldsymbol{x} \in \mathcal{S}$ corresponds a non-zero probability that the prediction mean $\mu_{\widehat{Y}}(\boldsymbol{x}) > a$ (safe domain) whereas the true value satisfies $\mathcal{M}(\boldsymbol{x}) \leq a$ (failure domain) or vice versa. Due to the fact that the prediction in PC-Kriging is a Gaussian random variable described by the first two moments $\{\mu_{\widehat{Y}}(\boldsymbol{x}),\ \sigma_{\widehat{Y}}(\boldsymbol{x})\}$, the probability of misclassification P_m can be written as (Bect et al., 2012):

$$P_m(\boldsymbol{x}) = \min\left[\Phi\left(\frac{\mu_{\widehat{Y}}(\boldsymbol{x}) - a}{\sigma_{\widehat{Y}}(\boldsymbol{x})}\right), \Phi\left(\frac{a - \mu_{\widehat{Y}}(\boldsymbol{x})}{\sigma_{\widehat{Y}}(\boldsymbol{x})}\right)\right] \equiv \Phi\left(-\frac{|\mu_{\widehat{Y}}(\boldsymbol{x}) - a|}{\sigma_{\widehat{Y}}(\boldsymbol{x})}\right). \tag{17}$$

The probability of misclassification is maximized when the fraction tends to zero, *i.e.* $-|\mu_{\widehat{Y}}(\boldsymbol{x}) - a|/\sigma_{\widehat{Y}}(\boldsymbol{x}) \sim 0 \rightarrow P_m(\boldsymbol{x}) = 0.5$, and is small when either the prediction mean is far away from the limit state parameter a and/or the prediction variance is small. Note that the probability measure used to define misclassification is the one associated to the Gaussian nature of the predictor $\mathcal{N}(\mu_{\widehat{Y}}(\boldsymbol{x}), \sigma_{\widehat{Y}}(\boldsymbol{x}))$ in each point $\boldsymbol{x}$. It shall not be confused with the probability $\mathbb{P}(\cdot)$ in Eq. (2) which corresponds to the input random vector $\boldsymbol{X}$.

In this formalism, the U-function is defined as the "reliability index" attached to the probability of misclassification (Echard et al., 2011):

$$U(\boldsymbol{x}) = \frac{|\mu_{\widehat{Y}}(\boldsymbol{x}) - a|}{\sigma_{\widehat{Y}}(\boldsymbol{x})}. \tag{18}$$

The optimal sample to enrich the experimental design is the one minimizing the U-function among $\boldsymbol{x} \in \mathcal{S}$, thus maximizing the probability of misclassification:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}\in\mathcal{S}}{\arg\min}\, U(\boldsymbol{x}) \equiv \underset{\boldsymbol{x}\in\mathcal{S}}{\arg\max}\, P_m(\boldsymbol{x}). \tag{19}$$

Due to this similarity, both the probability of misclassification and the U-function can be used as learning functions.

4.3.2 Multiple sample selection

In cases where parallel computing is available, it can be beneficial to add *multiple points* at the same time even to speed up the overall computation even if the procedure may be slightly suboptimal. Assuming that K samples can be determined at the beginning of each iteration, the corresponding model responses could be computed simultaneously on K independent CPUs.

There are several ways to sample multiple points from the set of candidates $\mathcal{S}$. Generally speaking though, attractive candidates are found close to the limit state surface. Thus we define a lower and upper boundary of the limit state surface defined by $\mu_{\widehat{Y}}(\boldsymbol{x}) = a$ which takes into account the prediction uncertainty in the Kriging model. The lower boundary reads:

$$\mu_{\widehat{Y}}(\boldsymbol{x}) - k \cdot \sigma_{\widehat{Y}}(\boldsymbol{x}) = a, \tag{20}$$

and the upper boundary:

$$\mu_{\widehat{Y}}(\boldsymbol{x}) + k \cdot \sigma_{\widehat{Y}}(\boldsymbol{x}) = a, \tag{21}$$

where k sets the confidence level typically equal to $1.96 = \Phi^{-1}(97.5\%)$. In other words, *e.g.*, the lower boundary estimates the limit state surface assuming that the real value of every sample $\boldsymbol{x} \in \boldsymbol{X}$ is $\mu_{\widehat{Y}}(\boldsymbol{x}) - k \cdot \sigma_{\widehat{Y}}(\boldsymbol{x})$ instead of the mean value $\mu_{\widehat{Y}}(\boldsymbol{x})$.

Analogously, we define the "mean" failure domain in terms of the prediction mean value:

$$\mathcal{D}_f^0 \stackrel{\text{def}}{=} \left\{ \boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} : \ \mu_{\widehat{Y}}(\boldsymbol{x}) \leq a \right\}, \tag{22}$$

and the corresponding lower and upper bounds to the failure domain:

$$\mathcal{D}_f^- \stackrel{\text{def}}{=} \left\{ \boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} : \ \mu_{\widehat{Y}}(\boldsymbol{x}) + k \cdot \sigma_{\widehat{Y}}(\boldsymbol{x}) \leq a \right\}, \tag{23}$$

$$\mathcal{D}_f^+ \stackrel{\text{def}}{=} \left\{ \boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} : \ \mu_{\widehat{Y}}(\boldsymbol{x}) - k \cdot \sigma_{\widehat{Y}}(\boldsymbol{x}) \leq a \right\}, \tag{24}$$

so that $\mathcal{D}_f^- \subset \mathcal{D}_f^0 \subset \mathcal{D}_f^+$. The lower and upper bounds can be interpreted as the least and most conservative estimate of the failure domain, respectively.

The *limit state margin* $\mathbb{M}_f$ is defined as the intersection between the lower and the upper boundaries of the failure domain (Dubourg, 2011):

$$\mathbb{M}_f \stackrel{\text{def}}{=} \mathcal{D}_f^+ \cap \mathcal{D}_f^- . \tag{25}$$

The limit state margin is a natural region where to focus for candidate samples for the design enrichment. Considering a large set of samples of the input vector $\mathcal{S} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, we define the following sets corresponding to the domains in Eq. (22)-(25):

$$\mathcal{S}_f^0 \stackrel{\text{def}}{=} \left\{\boldsymbol{x} \in \mathcal{S} : \ \mu_{\widehat{Y}}(\boldsymbol{x}) \leq a\right\}, \tag{26}$$

$$\mathcal{S}_f^+ \stackrel{\text{def}}{=} \left\{\boldsymbol{x} \in \mathcal{S} : \ \mu_{\widehat{Y}}(\boldsymbol{x}) - k \cdot \sigma_{\widehat{Y}}(\boldsymbol{x}) \leq a\right\}, \tag{27}$$

$$\mathcal{S}_f^- \stackrel{\text{def}}{=} \left\{\boldsymbol{x} \in \mathcal{S} : \ \mu_{\widehat{Y}}(\boldsymbol{x}) + k \cdot \sigma_{\widehat{Y}}(\boldsymbol{x}) \leq a\right\}, \tag{28}$$

$$\mathcal{S}_{\mathbb{M}} \stackrel{\text{def}}{=} \mathcal{S}_f^+ \cap \mathcal{S}_f^- , \tag{29}$$

where $\mathcal{S}_f^- \subset \mathcal{S}_f^0 \subset \mathcal{S}_f^+$. These sets are the discrete corresponding counterparts to the domains defined in Eq. (22)-(25). The margin set $\mathcal{S}_{\mathbb{M}}$ contains points of interest for enriching the experimental design since they lie close to the true limit state surface. Virtually any point in $\mathcal{S}_{\mathbb{M}}$ could be added to $\boldsymbol{\mathcal{X}}$ and a simple approach would be to sample the K different points randomly. However a better coverage of $\mathbb{M}_f$ can be obtained using clustering techniques.

