



HAL
open science

A new adaptive response surface method for reliability analysis

Nicolas Roussouly, Frank Petitjean, Michel Salaün

► **To cite this version:**

Nicolas Roussouly, Frank Petitjean, Michel Salaün. A new adaptive response surface method for reliability analysis. Probabilistic Engineering Mechanics, 2013, vol. 32, pp. 103-115. 10.1016/j.probengmech.2012.10.001 . hal-00813479

HAL Id: hal-00813479

<https://hal.science/hal-00813479>

Submitted on 15 Apr 2013

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.



Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <http://oatao.univ-toulouse.fr/>
Eprints ID: 8873

To link to this article: DOI: 10.1016/j.probengmech.2012.10.001
URL: <http://dx.doi.org/10.1016/j.probengmech.2012.10.001>

To cite this version: Roussouly, Nicolas and Petitjean, Frank and Salaün, Michel *A new adaptive response surface method for reliability analysis.* (2013) Probabilistic Engineering Mechanics, vol. 32 . pp. 103-115. ISSN 0266-8920

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@inp-toulouse.fr

A new adaptive response surface method for reliability analysis

N. Roussouly^{a,b,*}, F. Petitjean^b, M. Salaun^a

^a Université de Toulouse, Institut Clément Ader (ICA), ISAE, 10 av. Edouard Belin, 31055 Toulouse, France

^b ICAM, 75 avenue de Grande Bretagne, 31300 Toulouse, France

ARTICLE INFO

Keywords:

Structural reliability
Adaptive response surface
Regression
Prediction error
Bootstrap

ABSTRACT

Response surface method is a convenient tool to assess reliability for a wide range of structural mechanical problems. More specifically, adaptive schemes which consist in iteratively refine the experimental design close to the limit state have received much attention. However, it is generally difficult to take into account a lot of variables and to well handle approximation error. The method, proposed in this paper, addresses these points using sparse response surface and a relevant criterion for results accuracy. For this purpose, a response surface is built from an initial Latin Hypercube Sampling (LHS) where the most significant terms are chosen from statistical criteria and cross-validation method. At each step, LHS is refined in a region of interest defined with respect to an importance level on probability density in the design point. Two convergence criteria are used in the procedure: The first one concerns localization of the region and the second one the response surface quality. Finally, a bootstrap method is used to determine the influence of the response error on the estimated probability of failure. This method is applied to several examples and results are discussed.

1. Introduction

1.1. Problem statement

In mechanical structures, consideration of uncertainties in modeling is a growing topic because it provides valuable information in industrial applications. Indeed, sensitivity, risk or financial analyses, which enable to take relevant decisions by comparing cost and precision of design solution or manufacturing process, are all common problematics with uncertainty analysis. These uncertainties operate in several stages of a modeling procedure, but we shall only consider in this paper those which affect input parameters. More precisely, modeling of a physical system can be seen as a mathematical function which depends on input parameters and which provides one or several responses. In mechanical structures, the mathematical function is generally defined using finite element method. Input variables characterize geometry, materials or loads and responses can be any kind of mechanical quantities. Here, the probabilistic framework is considered, which means that input parameters are realizations of random variables or random fields. The stake of the analysis is thus to propagate uncertainties through a mathematical model in order

to get statistical informations on outputs. Our main interest concerns reliability analysis, which is performed through the probability of failure calculation.

Let us introduce $\mathbf{X} = (X_1, \dots, X_M)$ a vector of M random variables, characterized by its joint probability density function known as $f_{\mathbf{X}}(\mathbf{x})$. Suppose that one studies a mechanical system, described by a performance function, say $G(\mathbf{X})$, defined such as, for a realization \mathbf{x} of vector \mathbf{X} , $G(\mathbf{x}) < 0$ is called the domain of failure, $G(\mathbf{x}) = 0$ is the limit state and $G(\mathbf{x}) > 0$ is the domain of success. The probability of failure relative to the failure mode described by function G reads

$$P_f = \int_{G(\mathbf{x}) \leq 0} f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_M. \quad (1)$$

A simple and robust estimation of this probability is given by Monte Carlo method. However, in a lot of engineering problems, the limit state function comes from finite element discretization and is thus very expensive to evaluate. Moreover, when the probability of failure is low, crude Monte Carlo involves a large number of numerical simulations.

In order to overcome this difficulty, alternative methods have been proposed. Some of them use variance reduction techniques like Importance Sampling, Line Sampling [38] or Subset Simulation [2]. These methods are generally still expensive and some of them can be less efficient if a large number of input variables is taken into account. In order to be independent of the dimension, it is proposed in [9,42,43,34] to keep a crude Monte Carlo method but to make use of the regularity of tail probabilities. The principle

* Corresponding author.

E-mail addresses: nicolas.roussouly@gmail.com, nicolas.roussouly@icam.fr (N. Roussouly), frank.petitjean@icam.fr (F. Petitjean), michel.salaun@isae.fr (M. Salaun).

is to estimate probabilities of failure at moderate levels and to approximate the targeted probability of failure at low level. Although these approaches have attractive features, they can require a large number of simulations for industrial applications. Famous reliability methods, known as first and second reliability method (FORM, SORM), allow also to calculate low probability of failure [24]. Based on optimization procedures and Taylor expansions, these methods can be difficult to validate. Another kind of methods consists in substituting the initial expensive model by a simpler and fast to evaluate one. This new model, called meta-model, can be of various kinds: quadratic response surfaces [21], polynomial chaos [23,6], kriging [40,27], neural network [36,13,20] or even support vector machine [25]. In this paper, quadratic response surfaces, simply called response surfaces, will be used.

1.2. Short review of response surface approach

Within the current literature of reliability, analysis based on meta-models such as kriging, neural network or support vector machine is increasing since mechanical models are more and more complicated. Indeed, these meta-models can deal with highly non-linear relations between inputs and outputs and are thus efficient when several domains of failure and/or several design points exist. However, response surfaces (RS) are still widely used in structural reliability analysis because of their simplicity and effectiveness. Actually, in a lot of industrial applications, mechanical models have not always complicated behavior and the relation between inputs and outputs is not highly non-linear. Generally, difficulties are more related to the number of variables and sizing criteria which can be relatively or even very large.

RS method is an interesting tool when relation between inputs and outputs is quite regular. But problems of computational time may arise even when a moderate number of input variables is taken into account. Also, although this method has been widely used, it is often difficult to be very confident in results because of the difficulty to quantify error due to approximation. The present paper is oriented in these two issues.

Whatever the used meta-model, the current trend is to consider an adaptive strategy. Its principle is to build a meta-model which is refined, iteratively, in some regions of interest. These regions are close to the limit state and contribute significantly to the probability of failure. Recent contributions have been made on kriging [37,14,4] and support vector machine [8,12].

About RS, a lot of improvements have been proposed from the first use in reliability [21]. Adaptive schemes are also of great interest since they allow to get more precise results with a reduced number of simulations. Despite their differences, main steps of procedures are always the same and can be summarized as follows:

1. choose an initial experimental design (ED);
2. build a RS;
3. find the design point (and the reliability index) based on the RS;
4. add new sampling points around the design point;
5. repeat from the step 2 until a convergence criterion is satisfied.

The idea was first introduced by Bucher and Bourgund [10] who used, in turn, two RS. The first one is built from a star shaped ED around the mean point. A linear interpolation between the mean point and the design point is then used to determine the central point of the second and smaller ED. A second RS is finally built up, closer than the first one, to the design point. In this approach, only linear and square terms are used because cross terms would involve a too large number of additional sampling points. Rajashekhar and Ellingwood [39] suggest to improve this method by considering several iterations

until a convergence criterion is satisfied. This criterion is the distance between the design point and the central point of the ED, which must be small enough. Authors also take into account that the failure region is located in the tails of probability distributions and thus introduce several techniques to start the algorithm from this region. Finally, they consider cross terms in RS. In a similar way, Kim and Na [28] suggest to project star shaped sampling points of the ED on a linear approximation of the limit state, in order to be close to the design point. Das and Zheng [11] improve the previous methodology with a cumulative use of sampling points to build RS with square and cross terms. Gayton et al. [22] select the region around the design point from confidence intervals. These ones are estimated with a cross-validation technique with which several design points are calculated. This procedure is repeated until small enough confidence intervals are obtained. Nguyen et al. [35] locate additional sampling points in half-star shaped regions, with respect to the sensitivity of variables on the performance function and the size of the initial ED. RS are built with a double weighted regression, first weights being calculated with respect to the limit state value and second ones depending on distance between sampling points and design point. Kang et al. [26] use moving least square method but the size of additional ED around the design point is empirically chosen.