To account for the relative importance of the samples in $\mathcal{S}_{\mathbb{M}}$, a weighted K-means clustering algorithm is used (Zaki and Meira, 2014). In this algorithm samples with high information value according to the learning function have larger weights, whereas samples with low information value have low weights. The weights are set equal to the probability of misclassification (see Eq. (17)) of each sample $\boldsymbol{x} \in \mathcal{S}_{\mathbb{M}}$. By definition of the probability of misclassification, the weights are bounded on $[0, 0.5]$. The additional K samples $\boldsymbol{\chi}^{(i)}$ are then determined as the samples $\boldsymbol{x} \in \mathcal{S}_{\mathbb{M}}$ closest to the nuclei of each ith cluster.

To showcase the process of selecting additional samples, we consider the following function with two-dimensional input vector:

$$y = \mathcal{M}(\boldsymbol{x}) = 20 - (x_1 - x_2)^2 - 8 \cdot (x_1 + x_2 - 4)^3, \tag{30}$$

where the input variables have uniform distributions $X_i \sim \mathcal{U}(-5, 5),\ i = 1, 2$. As an example we compute the failure probability of the function associated with the criterion $\{\mathcal{M}(\boldsymbol{x}) \leq 0\}$, *i.e.* $a = 0$.

The function is meta-modeled with an ordinary Kriging model. Figure 2(a) displays the initial experimental design ("$\circ$" for failure samples and "+" for safe samples), the exact failure domain (thin line) and the failure region estimated from the prediction mean values $\mu_{\widehat{Y}}(\boldsymbol{x})$ (thick black line). Figures 2(b)-2(d) show the lower and upper bounds of the failure domain and the domain of the limit state margin, respectively. Finally, Figure 2(e) shows the probability of misclassification in the limit state margin and Figure 2(f) shows the selected candidates. The gray area in Figure 2(f) represents the samples in the limit state margin, the large black diamond represents the optimal sample (with highest probability of misclassification) and the hollow black diamonds represent the $K = 5$ samples obtained by weighted K-means clustering.

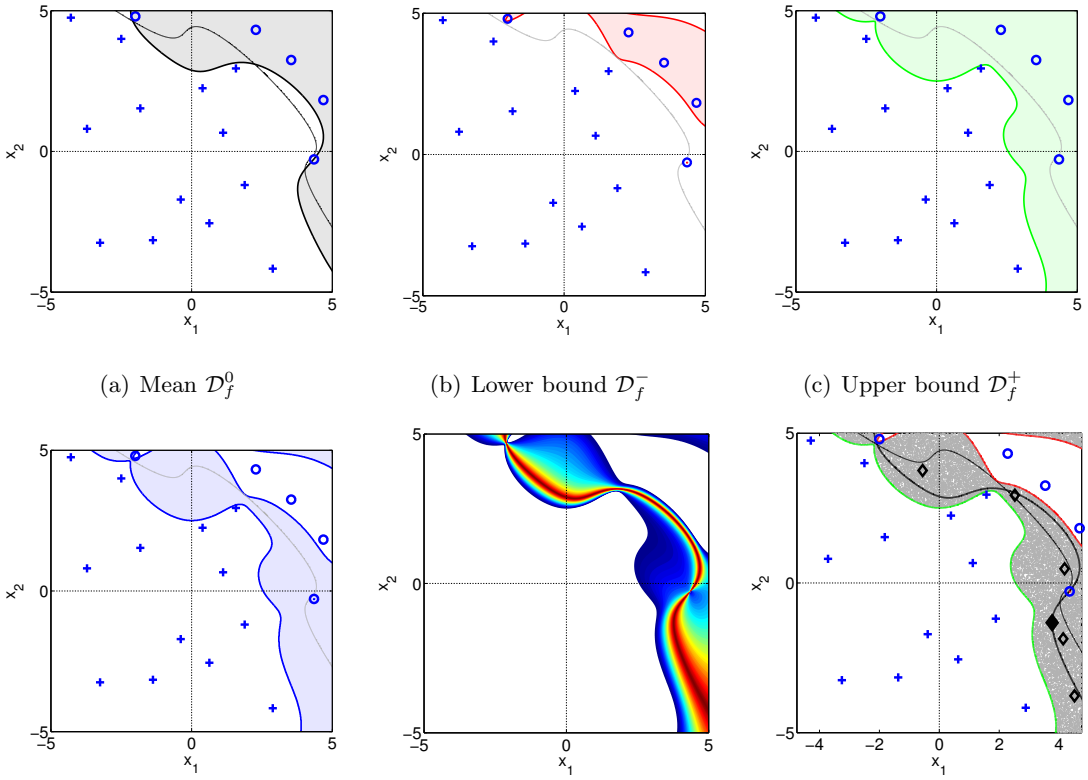
It can be observed that the $K = 5$ samples in Figure 2(f) cover the domain of the limit state margin well and are close to the limit state surface determined by the prediction mean value. This leads to an efficient selection of additional samples for enriching the experimental design. Interestingly, however, the optimal sample in terms of the probability of misclassification is not part of the $K = 5$ sample set. This implies that each of the $K = 5$ samples is suboptimal in terms of the probability of misclassification.

The optimal sample (filled diamond marker in Figure 2(f)) depends highly on the sampling of the limit state margin. In principle any sample with $\mu_{\widehat{Y}}(\boldsymbol{x}) = 0$ could be optimal in terms of the probability of misclassification. This implies that the choice of a single sample with respect to the maximal probability of misclassification is non-unique in theory. However it is unique and well-defined in the discrete optimization problem Eq. (19).

4.4 Stopping criterion

In Echard et al. (2011), the convergence measure (a.k.a. stopping criterion) is based on the accuracy of the meta-model around the limit state surface rather than on the estimation of the statistics of interest. Thus, we propose two stopping criteria for the estimation of failure probabilities and quantiles designed to maximize the accuracy of the statistics of interest while minimizing computational costs.

The stability of the estimate of the statistics of interest can be measured by the size of the limit state margin $\mathbb{M}_f$ and consequently by the associated values of the upper and lower



(a) Mean $\mathcal{D}_f^0$ (b) Lower bound $\mathcal{D}_f^-$ (c) Upper bound $\mathcal{D}_f^+$

(d) Limit state margin $\mathbb{M}_f$ (e) Probabiliy of misclassification P_m (f) Additional samples

Figure 2: Design enrichment when adding $K = 5$ samples – definition of failure domains and illustration of the candidate selection algorithm.

boundaries of the limit state surface. When the boundaries are close to each other (*i.e.* a small limit state margin), then the estimate of the statistics of interest is accurate. Therefore we define the stopping criterion for estimating failure probabilities as:

$$\frac{\widehat{P}_f^+ - \widehat{P}_f^-}{\widehat{P}_f^0} \leq \epsilon_{\widehat{P}_f}, \tag{31}$$

for two consecutive iteration steps where $\epsilon_{\widehat{P}_f} = 5\%$ is used in applications. The upper and lower bound failure probabilities are defined as:

$$\widehat{P}_f^{\pm} \stackrel{\text{def}}{=} \mathbb{P}\left(\mu_{\widehat{Y}}(\boldsymbol{X}) \mp k\, \sigma_{\widehat{Y}}(\boldsymbol{X}) \leq a\right) \tag{32}$$

and:

$$\widehat{P}_f^0 \stackrel{\text{def}}{=} \mathbb{P}\left(\mu_{\widehat{Y}}(\boldsymbol{X}) \leq a\right). \tag{33}$$

The stopping criterion defined in Eq. (31) can be adapted to the context of quantile estimation as follows:

$$\frac{\widehat{q_\alpha^+} - \widehat{q_\alpha^-}}{\widehat{\Sigma}_Y} \leq \epsilon_{\widehat{q}_\alpha}, \tag{34}$$

for two consecutive iteration steps where $\epsilon_{\widehat{q}_\alpha} = 5\%$ practical applications. The quantiles $\widehat{q}_\alpha^{\pm}$ are computed from a large Monte Carlo sampling of the bounds $\mu_{\widehat{Y}}(\boldsymbol{x}) \pm k\,\sigma_{\widehat{Y}}(\boldsymbol{x})$, where $\widehat{\Sigma}_Y$ is the empirical standard deviation of the sample $\mathcal{Y}_{\text{val}} = \left\{\mu_{\widehat{Y}}(\boldsymbol{x}_i),\, i = 1, \ldots, n\right\}$. The normalization in Eq. (34) makes use of the standard deviation $\widehat{\Sigma}_Y$ rather than by $\widehat{q}_\alpha^0$ in analogy with Eq. (31). Indeed, unlike $\widehat{P}_f^0$ in Eq. (31), $\widehat{q}_\alpha^0$ can be any real number in $\mathbb{R}$, which makes it inappropriate for normalization.

5 Applications

5.1 Four-branch function

5.1.1 Problem statement

The four-branch function is a common benchmark in structural reliability analysis that describes the failure of a series system with four distinct component limit states. Its mathematical formulation reads (Waarts, 2000; Schueremans and Van Gemert, 2005a,b):

$$f_1(\boldsymbol{x}) = \min \left\{ \begin{array}{c} 3 + 0.1\,(x_1 - x_2)^2 - \frac{x_1 + x_2}{\sqrt{2}} \\ 3 + 0.1\,(x_1 - x_2)^2 + \frac{x_1 + x_2}{\sqrt{2}} \\ (x_1 - x_2) + \frac{6}{\sqrt{2}} \\ (x_2 - x_1) + \frac{6}{\sqrt{2}} \end{array} \right\} \tag{35}$$

where the input variables are modeled by two independent Gaussian random variables $X_i = \mathcal{N}(0, 1)$. The failure event is defined as $f_1(\boldsymbol{x}) \leq 0$, *i.e.* the failure probability is $P_f = \mathbb{P}\left(f_1(\boldsymbol{X}) \leq 0\right)$.

5.1.2 Failure probability estimation

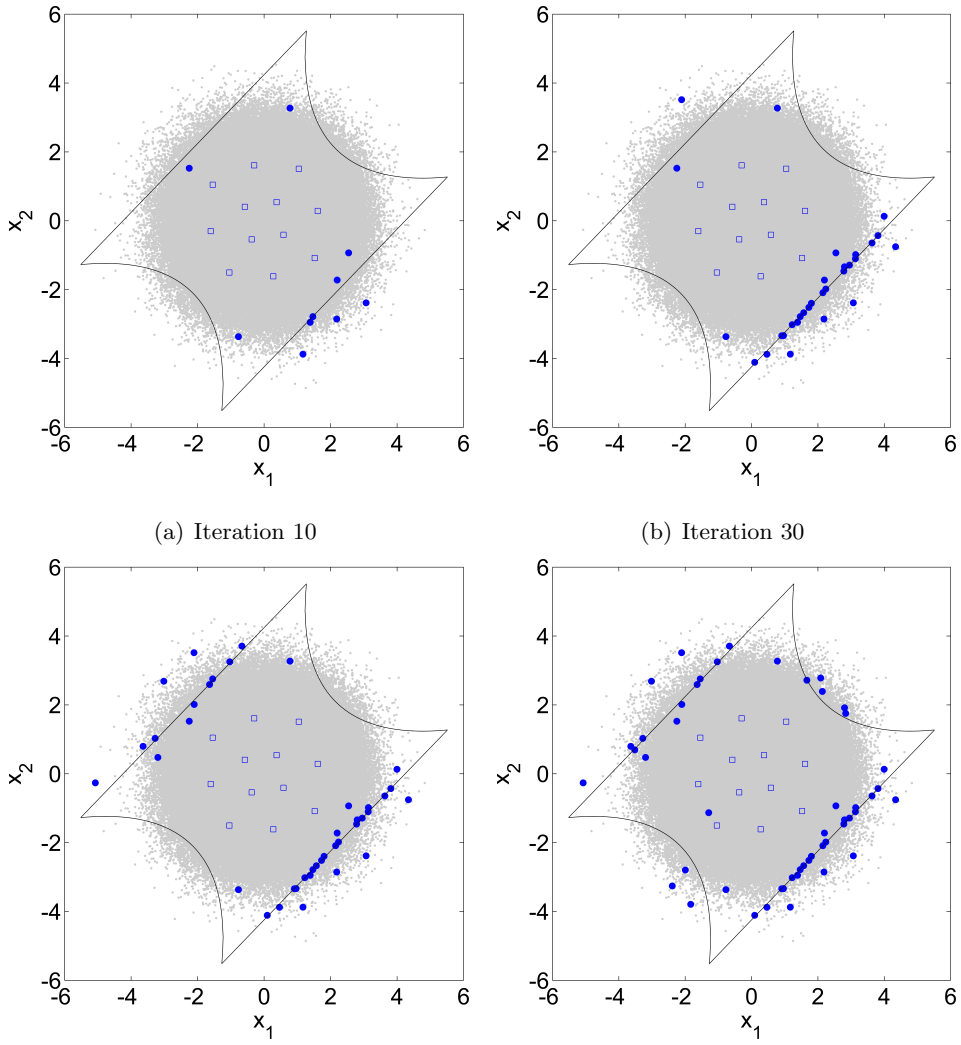
Setting The adaptive experimental design algorithm is initiated with $N_0 = 12$ Latin-hypercube samples (LHS) and a candidate Monte Carlo population $\mathcal{S}$ of $n = 10^6$ samples. The performances of PC-Kriging (PCK) are compared to those of ordinary Kriging (OK).

The trend in PC-Kriging consists of Hermite polynomials of maximal degree of eight and a Gaussian autocorrelation function is chosen for the Gaussian process. Both meta-modeling techniques are calibrated using the Matlab-based toolbox UQLab (Marelli and Sudret, 2014). The adaptive experimental design algorithm is tested in both single and multiple ($K = 6$) samples mode.