1.3. Short description of proposed method

Our approach follows the same general scheme but differs at each step. First, based on some classical statistical tools of linear regression, it is proposed to build a sparse RS instead of a complete one. It allows to reduce the number of simulations, and it also goes towards the minimization of approximation error, which is related to the number of terms. For this step, *forward* regression method is used in order to select most important terms based on statistical criteria. Second, quantification of effect of the approximation error on the probability of failure is highlighted. A *bootstrap* method is used in order to estimate variations of RS predictions and an interval of the estimated probability of failure is deduced. From these two points, an adaptive scheme is proposed. The ED is a Latin Hypercube as it is widely used among space filling designs and can be easily enriched. The region, in which the ED is refined, is determined from an importance level on the probability density in the design point. The final criterion used to validate results is the size of the interval around the estimated probability of failure, calculated with the bootstrap method.

So, this paper is organized as follows. Section 2 presents all tools used in the adaptive procedure which is detailed in Section 2.4. First, a general approach of response surface is given and statistical methods are presented to estimate approximation error (Section 2.1). This is followed by the description of response surface construction, where most significant terms are iteratively chosen (Section 2.2). Then, a final validation procedure, based on bootstrap, is introduced (Section 2.3). Section 2.4 describes how these tools are nested to build an adaptive method where experimental design is iteratively refined in a region of interest. The method is illustrated and discussed in Section 3 on one analytical example and two finite element cases.

2. Description of the adaptive procedure

2.1. Some tools related to response surfaces

2.1.1. General description of response surfaces

Let us consider a physical model represented by a mapping $y = \phi(\mathbf{x})$, where $y \in \mathbb{R}$ is a response of interest, $\mathbf{x} = (x_1, \dots, x_M) \in \mathbb{R}^M$ are input variables and ϕ the deterministic function of the model, which is, in our case, the finite element model. The physical

model is the reference one in the sense that it is a description of the physical phenomenon and it is considered to be correct. The meta-model, given by the mapping $y = \hat{\phi}(\mathbf{x}) + \varepsilon$, is then a simplified representation of the physical model, where $\hat{\phi}$ is the approximation function and ε an error. As mentioned before, the words “response surface” are used afterwards for quadratic response surfaces, i.e. functions which read $\hat{\phi}(\mathbf{x}) = \sum_{j=1}^P a_j z_j$, where terms z_j may be linear (x_i), square (x_i^2) or cross terms ($x_i x_k$) of variables \mathbf{x} .

The RS is obtained from a statistical sample of size N , with $N \geq P$, which comes from simulations of the reference model. In a matrix form, the linear regression equation reads

$$\mathbf{y} = \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}, \quad (2)$$

where $\mathbf{y} = (y_1, \dots, y_N)^T$ is the vector of response values in each sampling point, \mathbf{Z} is the matrix of regressors of size $N \times P$, with maximum rank, $\mathbf{a} = (a_1, \dots, a_P)^T$ is the vector of coefficients to determine and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_N)^T$ is the vector of errors. In classical linear regression theory, these terms of error are supposed to be normal random variables with null expectation and constant standard deviation denoted σ (normality assumption). Then, the least square minimization method gives an estimate, say $\hat{\mathbf{a}}$, of coefficients \mathbf{a} from normal equations

$$\hat{\mathbf{a}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y}. \quad (3)$$

As this linear system is generally ill-conditioned, it is often preferred to solve the least square problem with a QR decomposition of matrix \mathbf{Z} [29].

In a simple way, the quality of the meta-model can be measured through the empirical mean of square errors

$$\mathcal{E}_{\text{emp}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\phi}(\mathbf{x}_i))^2. \quad (4)$$

Another more convenient measure is given by the *coefficient of determination*, known as R^2 , widely used in literature and defined by

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{\phi}(\mathbf{x}_i))^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (5)$$

where \bar{y} is the empirical mean of responses: $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$. R^2 coefficient varies between 0 and 1 and represents how the model explains the response: closer it is to 1, better is the model.

Suppose that from a set of potential terms of size P , one tries to find a subset of size ξ which best fits sampling datas. If the selection criterion used is either R^2 or \mathcal{E}_{emp} , the number of terms in the RS will always be maximum, i.e. $\xi = P$. Moreover, if $P=N$, obtained values will be $R^2 = 1$ and $\mathcal{E}_{\text{emp}} = 0$ as the RS will perfectly fit sampling points. The risk involved when the number of terms is large, is that approximation could be very good on learning points but very bad elsewhere. This situation called *overfitting* means that the RS is flexible to be adjusted on learning sample but it is not stable enough, i.e. the RS would be too different if it is built with another learning sample. This can be avoided by minimizing the number of terms ξ in order to obtain a parsimonious RS. In this manner, predictions are generally better on the entire studied domain.

This highlights two main steps of a RS construction. First, the criterion used to estimate the prediction error must be robust and, second, it must be minimized in order to find the best RS from sets of potential terms. The two next sections address these points.

2.1.2. Estimates of prediction error

Prediction error is theoretically defined by the expectation, known as $E[\cdot]$, of the square error. This amounts to consider that input variables \mathbf{x} are realizations of random variables \mathbf{X} , and therefore y is also a realization of a random variable Y . The

prediction error, also known as Mean Square Error (MSE), reads

$$\text{MSE} = E[(\hat{\phi}(\mathbf{X}) - \phi(\mathbf{X}))^2], \quad (6)$$

which can be decomposed as

$$\text{MSE} = (E[\hat{\phi}(\mathbf{X})] - \phi(\mathbf{X}))^2 + V[\hat{\phi}(\mathbf{X})], \quad (7)$$

where $E[\hat{\phi}(\mathbf{X})] - \phi(\mathbf{X})$ is the bias, illustrating precision, and $V[\cdot]$ is the variance, representing stability. As these statistics evolve in an opposite way (larger is the RS, better is the bias but larger is the variance), the selection of a “good” RS involves a bias-variance trade-off [3]. Two main techniques exist to estimate this prediction: Penalty methods based on statistical considerations and simulation methods based on intensive calculation. Some of them are presented in the two following paragraphs. They are the most used in basic statistical literature.

Penalty methods. Mallows’ C_p [31] is a well-known estimate of MSE. If ξ terms, among P , are selected in the RS and if the statistical sample has N points, the sum of squared error is $\text{SSE}(\xi) = \sum_{i=1}^N \varepsilon_i^2$ and the C_p is defined by

$$C_p(\xi) = \frac{\text{SSE}(\xi)}{\hat{\sigma}^2} - N + 2\xi, \quad (8)$$

where $\hat{\sigma}$ is the estimator of the standard deviation σ of residuals ε_i , estimated on the complete RS. The C_p criterion can be viewed as a penalization of SSE by twice the number of terms. It enables to balance the reduction of SSE when the number of terms increases. Comparing models of different sizes, the one which minimizes C_p will be preferred.

Other criteria has been derived from the likelihood function such as *Akaike Information Criterion* (AIC) [1] and *Bayesian Information Criterion* (BIC) [41]. In case of normality assumption, they read

$$\begin{aligned} \text{AIC}(\xi) &= N \log \frac{\text{SSE}(\xi)}{N} + 2\xi \\ \text{BIC}(\xi) &= N \log \frac{\text{SSE}(\xi)}{N} + \xi \log N. \end{aligned} \quad (9)$$

Like C_p , they have to be minimized in order to select the best model from P potential terms.

Finally, a penalized form of the coefficient of determination exists. It is called *adjusted coefficient of determination* and, like previous criteria, depends on the number of terms in the RS. It is defined by

$$R_{\text{adj}}^2(\xi) = 1 - \frac{N-1}{N-\xi-1} (1-R^2(\xi)). \quad (10)$$

The idea still remains to compensate for reduction of SSE with the number of terms in the RS. It must be maximized contrary to C_p , AIC and BIC and varies between 0 and 1 like R^2 .