Visualization Figure 3 and 4 visualize several iterations of the algorithm for OK and PCK, respectively, for single sample enrichment. The gray dots represent the candidate MC population $\mathcal{S}$. The empty squares mark the initial experimental design, whereas the blue filled circles mark the additional samples. The solid black line represents the exact limit state surface.

Both variants discover the four distinct failure modes in the computational model within the first 50 iterations. The main difference lies in the limit state exploration pattern. When using ordinary Kriging in Figure 3, the distinct failure mechanisms are explored one-by-one. As soon as a failure mechanism is characterized, the adaptive sampling algorithm moves to the next. At iteration 50 all four failure mechanisms are discovered. In the case of PC-Kriging, all four failure mechanisms are already discovered after 10 iterations (Figure 4(a)). Thus, PC-Kriging converges on average faster to an accurate estimate of the failure probability as compared to ordinary Kriging. This is to be expected due to the more accurate global trend in PC-Kriging. Figure 5 illustrates the convergence of the estimate of the failure probability by showing $\widehat{P}_f^{\pm}$ ($k = 2$) and $\widehat{P}_f^0$. Further, the final iteration is marked where the iterative algorithm is stopped due to the stopping criterion in Eq. (31). Figure 5 confirms the faster convergence of the PC-Kriging meta-models. Note that for the Kriging meta-model, Figure 5 shows four plateaus in $\widehat{P}_f^0$ which correspond to the exploration of the four branches of the limit state function.

In order to test the statistical significance of the convergence, the analysis is replicated 50 times with different initial experimental designs. Figure 6 illustrates the convergence of the failure probability estimates $\widehat{P}_f^0$ based on ordinary Kriging and PC-Kriging up to iteration 100, respectively. The solid line represents the mean value of 50 independent replications of the same analysis with different initial Latin-hypercube experimental designs. The dashed



(a) Iteration 10

(b) Iteration 30

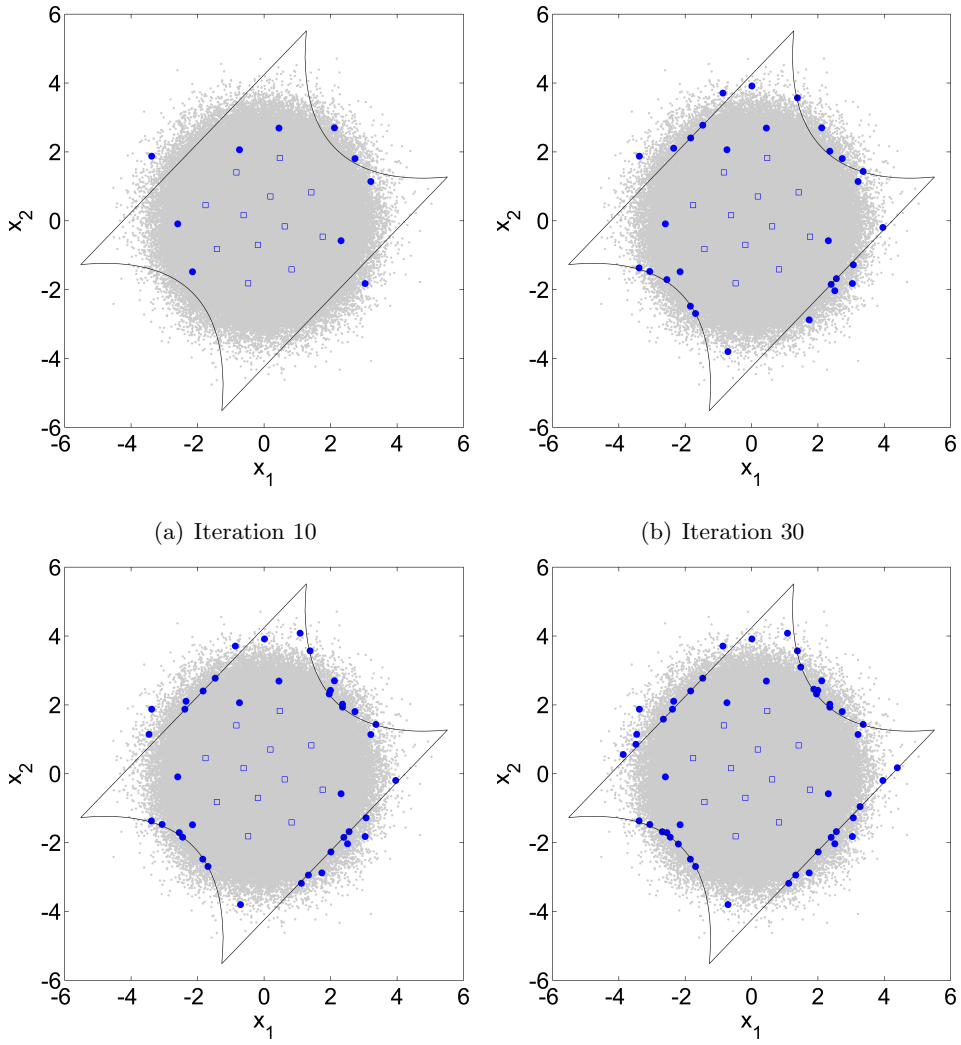
(c) Iteration 40

(d) Iteration 50

Figure 3: Four-branch function – iterations of the adaptive experimental design algorithm with ordinary Kriging meta-models.

lines represent the 5% and 95% quantiles of the 50 independent replications, *i.e.* the 90% confidence interval (CI). The results show that PC-Kriging converges faster than ordinary Kriging. Note that the variation in $\widehat{P}_f$ after convergence (iterations ≥ 80) originates from the finite size of $\mathcal{S}$ and the corresponding variance in Monte Carlo simulations (see also Eq. (5)).

Parametric study In this paragraph we discuss the candidate selection algorithm and the choice of stopping criterion. The candidate selection algorithm compares the selection of the



(a) Iteration 10

(b) Iteration 30

(c) Iteration 40

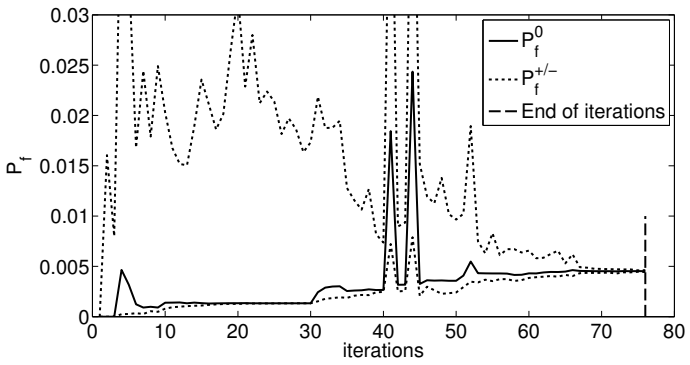
(d) Iteration 50

Figure 4: Four-branch function – iterations of the adaptive experimental design algorithm with PC-Kriging meta-models.