Simulation methods. Their principle is to test a RS with sampling points not used in the building procedure. The well-known *cross-validation* consists in splitting the initial statistical sample of size N in K groups. Iteratively, for $k = 1, \dots, K$, the k th group is removed from the initial sample, the RS is built with the rest of sample and tested on the left out group, i.e. the k th. The error is thus only calculated on points not used in the building procedure. This error is denoted *cross-validated error* and reads

$$\mathcal{E}_{\text{cv}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\phi}^{(-D_i)}(\mathbf{x}_i))^2, \quad (11)$$

where $\hat{\phi}^{(-D_i)}$ defines the RS built when the subset which contains point i is removed from the initial sample. In analogy with R^2 coefficient, a more convenient measure, often denoted Q^2 and called *predictive coefficient*, can be used. It is defined by

$$Q^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{\phi}^{(-D_i)}(\mathbf{x}_i))^2}{\sum_{i=1}^N (y_i - \bar{y})^2}. \quad (12)$$

The number K is often taken to 10 in literature as it gives good estimate [33]. The limit is $K=N$ which means that, in turn, one point is removed from the initial sample. This case is the *leave-one-out cross-validation*.

Another tool based on simulation, known as *bootstrap* [17], consists in computing several learning sets of size N by sampling with replacement in the initial sample. Bootstrap samples thus may include redundant points of the initial sample, while some others are missing. This resampling technique can be used to estimate mean, standard deviation, confidence interval or even distribution law. About prediction error, several estimates have been proposed in Efron and Tibshirani [19]. No more explanation is given here because, in this paper, the bootstrap will be rather used to estimate variations of RS predictions (cf. Section 2.3.2).

2.1.3. Terms selection

It has been already noticed that the number of terms in RS should be limited in order to make good predictions (to avoid overfitting). As previous section has presented tools which quantify prediction error, the aim of this one concerns selection of terms which minimize it.

This procedure is classical in statistics and consists in testing, one by one, subsets of terms [32]. Since it is not possible to do it with all subsets, several methods have been proposed to add or remove terms, in order to test them, in an iterative manner. The most used ones are known as *forward*, *backward* or *stepwise* selections. Forward selection consists in adding terms one by one in the RS with respect to a given criterion: in turn the term which improves at best the criterion is included. Backward selection starts from a given RS (generally with all terms) and remove terms in turn with respect to the criterion. Finally, stepwise selection uses both techniques and thus can add or remove terms at each step. At the end of the procedure, the RS is often sparse as all terms are not generally useful. The selection criterion can be any of the previously described estimates of prediction error. The choice made in this paper will be discussed in next section.

2.2. Response surface construction

In order to build a predictive RS, a strategy of terms selection and an estimate of prediction error must be chosen. Penalized criteria are convenient for selection since they depend on the number of terms and do not require more computations than one RS construction. However, they are based on normality assumption which is not justified in our applications. Indeed, this assumption holds for physical experiments because repeating same experiments does not give same results. In our case, experiments are deterministic simulations and residuals between predicted responses and true ones only represent the RS lack of fit. Therefore, penalty methods cannot be used to validate a RS in an absolute manner. Conversely, simulation methods do not require any assumption but are more computationally demanding as the RS is built several times. Consequently, they are not convenient for terms selection in RS.

For these reasons, a mixture of all estimates of prediction error will be used: penalty methods for terms selection and simulation methods for RS validation. Among penalized criteria, none is really better than another and it is not possible to know *a priori* which one provides the best RS. Therefore, they will be all used. Finally, selection procedure will provide four potential RS and the best one will be chosen with cross-validation method, as it is generally convenient to discriminate between several models.

About selection procedure, the backward selection has the drawback to start from a given *a priori* RS (generally the complete RS), which needs a high enough statistical sample. With forward

and stepwise selections, a small sample can be first formed, even smaller than the number of potential terms. However, if four penalized criteria are used, stepwise selection needs repeating four times the procedure, whereas only one can be performed with forward selection. Indeed, since criteria depend on SSE and number of terms in the RS, the selection can be performed with respect to SSE in order to build RS of different sizes (with one term, two terms, three terms, etc) and criteria can be calculated on each RS. The four best RS, with respect to the four criteria, are then selected. Consequently, forward selection will be preferred in order to reduce computational time. As said before, the best RS among the four, will be selected with cross-validation, i.e. the RS with the highest predictive coefficient Q^2 will be chosen. This RS construction is summarized in Fig. 1.

2.3. Response surface validation

2.3.1. Motivations

When a reliability analysis is performed with RS method and if no information is available on the domain of failure location, it is natural to sample the input space around mean values. However, failure of mechanical structures is generally a consequence of extreme behavior, which means that probability of failure is often low. Thus, the limit state function is located in tail probabilities. Consequently, even if predictive coefficient is a robust global criterion of the RS quality, it is not representative of its quality close to the limit state. Therefore, a more precise validation criteria seems necessary; it is the objective of present section. Another important point is the difficulty to validate a probability of failure, estimated with RS, when only a measure of the RS quality is available. The validation step described here aims also

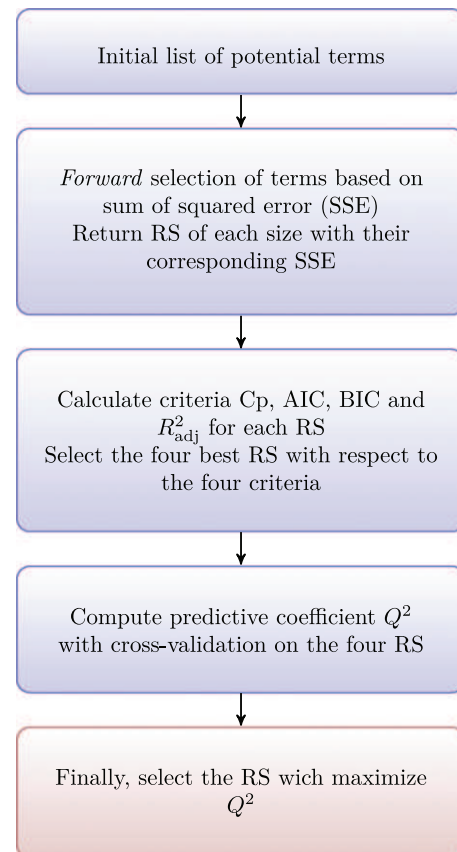


Fig. 1. Procedure of the response surface construction.

at calculating the influence of RS error on the probability of failure. The bootstrap method will be used for this purpose.

Let us recall that bootstrap method [17] is a resampling technique, which consists in building several data samples by sampling with replacement in the original data set. It thus enables to estimate a lot of statistical properties. Here, it will be used in order to estimate variations of RS predictions. Then, two indicators will be deduced and will allow to determine how the probability of failure is affected by RS error.

2.3.2. Bootstrap indicator

Let \mathbf{x}^{ref} be the learning sample of size N and \mathbf{x}^{*b} with $b = 1 \dots B$ the B bootstrap samples of size N obtained by sampling with replacement in \mathbf{x}^{ref} . For each bootstrap sample \mathbf{x}^{*b} , the RS, whose terms have been selected with the procedure described in previous section, is determined and used in order to predict responses on the learning sample. The RS, built with the b -th bootstrap sample \mathbf{x}^{*b} , is denoted $\hat{\phi}^{*b}$ and predictions on \mathbf{x}^{ref} are $\hat{\mathbf{y}}^{*b} = (\hat{y}_1^{*b}, \dots, \hat{y}_N^{*b})$. After B repetitions, variations of the i -th prediction can be observed and, particularly, minimum and maximum values: $\min_{b=1, \dots, B}(\hat{y}_i^{*b})$ and $\max_{b=1, \dots, B}(\hat{y}_i^{*b})$. However, if only one bootstrap sample is badly formed, these minimum and maximum predictions could be very pessimistic. It is thus preferred to introduce the 95% confidence intervals, widely used in statistical literature. An easy and intuitive manner is to sort \hat{y}_i^{*b} , for $b = 1, \dots, B$, in order to obtain the 0.025 B -th and the 0.975 B -th values, which will be denoted $\hat{y}_{i_{\text{low}}}^*$ and $\hat{y}_{i_{\text{up}}}^*$. This confidence interval is known as *percentile bootstrap confidence interval* [18]. From this interval, bootstrap indicators are derived. They correspond to the maximum variations between extreme and nominal predictions

$$\begin{cases} \hat{\varepsilon}_{\text{low}} = \max_{i=1, \dots, N} (|\hat{y}_i - \hat{y}_{i_{\text{low}}}^*|) \\ \hat{\varepsilon}_{\text{up}} = \max_{i=1, \dots, N} (|\hat{y}_i - \hat{y}_{i_{\text{up}}}^*|). \end{cases} \quad (13)$$