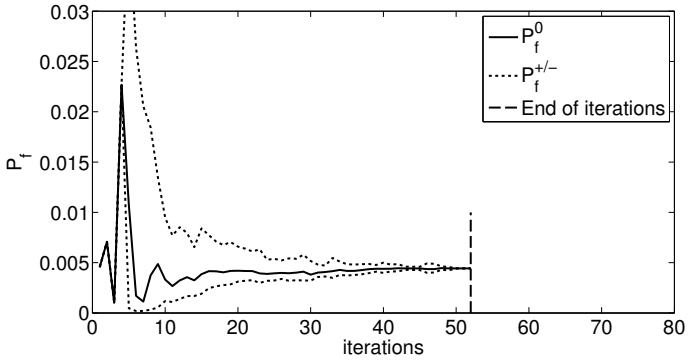
single optimal candidate to the selection of multiple candidates through weighted K-means clustering. In terms of stopping criterion we compare Eq. (31) ($k = 2$) to the existing stopping criterion defined in Echard et al. (2011):

$$\min[U(\boldsymbol{x})] \geq 2 \; \forall \boldsymbol{x} \in \mathcal{S}, \tag{36}$$

which indicates that the probability of misclassification must be smaller than $\Phi(-2) \approx 2$ % for *all candidate samples* in order to stop the iterations.



(a) Ordinary Kriging



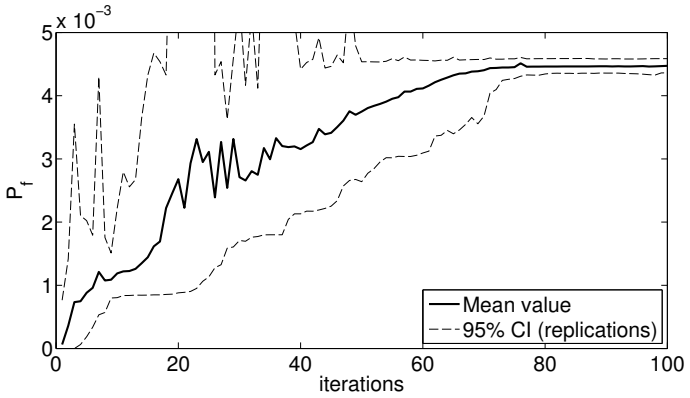
(b) PC-Kriging

Figure 5: Four-branch function – convergence curves of the adaptive experimental design algorithm for the single run in Figure 3 and 4

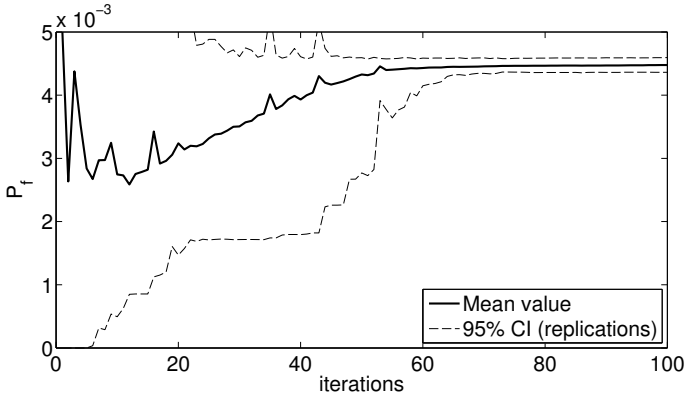
The full failure probability analysis is replicated 50 times with different initial LHS experimental design in order to assess the statistical significance (*e.g.* standard deviation of the resulting $\widehat{P}_f$). The analysis is focused on the total number of computational model evaluations and on the accuracy of the estimate of the failure probability. The total number of computational model runs is $N_{tot} = N_0 + N_1$, where N_1 is the number of samples added by the adaptive algorithm.

Table 1 summarizes the results for the four-branch function. The results of the adaptive experimental design algorithm are presented for both stopping criteria (Eq. (31) and Eq. (36)) and meta-models (ordinary Kriging and PC-Kriging). The reference value for the failure probability is computed by Monte Carlo simulation with 10^8 samples.

Considering the cases of adding a single candidate, all configurations estimate accurately



(a) Ordinary Kriging



(b) PC-Kriging

Figure 6: Four-branch function – convergence curves of the adaptive experimental design algorithms (statistics of 50 replications of the analysis with different initial Latin-hypercube experimental design).

the failure probability. The coefficient of variation of the estimate of the failure probability $\mathrm{CoV}\left[\widehat{P}_f\right]$ is comparable with the theoretical coefficient of variation of a Monte Carlo sampling (see Eq. (5)) with $n_{MC} = 10^6$ samples, *i.e.* $\mathrm{CoV}\left[\widehat{P}_f = 4.46 \cdot 10^{-3}, n_{MC} = 10^6\right] \approx 1.5$ %.

The results in Table 1 show that the stopping criterion in Eq. (31) leads to accurate results despite the smaller number of computational model runs N_{tot} when compared to the stopping criterion in Eq. (36). This indicates that Eq. (31) is more suitable to the estimation of failure probabilities than Eq. (36). It is interesting though that PC-Kriging converges faster than ordinary Kriging with the stopping criterion in Eq. (31), as opposed to the one in Eq. (36). The

Table 1: Four-branch function – results using different meta-models, candidate selection algorithms and stopping criteria. Coefficients of variation CoV are computed based on the 50 replications of the analysis with different initial experimental designs.

Method	Enrich.	Stop. crit.	$\mathbb{E}\left[\widehat{P}_f\right]$	CoV $\left[\widehat{P}_f\right]$	$\mathbb{E}\left[N_{tot}\right]$	CoV $\left[N_1\right]$
MC			$4.460 \cdot 10^{-3}$	0.15 %	10^8	-
OK	single	Eq. (36)	$4.464 \cdot 10^{-3}$	1.4 %	$12 + 96.4 = 108.4$	6.7 %
PCK	single	Eq. (36)	$4.471 \cdot 10^{-3}$	1.4 %	$12 + 115.8 = 127.8$	24.5 %
OK (*)	single	Eq. (36)	$4.416 \cdot 10^{-3}$	-	126	-
OK	single	Eq. (31)	$4.440 \cdot 10^{-3}$	1.6 %	$12 + 66.3 = 78.3$	11.0 %
PCK	single	Eq. (31)	$4.457 \cdot 10^{-3}$	1.5 %	$12 + 61.2 = 73.2$	28.5 %
PCK	$K = 6$	Eq. (31)	$4.458 \cdot 10^{-3}$	1.5 %	$12 + 14.4 \cdot 6 = 98.4$	8.9 %

(*) single run results from Echard et al. (2011)

explanation lies in the fact that the stopping criterion in Eq. (31) measures the convergence of the statistics of interest whereas Eq. (36) assesses the overall accuracy of the approximation of the Kriging surrogate.

For the case of PCK and the stopping criterion in Eq. (31), Table 1 includes also the results obtained by adding $K = 6$ candidates at each iteration. The comparison of the total number of computational model runs N_{tot} to the case of adding a single point in each iteration shows that the addition of multiple points at each iteration requires a larger N_{tot}: indeed in the case of single point enrichment, the total number of model runs is $N_{tot} = 73$, whereas adding $K = 6$ points in each step leads to $N_{tot} = 98$. Considering however that the runs of the computational model can be performed in parallel ($K = 6$ on six independent CPUs), the total time for the K-point enrichment algorithm is approximately four times smaller.

5.1.3 Quantile estimation

Consider now the quantile estimation problem related to the four-branch function. The quantiles are estimated for $\alpha = \{0.01, 0.001, 0.0001\}$. PC-Kriging is used, a single sample is added at each iteration out of a candidate Monte Carlo population of $n = 10^6$ samples and the algorithm is stopped with the stopping criterion in Eq. (34).

Table 2 summarizes the results and compares the estimate of the quantiles $\widehat{q}_\alpha$ with a reference solution from a Monte Carlo simulation ($n_{MC} = 10^8$). The three α values are accurately estimated with the adaptive algorithm despite the small initial experimental design of $N_0 = 12$ samples. The coefficient of variation of the results increases with decreasing α due to the decreasing size of the failure domain.

It is interesting to notice that the smaller the α value is, the fewer samples are required to estimate the quantile. This phenomenon can be explained by considering the failure domain defined by Eq. (35). For small values of α, indeed the failure domain is also small. Hence to meta-model the limit state surface, a smaller number of experimental design points is required.

Table 2: Four-branch function – quantile estimation results. MC is based $n_{MC} = 10^8$ samples. Coefficients of variation CoV are computed based on 50 replications of the analysis with different initial experimental designs.

α	MC	$\mathbb{E}\left[\widehat{q}_\alpha\right]$	CoV $\left[\widehat{q}_\alpha\right]$	$\mathbb{E}\left[N_{tot}\right]$	CoV $\left[N_1\right]$
0.01	0.303	0.305	1.4 %	$12 + 62.4 = 74.4$	21.0 %
0.001	-0.528	-0.531	2.1 %	$12 + 57.7 = 69.7$	23.3 %
0.0001	-1.299	-1.292	2.4 %	$12 + 47.9 = 59.9$	13.6 %

5.2 Borehole model

5.2.1 Definition

The first realistic engineering problem considered is the so-called *borehole-function*, which describes the water flow through a borehole. This benchmark function has been discussed in papers such as Harper and Gupta (1983), Morris et al. (1993), An and Owen (2001), Kersaudy et al. (2015). It is a fast-to-evaluate function (Harper and Gupta, 1983) depending on an eight-dimensional input vector $\boldsymbol{x} = [r_w, r, T_u, H_u, T_l, H_l, L, K_w]^\mathsf{T}$:

$$v(\boldsymbol{x}) = \frac{2\pi T_u (H_u - H_l)}{\ln(r/r_w)\left(1 + \frac{2LT_u}{\ln(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l}\right)}, \tag{37}$$

where $v(\boldsymbol{x})$ is the fluid water flow measured in m^3/year, r_w is the radius of the borehole, r the radius of influence, T_u the transmissivity of the upper aquifer, H_u the potentiometric head of

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

The uncertainties in the input vector are modeled as independent random variables whose properties are summarized in Table 3. For the lognormal distribution, the parameters are the mean and standard deviation of the natural logarithm of the variable. For the other variables, they describe the range of uniform distributions.

Table 3: Borehole model – definition of the probabilistic model of the input variables. For the uniform distributions the parameters denote the range, whereas for the lognormal distribution, the parameters denote the mean and standard deviation of the natural logarithm of the variable.

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

5.2.2 Quantile estimation

The statistics of interest are the quantiles of the water flow $v(\boldsymbol{x})$ corresponding to $\alpha = \{0.99, 0.999, 0.9999\}$. An initial experimental design of $N_0 = 12$ Latin-hypercube samples in used. The number of samples in the candidate Monte Carlo population is set to $n = 10^6$ and the remaining settings are the same as in the four-branch function.

The results of the adaptive algorithm and the reference values are reported in Table 4 for the various values of α, the meta-modeling method (PCK and OK) and enrichment strategies (single or multiple) used.

The prediction of the quantiles is accurate for both meta-modeling techniques (OK and PCK)

Table 4: Borehole model – results of quantile estimation based on a single run of the adaptive experimental design algorithm with stopping criterion defined in Eq. (34).

α	Method	Enrichment	$\widehat{q}_\alpha$	N_{tot}
0.99	MC	-	157.5 m^3/year	10^6
	OK	single	157.3 m^3/year	$12 + 199 = 211$
	PCK	single	157.4 m^3/year	$12 + 34 = 46$
	OK	$K = 6$	157.5 m^3/year	$12 + 41 \cdot 6 = 258$
	PCK	$K = 6$	157.7 m^3/year	$12 + 10 \cdot 6 = 72$
0.999	MC	-	197.8 m^3/year	10^6
	OK	single	197.8 m^3/year	$12 + 137 = 149$
	PCK	single	198.2 m^3/year	$12 + 30 = 42$
	OK	$K = 6$	198.0 m^3/year	$12 + 32 \cdot 6 = 204$
	PCK	$K = 6$	198.2 m^3/year	$12 + 4 \cdot 6 = 32$
0.9999	MC	-	235.0 m^3/year	10^6
	OK	single	235.2 m^3/year	$12 + 143 = 155$
	PCK	single	235.2 m^3/year	$12 + 26 = 38$
	OK	$K = 6$	235.3 m^3/year	$12 + 33 \cdot 6 = 210$
	PCK	$K = 6$	235.3 m^3/year	$12 + 10 \cdot 6 = 72$

compared to the Monte Carlo ($n_{MC} = 10^6$) solution when adding a single sample and multiple samples to the experimental design in each iteration. There is, however, a noticeable difference in the total number of evaluations of the computational model. Ordinary Kriging (OK) requires more model evaluations than PC-Kriging (PCK) for all values of α. The addition of $K = 6$ samples at a time induces slightly more model evaluations N_{tot} but at the same time reduces the total computational time due to the parallel evaluation of $K = 6$ samples.

5.3 Two-dimensional truss structure

5.3.1 Definition

For the purpose of illustrating a realistic structural engineering application, a two-dimensional truss structure is analyzed. This specific truss example structure has been presented and discussed previously in Lee and Kwak (2006), Sudret et al. (2007), Blatman and Sudret (2008) and Blatman and Sudret (2010).

Consider the simply supported two-dimensional truss structure sketched in Figure 7 consisting of 23 bars and 13 nodes. The geometry is known deterministically whereas the material properties and the loadings are modeled stochastically. Ten stochastic input variables form the input vector $\boldsymbol{X}$:

$$\boldsymbol{X} = [E_1, E_2, A_1, A_2, P_1, P_2, P_3, P_4, P_5, P_6]^{\mathsf{T}}, \tag{38}$$

where E_1, E_2 are the Young's moduli of the linear elastic material, A_1, A_2 are the cross-sections of the bars and $P_1, \ldots, P_6$ are the vertical loads acting on the nodes of the upper part of the structure. The horizontal bars have the properties $\{E_1, A_1\}$ whereas the diagonal bars have the properties $\{E_2, A_2\}$. The input variables are modeled by the probability distributions summarized in Table 5. It is assumed that the input variables are statistically independent. A finite element model of the structure is used to calculate the mid-span deflection, denoted by $u(\boldsymbol{x})$, as a function of the ten variables in the input vector $\boldsymbol{x}$, which is defined as positive in the direction indicated in Figure 7.