These bootstrap-based indicators can be added to the limit state function in order to evaluate their influence on the probability of failure. Let us consider \mathbf{X} the input random vector, $\hat{\phi}$ the RS function and $G(\hat{\phi}(\mathbf{X}))$ the performance function, such as $G(\hat{\phi}(\mathbf{X})) \leq 0$ is the domain of failure. The following probability interval is defined

$$\begin{cases} P_{f_{\text{low}}} = \mathbb{P}(G(\hat{\phi}(\mathbf{X})) - \hat{\varepsilon}_{\text{low}} \leq 0) \\ P_{f_{\text{up}}} = \mathbb{P}(G(\hat{\phi}(\mathbf{X})) + \hat{\varepsilon}_{\text{up}} \leq 0). \end{cases} \quad (14)$$

This interval is the final information to validate the probability of failure estimated with RS. Consequently, it should be low enough if the RS must be accepted. Of course, this interval does not contain, for sure, the true probability of failure but it reflects the quality of the estimated probability. However, since bootstrap-based indicators are determined over the whole sampling space, they could be very constraining if selected points are very far from the limit state area. To avoid this, indicators should be searched in the neighborhood of the design point. The following section aims at defining such a region of interest.

2.3.3. Definition of the region of interest

The region of interest is the area where it is relevant to evaluate bootstrap-based indicators because it mostly contributes to the probability of failure. Therefore, if this region is denoted S , bootstrap-based indicators defined in (13) become

$$\begin{cases} \varepsilon_{\text{low}} = \max_{i \in S} (|\hat{y}_i - \hat{y}_{i_{\text{low}}}^*|) \\ \varepsilon_{\text{up}} = \max_{i \in S} (|\hat{y}_i - \hat{y}_{i_{\text{up}}}^*|), \end{cases} \quad (15)$$

which means that sampling points are searched inside S . This region has been determined in relation with the design point neighborhood as defined by Dutfoy and Lebrun [16]. Let us recall that design point is the nearest point to the origin in the normal standard space of random variables (\mathbf{u} -space) which belongs to the limit state. The change from physical space to standard normal space can be done with an isoprobabilistic transformation [30]. The design point is obtained through an optimization procedure which can be written

$$\mathbf{u}^* = \operatorname{argmin} \|\mathbf{u}\|^2 \quad \text{with} \quad H(\mathbf{u}) = 0, \quad (16)$$

where \mathbf{u}^* is the design point, $\|\cdot\|$ the euclidian norm and H the performance function in \mathbf{u} -space, such as $H(\mathbf{u}) = 0$ is the limit state. The distance between the design point and the \mathbf{u} -space origin is the reliability index, known as β . It is generally used in the First Order Reliability Method (FORM) in order to approximate the probability of failure. The design point probability density is $\varphi_n(\mathbf{u}^*)$ where φ_n is the multivariate standard normal distribution. An importance level, denoted ε_{pdf} , is defined such as all points \mathbf{u} with $\varphi_n(\mathbf{u}) \leq \varepsilon_{\text{pdf}} \varphi_n(\mathbf{u}^*)$ are considered to have a negligible probability density. The function ψ is defined such as $\psi(\|\mathbf{u}\|) = \varphi_n(\mathbf{u})$ and the previous inequality is equivalent to $\psi(\|\mathbf{u}\|) \leq \varepsilon_{\text{pdf}} \psi(\beta)$. Introducing the factor $\delta_{\varepsilon_{\text{pdf}}}$ such that $\psi(\beta(1 + \delta_{\varepsilon_{\text{pdf}}})) = \varepsilon_{\text{pdf}} \psi(\beta)$, we have

$$\delta_{\varepsilon_{\text{pdf}}} = \sqrt{1 - \frac{2\ln(\varepsilon_{\text{pdf}})}{\beta^2}} - 1, \quad (17)$$

Then, for a given value of importance level ε_{pdf} , the region of interest S is defined as the hypercube $\prod_{i=1}^M [b_{i_{\text{low}}}, b_{i_{\text{up}}}]$ which framed the intersection between the ball $\mathcal{B}(\mathbf{0}, \beta(1 + \delta_{\varepsilon_{\text{pdf}}}))$ and the limit state $H(\mathbf{u}) = 0$ (see an example on Fig. 2).

In practice, hypercube boundaries are determined by optimization. More precisely, for all component u_i , with $i = 1, \dots, M$, the problem to solve reads

$$b_{i_{\text{low}}} = \operatorname{argmin} u_i \quad \text{with} \quad \begin{cases} \|\mathbf{u}\| - \beta(1 + \delta_{\varepsilon_{\text{pdf}}}) = 0 \\ H(\mathbf{u}) = 0 \end{cases} \quad (18)$$

$b_{i_{\text{up}}}$ is obtained exactly in the same way, except that u_i must be maximized. Finally, the importance level ε_{pdf} is the only parameter which controls the size of the region of interest. It will be discussed further.

The region of interest is used to calculate bootstrap-based indicators in a more relevant manner. However, it is necessary that this region contains sampling points. If it is not the case, additional sampling points should be included inside in order to compute the probability interval. If this interval can be calculated but is too high to validate the RS, this latter has to be refined in the neighborhood of the design point. In both cases, ED must be enriched in the region of interest and a new RS must be evaluated.

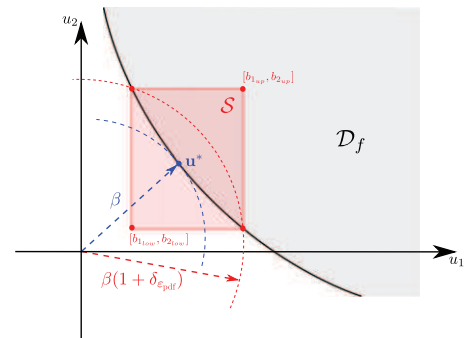


Fig. 2. Definition of the region of interest.

This refinement is described in the next section as an adaptive scheme.

2.4. Adaptive scheme

2.4.1. Short description

The principle of the proposed method is to iteratively enrich the ED in the region of interest determined thanks to the RS. So, at each step, sampling points are added to ED, the RS is updated and the region of interest is thus modified. Consequently, procedure is divided in three main stages:

1. The region of interest must be first located, i.e. it does not have to move too much from one iteration to another. It is called *region convergence*.
2. Once the region is found, the RS must be globally good enough so that a probability of failure can be estimated. It is the *response surface convergence*.
3. This estimated probability of failure must be validated with the bootstrap method, whose methodology is described in section 2.3. It is the *validation*.

The following sections describe the enrichment method and all steps of the adaptive scheme. But, before all, the choice of an ED is explained.

2.4.2. Choice of experimental design

In deterministic computer experiments, space filling design are often preferred compared to classical design of experiments used in physical experiments [40,45,44]. This kind of experimental designs tries to fill the design space uniformly. A widely used experimental design, in computer experiments, is Latin Hypercube Sampling (LHS). It is very easy and cheap to generate and

can be used with a lot of input variables. Also, as it will be seen in the next section, it can be easily enriched, keeping quasi-LHS properties. That is why it has been chosen here.

In order to build a RS with the statistical sample formed by LHS, the number of sampling points must be $N \geq \zeta$ where ζ is the number of terms in the RS. However, the use of bootstrap method involves a larger number of sampling points in order to avoid conditioning problems. Indeed, bootstrap samples are generated by sampling with replacement, which means that some points are redundant whereas other points do not appear. Empirical studies made on some examples showed that a number of sampling points $N = 3\zeta$ is a good trade-off between accuracy and computational time. If the number is lower, RS predictions based on bootstrap samples could be deteriorated; if the number is higher, computational time is uselessly larger.

The number of terms ζ is not known before choosing the number of points in the ED, since terms are selected with the iterative method described in Section 2.2. Therefore, it is necessary to set an ED of size N and to restrict the number of terms ζ in the RS construction such as $N \geq 3\zeta$. In order to reduce computational time, it is better to start the adaptive procedure with a quite small ED. This one will be enriched at each step if necessary. So, the initial size of ED has been arbitrary fixed to $N_{\text{init}} = 3M$, where M is the number of input variables, because it enables to potentially include all linear terms in the RS.

2.4.3. Sequential latin hypercube sampling

At each iteration of our adaptive scheme, the LHS initially generated is enriched in the region of interest. This enrichment is described now and is inspired from [46,7]. It consists in adding sampling points to the initial LHS so that it keeps LHS structure (or quasi-LHS as mentioned by Blatman and Sudret [7]). Since ED is only increased in the region of interest \mathcal{S} , which is an

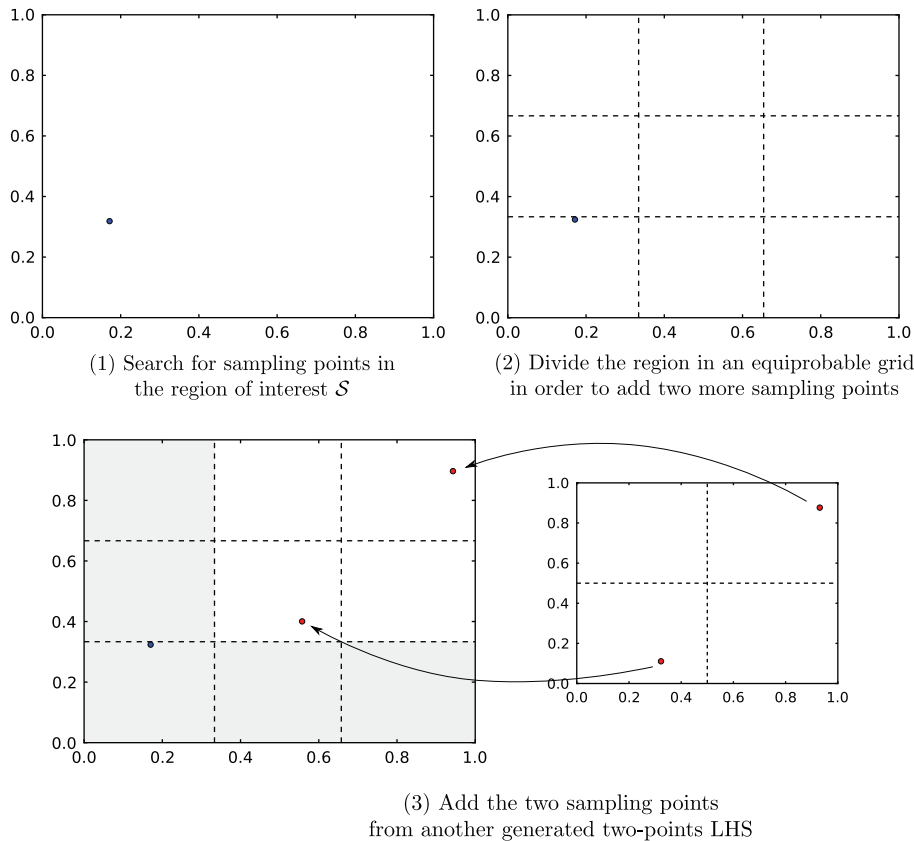


Fig. 3. Method to increase LHS - Ideal case.

hypercube, this section considers that design space is restricted to a unit hypercube $\prod_1^M [0,1]$, where M is the number of variables. The method is explained in the two dimensional case, illustrated in Fig. 3, but can be generalized to a M dimensional problem.

Assume that one point already exists in the design space and that two more points must be added. As the resultant number of points is three, the design space is split into a 3×3 equiprobable grid. Intervals (row and column), represented by the existing point, are shaded in Fig. 3. It means that no additional point can fall inside. Conversely, blank spaces point to potential areas for additional points. These new ones are generated with an independent LHS and mapped to the real design space with respect to the cell position. For example, as shown in Fig. 3, the point in cell corresponding to the first row and the first column, is set at the intersection of the first blank row and the first blank column, if one counts from the origin. Therefore, the new ED still depicts a LHS.

However, this previous case is ideal in the sense that no existing point is redundant in the intervals when the design space is split. A second example is illustrated in Fig. 4, where two points fall in the same interval (second row). Normally, only two additional points should be generated but, here, the red row would not be represented. Therefore, three points are generated instead of two. The two first ones are placed as presented above. The third one, say \mathcal{A} in Fig. 3, enables to represent the red row. But, since all columns are already filled, the cell along this row is randomly chosen. Although this method involves more additional points than the expected number, it enables to get a quasi-LHS scheme and thus better covers design space. In this example, only one additional point was needed but more could be necessary. In that case, the procedure automatically adjust the number of additional points in order to represent each interval.

2.4.4. Stages description of adaptive method

Initial step. A first LHS is generated with $N_{\text{init}} = 3M$ sampling points, where M is the number of input variables. A RS is determined where the number of allowed terms is such as $\xi_0 \leq M$.

Region convergence. This stage consists in locating the region of interest. Assume that a RS is built at step i . The reliability index, say β_i , and the region of interest, say \mathcal{S}_i , are determined on the RS (the region is calculated as explained in Section 2.3.3). For a given threshold, denoted ε_β , the criterion which specifies if the region has changed compared to the previous step, reads

$$\left| \frac{\beta_i - \beta_{i-1}}{\beta_i} \right| \leq \varepsilon_\beta. \quad (19)$$

If it is satisfied, the region of interest is located and can be fixed. The converged region is denoted $\mathcal{S}_{\text{conv}}$ and is not updated anymore. One can go to the *response surface convergence* stage. If criterion (19) is not satisfied, the region \mathcal{S}_i is used in order to enrich the ED at step $i+1$, following the procedure given in

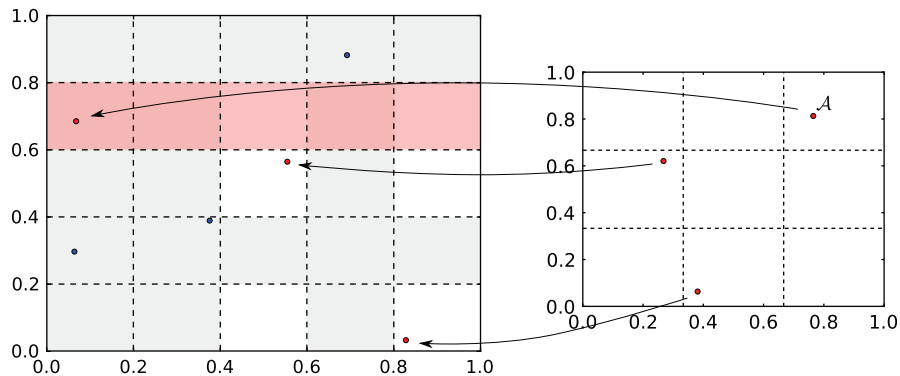


Fig. 4. Method to augment LHS - Case of redundant points.

Section 2.4.3, and a new RS is built. At each step, all sampling points from the initial ED are kept for the RS construction. Since the number of allowed terms in RS depends on ED size ($3\xi_i \leq N_i$), the RS can grow up iteratively. Steps of *region convergence* are summarized in Fig. 5. In this one, the number of additional points at each iteration is denoted N_{add} . In following applications, it will be arbitrarily fixed to the number of input variables, but it must be recalled that more sampling points can be added in order to keep a quasi-LHS structure (see Section 2.4.3).

Response surface convergence. Now, the region of interest, known as $\mathcal{S}_{\text{conv}}$, is located and the RS quality is of main interest. The chosen convergence criterion is based on the predictive coefficient Q^2 (cf. Section 2.1.2). Note that cross-validation does not involve more computation since it has already been performed in RS construction procedure (cf. Section 2.2). At the i -th iteration, Q_i^2 is compared to a threshold, say ε_{rs} , according to

$$Q_i^2 \geq \varepsilon_{\text{rs}}. \quad (20)$$

If it is satisfied, iterative procedure is stopped and *validation* step takes place. Otherwise, the RS must be improved by adding more sampling points in ED, inside region $\mathcal{S}_{\text{conv}}$.

Suppose now a case where limit state is far from the mean point. Sampling points are scattered over a large area because the region of interest has changed several times before convergence. Then, sampling points out of $\mathcal{S}_{\text{conv}}$ could prevent the RS quality to reach the required value ε_{rs} . To avoid it, a second RS is built on a local ED, i.e. which only contains sampling points inside $\mathcal{S}_{\text{conv}}$.