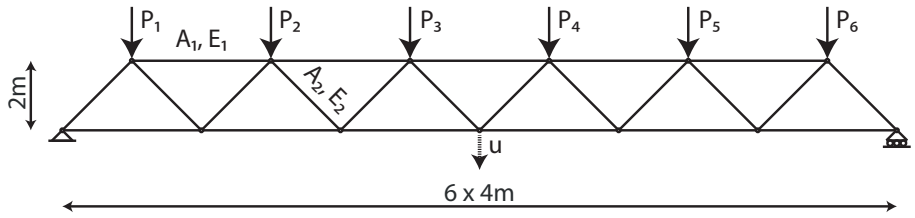


Figure 7: Truss structure – sketch of the geometry, material parameters $\{A_i, E_i\}$ and loads P_i.

Table 5: Truss structure – definition of the probabilistic model of the input variables

Variable		Distribution	Mean	Standard deviation
$E_1,\ E_2$	[Pa]	Lognormal	$2.1 \cdot 10^{11}$	$2.1 \cdot 10^{10}$
A_1	[m^2]	Lognormal	$2.0 \cdot 10^{-3}$	$2 \cdot 10^{-4}$
A_2	[m^2]	Lognormal	$1.0 \cdot 10^{-3}$	$1 \cdot 10^{-4}$
$P_1, \ldots, P_6$	[N]	Gumbel	$5.0 \cdot 10^4$	$7.5 \cdot 10^3$

5.3.2 Failure probability estimation

The failure probability is defined as the probability that the mid-span deflection u is larger than an admissible value u_{adm}, *i.e.* $P_f = \mathbb{P}\left(u_{adm} - u(\boldsymbol{X}) \leq 0\right)$.

The adaptive experimental design algorithm is started with an initial experimental design of $N_0 = 12$ Latin-hypercube samples and the candidate Monte Carlo population $\mathcal{S}$ has a size of $n = 10^6$ samples. A single sample or $K = 6$ samples are added to the experimental design at each iteration.

Table 6 summarizes the results of the failure probability estimation. The results are compared to a reference plain Monte Carlo simulation with $n_{MC} = 10^6$ samples and to the *first order reliability method* (FORM) in terms of failure probability and Hasofer-Lind reliability index (Zhang and Der Kiureghian, 1995). For comparison purposes the generalized reliability index $\widehat{\beta} = \Phi^{-1}(\widehat{P}_f)$ is also given for the various sampling-based approaches.

The values of $\widehat{P}_f$ shown in Table 6 are consistent with the reference for all the adaptive experimental design configurations. The number of runs of the computational model, however, differ considerably. For the cases of $u_{adm} = 10$ cm and 12 cm the number of exact model runs are similar for both ordinary Kriging (OK) and PC-Kriging (PCK). In the case of $u_{adm} = 0.14$ cm, PCK requires significantly fewer runs than OK, for both single and multiple sample selection modes. Similarly to the borehole example, multiple sample selection results in more computational model runs, but overall lower computational times (up to 5-times smaller) due to high performance computing.

The well-established FORM gives an estimate worse than the adaptive experimental design algorithm despite the higher cost. It is thus preferable in this application to conduct an

Table 6: Truss structure – results of the failure probability estimate based on a single run of the adaptive experimental design algorithm for different meta-modelling techniques and enrichment modes.

u_{adm}	Method	Enrichment	$\widehat{P}_f$ (CoV $[P_f]$)	$\widehat{\beta}$	N_{tot}
10 cm	MC	-	$4.29 \cdot 10^{-2}$ (0.5 %)	1.72	10^6
	FORM	-	$2.81 \cdot 10^{-2}$	1.91	251
	OK	single	$4.32 \cdot 10^{-2}$	1.71	$12 + 135 = 147$
	PCK	single	$4.32 \cdot 10^{-2}$	1.71	$12 + 158 = 170$
	OK	$K = 6$	$4.31 \cdot 10^{-2}$	1.72	$12 + 26 \cdot 6 = 168$
	PCK	$K = 6$	$4.32 \cdot 10^{-2}$	1.71	$12 + 31 \cdot 6 = 198$
12 cm	MC	-	$1.55 \cdot 10^{-3}$ (2.5 %)	2.96	10^6
	FORM	-	$7.57 \cdot 10^{-4}$	3.17	236
	OK	single	$1.53 \cdot 10^{-3}$	2.96	$12 + 164 = 176$
	PCK	single	$1.52 \cdot 10^{-3}$	2.96	$12 + 157 = 169$
	OK	$K = 6$	$1.53 \cdot 10^{-3}$	2.96	$12 + 27 \cdot 6 = 174$
	PCK	$K = 6$	$1.53 \cdot 10^{-3}$	2.96	$12 + 25 \cdot 6 = 162$
14 cm	MC	-	$3.6 \cdot 10^{-5}$ (16.7 %)	3.97	10^6
	FORM	-	$1.29 \cdot 10^{-5}$	4.21	231
	OK	single	$3.7 \cdot 10^{-5}$	3.96	$12 + 110 = 122$
	PCK	single	$3.7 \cdot 10^{-5}$	3.96	$12 + 63 = 75$
	OK	$K = 6$	$3.4 \cdot 10^{-5}$	3.99	$12 + 27 \cdot 6 = 174$
	PCK	$K = 6$	$3.2 \cdot 10^{-5}$	4.00	$12 + 11 \cdot 6 = 78$

adaptive experimental design search rather than using FORM algorithm.

6 Conclusions

In this paper, a unified framework for the estimation of both failure probabilities and quantiles is introduced in the context of rare event estimation. The combination of PC-Kriging and design enrichment (adaptive experimental design) based on the probability of misclassification increases the accuracy in the estimation of the statistics of interest.

An efficient strategy to add multiple samples at each iteration of the adaptive experimental design algorithm is introduced to enable the use of available high-performance computing resources.

The adaptive experimental design algorithm is also equipped with a new stopping criterion which monitors the convergence of the statistics of interest better than existing stopping criteria. This results in further reduction of the computational resources needed.

The performance of the proposed algorithm is assessed and illustrated through a benchmark analytical function, which shows the algorithm's capabilities in a variety of settings. The results show that the proposed strategy provides an efficient way to estimate failure probabilities as well as quantiles. Two engineering applications (the water flow through a borehole and the deflection of a truss structure modeled by finite elements) confirm the strengths of the adaptive experimental design algorithm in the context of realistic high-dimensional engineering problem settings, and thus its significance for the engineering practice.

References

An, J. and A. Owen (2001). Quasi-regression. *J. Complexity 17*(4), 588–607.

Arnaud, A., J. Bect, M. Couplet, A. Pasanisi, and E. Vazquez (2010). Evaluation d'un risque d'inondation fluviale par planification séquentielle d'expériences. In *42èmes Journées de Statistique*, Marseille, France, pp. 1–6.

Bect, J., D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Stat. Comput. 22*(3), 773–793.

Bichon, B., M. Eldred, L. Swiler, S. Mahadevan, and J. McFarland (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal 46*(10), 2459–2468.

Blatman, G. and B. Sudret (2008). Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *Comptes Rendus Mécanique 336*(6), 518–523.