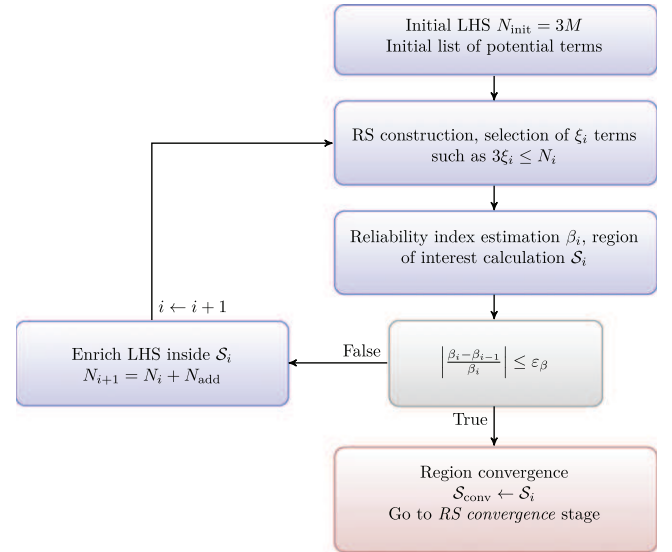


Fig. 5. Flowchart of region convergence steps.

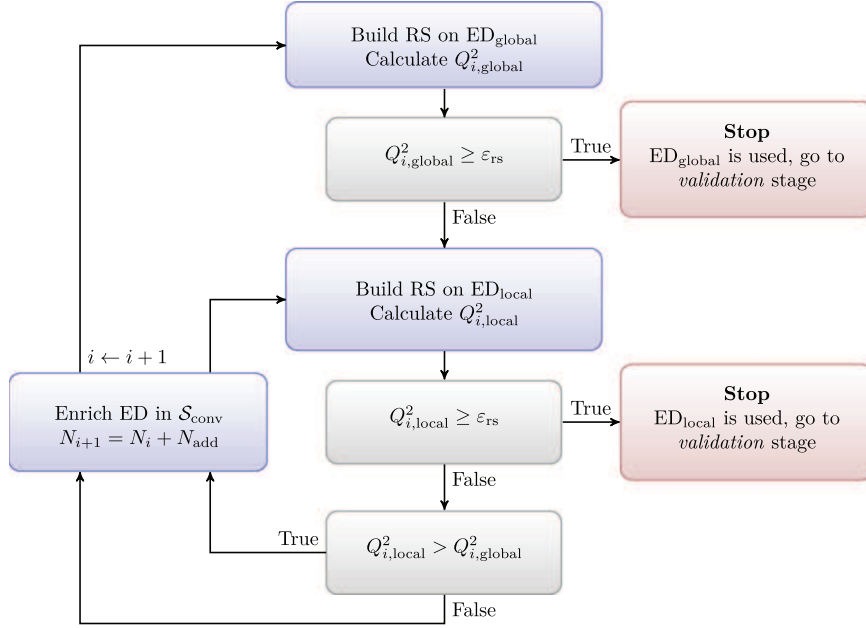


Fig. 6. Flowchart of response surface convergence steps.

This ED is denoted ED_{local} and the predictive coefficient estimated on its RS is Q^2_{local} . Conversely, ED_{global} is used to speak about the whole ED, i.e. with all sampling points, and Q^2_{global} is its associated predictive coefficient. If $Q^2_{local} \geq Q^2_{global}$, ED_{local} is used until the end of the procedure. It means sampling points out of S_{conv} are not used anymore to build RS. In such a case, criterion (20) is applied to Q^2_{local} . This “test of deletion” is applied at each iteration until convergence. Main steps of *response surface convergence* are summarized in Fig. 6.

Validation. This last step is based on Section 2.3. The bootstrap is performed on RS with the selected ED (global or local) and indicators ε_{low} and ε_{up} are determined. They are used to calculate $P_{f_{low}}$ and $P_{f_{up}}$ (cf. (14)) on RS with a simulation method (crude Monte Carlo or Importance Sampling). This probability interval is the final validation criterion. If it is not low enough, the algorithm can be restarted with a higher value of ε_{rs} , in order to obtain better quality RS. All steps of adaptive scheme are summarized in Fig. 7.

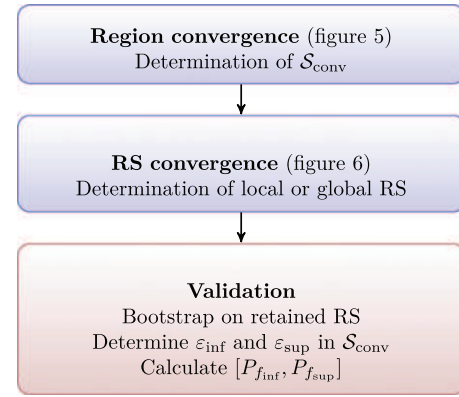


Fig. 7. Flowchart of adaptive method.

2.4.5. Discussion on parameters

Main parameters, which control the algorithm, are: Importance level ε_{pdf} , connected to the region of interest size, ε_{β} governing the region of interest location and ε_{rs} , for the RS quality. Validation step ensures that ε_{rs} is high enough. Indeed, if $[P_{f_{low}}, P_{f_{up}}]$ is too large, ε_{rs} can be adjusted and the algorithm is restarted so that RS quality is improved. ε_{pdf} characterizes the region of interest size: lower it is, bigger is the region of interest. Following its definition, the chosen value should be relatively low, i.e. 0.05 or 0.1. Actually, this importance level defines the region above which the probability density is considered as negligible. The interval $[P_{f_{low}}, P_{f_{up}}]$ is impacted by this parameter as the number of sampling points, used to determine ε_{low} and ε_{up} , may be different with respect to the region size. If the region is small (high ε_{pdf}), few confidence is given to $[P_{f_{low}}, P_{f_{up}}]$. ε_{β} is intermediate and must not be neither too high nor too low. If it is too high, the region can converge quickly towards a bad location. Conversely, a very low value could involve more computational time. However, additional iterations, which are done when ε_{β} is too low, should involve a faster convergence of RS quality in RS convergence step. Examples of classical values are presented in next section.

Another parameter, which does not clearly appear, is the spread of the initial ED. In literature, it is always arbitrarily chosen as a function of standard deviation of random variables, i.e. $\pm k\sigma$, where k is often taken equal to 3. Intuitively, one can imagine that different values of k may change the adaptive method progress during iterations. If k is low, a small region is initialized and moves at each step. If k is high, a larger region is initialized and then refined. The choice of k is difficult to rigorously justify but some numerical experiments, done by authors, have shown that, on common cases, the number of simulation runs and accuracy of estimated results are very close whatever the value of k is between 1 and 5.

3. Numerical applications

3.1. Analytical example

3.1.1. Description

First, a two dimensional analytical case is used in order to illustrate the adaptive scheme and its progress. This example has been already used by Kang et al. [26] and Nguyen et al. [35]. The two input random variables, say X_1 and X_2 , are assumed to follow

standard normal distributions. The performance function reads

$$G(\mathbf{x}) = \exp(0.4(x_1 + 2) + 6.2) - \exp(0.3x_2 + 5) - 200, \quad (21)$$

and the domain of failure is such as $G(\mathbf{x}) \leq 0$. The reference value of the probability of failure, say P_f^{REF} , is obtained with FORM followed by Importance Sampling with 500,000 simulations. Corresponding generalized reliability index is given by $\beta^{\text{REF}} = -\Phi^{-1}(P_f^{\text{REF}})$, where Φ is the normal cumulative distribution function, and its reference value is $\beta^{\text{REF}} = 2.686$.

3.1.2. Analysis and results

The previously described adaptive method is performed on this analytical function in order to estimate the probability of failure. The initial ED contains 6 sampling points and the number of additional points is set to 2 (the number of input variables as mentioned in Section 2.4.4). The importance level is chosen to $\varepsilon_{\text{pdf}} = 0.05$, the convergence criterion on reliability index is $\varepsilon_{\beta} = 0.01$ and the convergence criterion on RS quality is $\varepsilon_{\text{rs}} = 0.99$. The probability of

failure is calculated on the RS with FORM followed by Importance Sampling with 500,000 simulations. The generalized reliability index β_{gen} is then deduced. Results are presented in Table 1 and Fig. 8 shows the progress of the method at each iteration. In Table 1, $\beta_{\text{gen}_{\text{low}}}$ and $\beta_{\text{gen}_{\text{up}}}$ stand for generalized reliability indexes, which correspond to $P_{f_{\text{low}}}$ and $P_{f_{\text{up}}}$, calculated with bootstrap method during the validation step (cf. Section 2.4.4).

The algorithm converges in 4 iterations and the number of estimates of the original performance function, say N_{calc} , is 15. It can be observed that sequential LHS procedure has added more than 2 sampling points during steps 2 and 3. Estimated generalized reliability index β_{gen} is very close to the reference value and the interval given by $\beta_{\text{gen}_{\text{low}}}$ and $\beta_{\text{gen}_{\text{up}}}$ is small enough. The predictive coefficient of the final RS is $Q^2 = 0.998$, i.e. much larger than the required value $\varepsilon_{\text{rs}} = 0.99$. Actually, when the region of interest has converged, RS quality was already good enough to converge immediately, i.e. without adding more sampling points in ED. Finally, note that, in this example, Q^2 coefficient has been evaluated from leave-one-out cross-validation since first ED only contains 6 sampling points.

Table 1

Results of analytical example.

β^{REF}	β_{gen}	$\beta_{\text{gen}_{\text{low}}}$	$\beta_{\text{gen}_{\text{up}}}$	Q^2	N_{calc}
2.686	2.677 (0.3%) ^a	2.593 (3.1%) ^b	2.717 (1.5%) ^b	0.998	15

^a Relative gap with β^{REF} .

^b Relative gap with β_{gen} .

3.2. Truss structure

3.2.1. Description

This example is taken from [7] and consists in a truss structure with 23 bar elements. The structure is shown Fig. 9. There are 10 input random variables: Young's moduli and cross sections, denoted

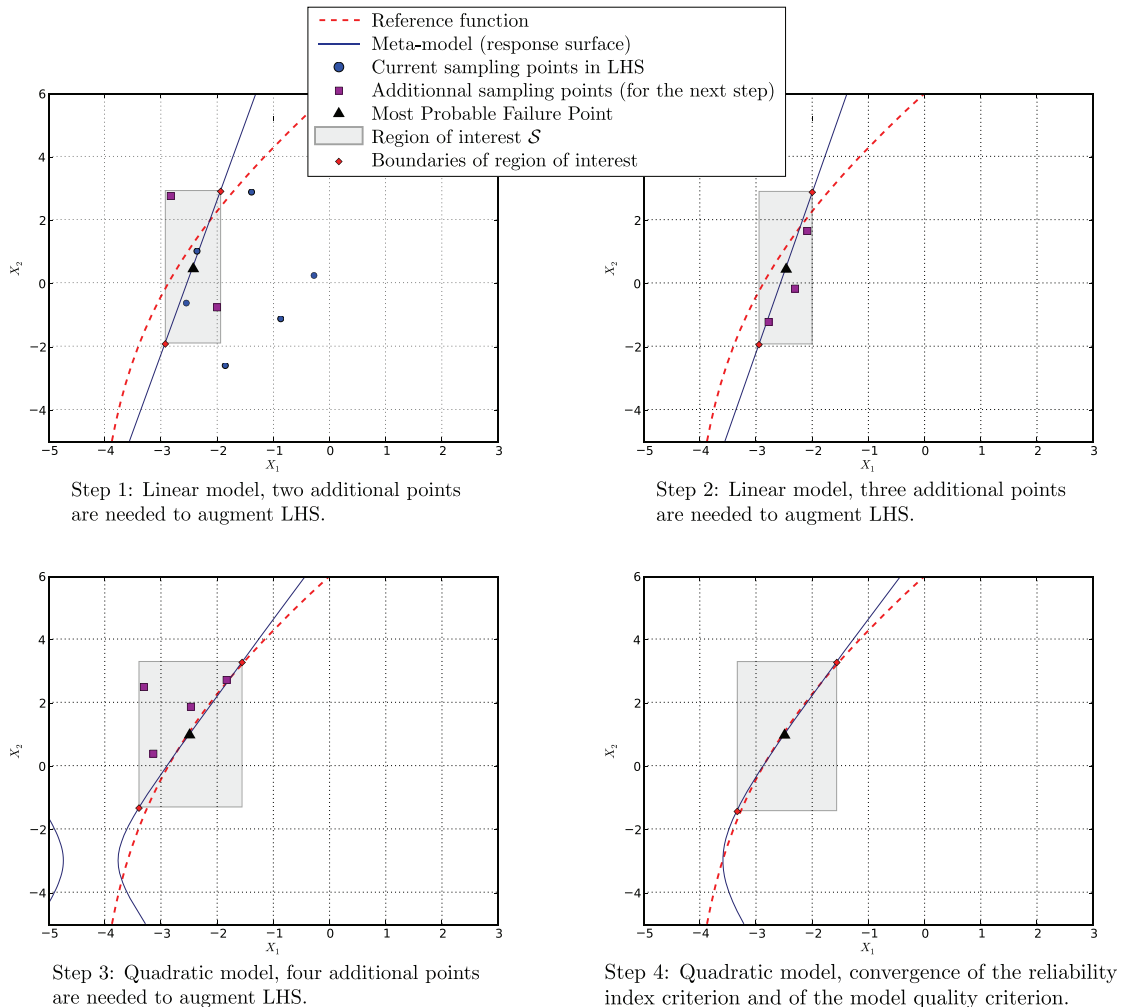


Fig. 8. Evolution of the method on an analytical example.

A_1, E_1 for horizontal bars and A_2, E_2 for sloping bars, and applied loads from P_1 to P_6 . All these variables are assumed to be independent. Their distributions, mean values and standard deviations are given in Table 2. Response of interest is the deflection of the midspan, say \mathbf{v} .

3.2.2. Analysis and results

The structure is studied for reliability analysis with the following performance function

$$G(\mathbf{x}) = 0.14 - |\mathbf{v}(\mathbf{x})|, \quad (22)$$

where \mathbf{x} are realizations of \mathbf{X} , which are standard random variables obtained with an isoprobabilistic transformation of physical random variables. The domain of failure is defined such as $G(\mathbf{x}) \leq 0$. Reference value of the probability of failure, say P_f^{REF} , has been estimated by Blatman and Sudret [7] with FORM followed by Importance Sampling with 500,000 simulations. The corresponding generalized reliability index is $\beta^{\text{REF}} = 3.98$.

The adaptive method has been used in order to estimate the probability of failure. The initial LHS contains 30 sampling points and the number of additional points at each step is 10, since 10 variables are considered here. The importance level is $\varepsilon_{\text{pdf}} = 0.05$, criterion on region convergence is $\varepsilon_{\beta} = 0.01$ and criterion which controls RS quality is set to $\varepsilon_{\text{rs}} = 0.99$. The probability of failure is still obtained by Importance Sampling with 500,000 simulations, and results are provided with respect to generalized reliability indexes. They are presented in Table 3. It can be observed that the global procedure takes place as follows:

- First, 62 computations of the finite element model have been performed to reach the targeted RS quality, i.e. $\varepsilon_{\text{rs}} = 0.99$. The

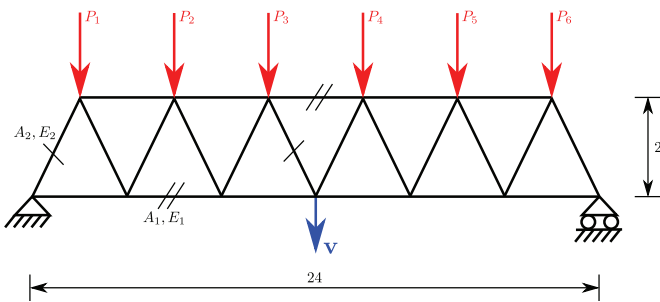


Fig. 9. Truss structure (unit m).

Table 2
Properties of random variables—Truss structure.