Blatman, G. and B. Sudret (2010). An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Prob. Eng. Mech. 25*, 183–197.

Blatman, G. and B. Sudret (2011). Adaptive sparse polynomial chaos expansion based on Least Angle Regression. *J. Comput. Phys 230*, 2345–2367.

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

Cadini, F., F. Santos, and E. Zio (2014). An improved adaptive Kriging-based importance technique for sampling multiple failure regions of low probability. *Reliab. Eng. Sys. Safety 131*, 109–117.

Dai, H., H. Zhang, K. Rasmussen, and W. Wang (2014). Wavelet density-based adaptive importance sampling method. *Structural Safety 52*, 161–169.

Dani, V., T. P. Hayes, and S. M. Kakade (2008). Stochastic linear optimization under bandit feedback. In *The 21st Annual Conference on Learning Theory (COLT 2008)*.

De Rocquigny, E., N. Devictor, and S. Tarantola (Eds.) (2008). *Uncertainty in industrial practice – A guide to quantitative uncertainty management*. John Wiley & Sons.

Dubourg, V. (2011). *Adaptive surrogate models for reliability analysis and reliability-based design optimization*. Ph. D. thesis, Université Blaise Pascal, Clermont-Ferrand, France.

Dubourg, V. and B. Sudret (2014). Metamodel-based importance sampling for reliability sensitivity analysis. *Structural Safety 49*, 27–36.

Dubourg, V., B. Sudret, and F. Deheeger (2013). Metamodel-based importance sampling for structural reliability analysis. *Prob. Eng. Mech. 33*, 47–57.

Echard, B., N. Gayton, and M. Lemaire (2011). AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation. *Structural Safety 33*(2), 145–154.

Echard, B., N. Gayton, M. Lemaire, and N. Relun (2013). A combined importance sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models. *Reliab. Eng. Syst. Safety 111*, 232–240.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *Annals of Statistics 32*, 407–499.

Ghanem, R. and P. Spanos (2003). *Stochastic Finite Elements: A Spectral Approach* (2nd ed.). Courier Dover Publications, Mineola.

Ginsbourger, D., B. Rosspopoff, G. Pirot, N. Durrande, and R. Renard (2013). Distance-based Kriging relying on proxy simulation for inverse conditioning. *Adv. Water Resources 52*, 275–291.

Gunn, S. (1998). Support vector machines for classification and regression. Technical Report ISIS-1-98, Dpt. of Electronics and Computer Science, University of Southampton.

Harper, W. V. and S. K. Gupta (1983). Sensitivity/uncertainty analysis of a borehole scenario comparing Latin hypercube sampling and deterministic sensitivity approaches. Technical Report No. BMI/ONWI-516, Battelle Memorial Inst. - Office of Nuclear Waste Isolation, Columbus, OH (USA).

Hurtado, J. E. and D. A. Alvarez (2001). Neural-network-based reliability analysis: a comparative study. *Comput. Method. Appl. M. 191*(1-2), 113–132.

Jala, M. (2013). *Plans d'expériences adaptatifs pour le calcul de quantiles et application à la dosimétrie numérique*. Ph. D. thesis, Institut Mines-Télécom, Paris, France.

Jones, D., M. Schonlau, and W. Welch (1998). Efficient global optimization of expensive black-box functions. *J. Global Optim. 13*(4), 455–492.

Kaymaz, I. (2005). Application of kriging method to structural reliability problems. *Structural Safety 27*(2), 133–151.

Kersaudy, P., B. Sudret, N. Varsier, O. Picon, and J. Wiart (2015). A new surrogate modeling technique combining Kriging and polynomial chaos expansions – Application to uncertainty analysis in computational dosimetry. *J. Comput. Phys 286*, 103–117.

Lebrun, R. and A. Dutfoy (2009a). A generalization of the Nataf transformation to distributions with elliptical copula. *Prob. Eng. Mech. 24*(2), 172–178.

Lebrun, R. and A. Dutfoy (2009b). An innovating analysis of the Nataf transformation from the copula viewpoint. *Prob. Eng. Mech. 24*(3), 312–320.

Lee, S. and B. Kwak (2006). Response surface augmented moment method for efficient reliability analysis. *Structural Safety 28*, 261–272.

Lemaire, M. (2009). *Structural reliability*. Wiley.

Marelli, S. and B. Sudret (2014). UQLab: A framework for uncertainty quantification in Matlab. In *Vulnerability, Uncertainty, and Risk (Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom)*, pp. 2554–2563.

Morris, M., T. Mitchell, and D. Ylvisaker (1993). Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics 35*, 243–255.

Ranjan, P., D. Bingham, and G. Michailidis (2012). Sequential experiment design for contour estimation from complex computer codes. *Technometrics 50*(4), 527–541.

Rasmussen, C. and C. Williams (2006). *Gaussian processes for machine learning* (Internet ed.). Adaptive computation and machine learning. Cambridge, Massachusetts: MIT Press.

Santner, T., B. Williams, and W. Notz (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.

Schöbi, R. and B. Sudret (2014). PC-Kriging: a new metamodelling method combining polynomial chaos expansions and Kriging. In *Proc. 2nd Int. Symposium on Uncertainty Quantification and Stochastic Modeling, Rouen, France*.

Schöbi, R., B. Sudret, and J. Wiart (2015). Polynomial-chaos-based Kriging. *Int. J. Uncertainty Quantification 5*(2), 171–193.

Schueremans, L. and D. Van Gemert (2005a). Benefit of splines and neural networks in simulation based structural reliability analysis. *Structural Safety 27*(3), 246–261.

Schueremans, L. and D. Van Gemert (2005b). Use of Kriging as meta-model in simulation procedures for structural reliability. In *Proc. 9th Int. Conf. Struct. Safety and Reliability (ICOSSAR'2005), Rome, Italy*, pp. 2483–2490.

Srinivas, N., A. Krause, S. Kakade, and M. Seeger (2012). Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Trans. Inform. Theory 58*(5), 3250–3265.

Sudret, B. (2007). Uncertainty propagation and sensitivity analysis in mechanical models – contributions to structural reliability and stochastic spectral methods. Technical report. Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand, France (229 pages).

Sudret, B., G. Blatman, and M. Berveiller (2007). Quasi random numbers in stochastic finite element analysis – Application to global sensitivity analysis. In J. Kanda, T. Takada, and H. Furuta (Eds.), *Proc. 10th Int. Conf. on Applications of Stat. and Prob. in Civil Engineering (ICASP10), Tokyo, Japan*.

Waarts, P.-H. (2000). *Structural reliability using finite element methods: an appraisal of DARS: Directional Adaptive Response Surface Sampling*. Ph. D. thesis, Technical University of Delft, The Netherlands.

Xiu, D. and G. Karniadakis (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput. 24*(2), 619–644.

Zaki, M. and W. Meira (2014). *Data Mining and Analysis: fundamental Concepts and Algorithms*. Cambridge University Press.

Zhang, Y. and A. Der Kiureghian (1995). Two improved algorithms for reliability analysis. In R. Rackwitz, G. Augusti, and A. Bori (Eds.), *Proc. 6th IFIP WG7.5 on Reliability and Optimization of Structural systems, Assisi, Italy*. Chapman & Hall, London.