Variable	Distribution	Mean	Standard deviation
E_1, E_2 (Pa)	Lognormal	2.10×10^{11}	2.10×10^{10}
A_1 (m ²)	Lognormal	2.0×10^{-3}	2.0×10^{-4}
A_2 (m ²)	Lognormal	1.0×10^{-3}	1.0×10^{-4}
$P_1 - P_6$ (N)	Gumbel	5.0×10^4	7.5×10^3

predictive coefficient obtained is $Q^2 = 0.992$. But the interval, estimated around β_{gen} with bootstrap-based indicator, is quite large. Note that no sampling point has been removed, i.e. final RS is built with ED_{global}.

- The algorithm is restarted from the already performed ED, with an increasing value of ε_{rs} , namely $\varepsilon_{\text{rs}} = 0.995$, in order to improve RS quality. Only 16 additional simulations were necessary to reach the criterion and ED_{global} is still conserved.
- Finally, the procedure is launched a last time with $\varepsilon_{\text{rs}} = 0.999$. The predictive coefficient obtained for the final RS is $Q^2 = 0.9994$ and 64 additional simulations have been performed. Here, sampling points out of region of interest have been removed to reach the required criterion, i.e. final RS is built on ED_{local}. Therefore, 142 computations have been performed from the first step but 103 sampling points are used to build the RS. Interval obtained around β_{gen} is considered as sufficiently small.

3.3. Frame structure

3.3.1. Description

Now is considered the frame structure presented in Fig. 10. It was already studied in several papers such as [7,47,35]. There are 21 studied variables, namely 3 horizontal loads (P_1 to P_3), 8 moments of inertia and cross-sections corresponding to different mechanical properties (from A_1, I_1 to A_8, I_8) and 2 Young's moduli (E_1 and E_2). These properties, associated to beam elements, are detailed in Table 4. Distributions, mean values and standard deviations of variables are given in Table 5. Here, some variables are assumed to be correlated:

- Cross section and moments of inertia of the same element with a coefficient $\rho_{A_i, I_i} = 0.95$;
- All others geometrical properties with coefficient $\rho_{A_i, A_j} = \rho_{I_i, I_j} = \rho_{A_i, I_j} = 0.13$;

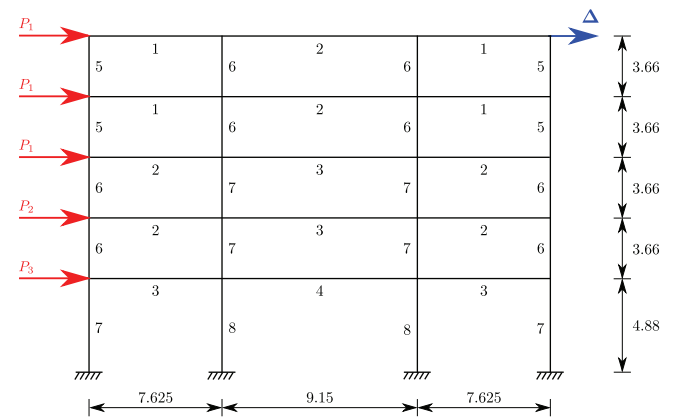


Fig. 10. Frame structure (unit m).

Table 3
Results for truss structure.

β^{REF}	Target Q^2	Obtained Q^2	β_{gen}	$\beta_{\text{gen}_{\text{low}}}$	$\beta_{\text{gen}_{\text{up}}}$	N_{calc}
3.98	0.990	0.9920	3.80 (4.7%) ^a	3.37 (11.3%) ^b	4.06 (6.8%) ^b	62
	0.995	0.9970	3.90 (2%) ^a	3.67 (5.8%) ^b	4.09 (4.9%) ^b	+ 16 = 78
	0.999	0.9994	3.99 (0.3%) ^a	3.87 (3%) ^b	4.09 (2.4%) ^b	+ 64 = 142 ^c

^a Relative gap with β^{REF} .

^b Relative gap with β_{gen} .

^c Only 103 sampling points (ED_{local}) are used in the RS construction.

Table 4
Finite element properties—Frame structure.

Elements	Young's modulus	Moment of Inertia	Cross section
1	E_1	I_5	A_5
2	E_1	I_6	A_6
3	E_1	I_7	A_7
4	E_1	I_8	A_8
5	E_2	I_1	A_1
6	E_2	I_2	A_2
7	E_2	I_3	A_3
8	E_2	I_4	A_4

Table 5
Properties of random variables—Frame structure.

Variable	Distribution	Mean	Standard deviation
P_1 (kN)	Lognormal	133.454	40.04
P_2 (kN)	Lognormal	88.97	35.59
P_3 (kN)	Lognormal	71.175	28.47
E_1 (kN/m ²)	Normal	2.1738×10^7	1.9152×10^6
E_2 (kN/m ²)	Normal	2.3796×10^7	1.9152×10^6
I_1 (m ⁴)	Normal	8.1344×10^{-3}	1.0834×10^{-3}
I_2 (m ⁴)	Normal	1.1509×10^{-2}	1.2980×10^{-3}
I_3 (m ⁴)	Normal	2.1375×10^{-2}	2.5961×10^{-3}
I_4 (m ⁴)	Normal	2.5961×10^{-2}	3.0288×10^{-3}
I_5 (m ⁴)	Normal	1.0812×10^{-2}	2.5961×10^{-3}
I_6 (m ⁴)	Normal	1.4105×10^{-2}	3.4615×10^{-3}
I_7 (m ⁴)	Normal	2.3279×10^{-2}	5.6249×10^{-3}
I_8 (m ⁴)	Normal	2.5961×10^{-2}	6.4902×10^{-3}
A_1 (m ⁴)	Normal	3.1256×10^{-1}	5.5815×10^{-2}
A_2 (m ⁴)	Normal	3.7210×10^{-1}	7.4420×10^{-2}
A_3 (m ⁴)	Normal	5.0606×10^{-1}	9.3025×10^{-2}
A_4 (m ⁴)	Normal	5.5815×10^{-1}	1.1163×10^{-1}
A_5 (m ⁴)	Normal	2.5302×10^{-1}	9.3025×10^{-2}
A_6 (m ⁴)	Normal	2.9117×10^{-1}	1.0232×10^{-1}
A_7 (m ⁴)	Normal	3.7303×10^{-1}	1.2093×10^{-1}
A_8 (m ⁴)	Normal	4.1860×10^{-1}	1.9537×10^{-1}

- Correlation of Young's modulus is equal to $\rho_{E_1, E_2} = 0.9$;
- All remaining variables have no correlation.

Response of interest is the horizontal top displacement, say Δ .

3.3.2. Analysis and results

The structure is studied for reliability analysis with the following performance function

$$G(\mathbf{x}) = 0.06 - \Delta(\mathbf{x}), \quad (23)$$

where \mathbf{x} are realizations of \mathbf{X} , which are independent standard random variables obtained with an isoprobabilistic transformation of physical random variables. The domain of failure is defined as $G(\mathbf{x}) \leq 0$. As in previous example, reference value of the generalized reliability index has been estimated by Blatman and Sudret [7] with FORM followed by Importance Sampling with 500,000 simulations, and is $\beta^{\text{REF}} = 3.51$.

This example shows somehow a limit of the proposed adaptive method. Actually, if we use the same parameters than in other examples, i.e. $\varepsilon_{\text{pdf}} = 0.05$, $\varepsilon_{\beta} = 0.01$ and $\varepsilon_{\text{rs}} = 0.99$, the procedure does not converge. The reason is that, with high uncertainties (large values of standard deviation) and high correlated input random variables, the region of interest, defined with $\varepsilon_{\text{pdf}} = 0.05$, is very large. Moreover, the response behaviour is too non-linear so that it cannot be well approximated with a quadratic RS. Consequently, the required RS quality is never reached. However, in [35], authors obtain good results with a complete quadratic RS and with 259 numerical simulations. Actually, it has been observed that

Table 6
Results of frame example.

β^{REF}	β_{gen}	$\beta_{\text{gen}_{\text{low}}}$	$\beta_{\text{gen}_{\text{up}}}$	Q^2	N_{calc}
3.51	3.63 (3.4%) ^a	3.49 (3.9%) ^b	3.75 (3.3%) ^b	0.991	149

^a Relative gap with β^{REF} .

^b Relative gap with β_{gen} .

parameters, used by these authors for their own adaptive ED, are equivalent to choose, in our case, $\varepsilon_{\text{pdf}} = 0.6$. Consequently, if our method is applied with this value, the procedure converges in 149 finite element simulations with a good accuracy. Corresponding results are presented in Table 6.

Although this example is particular with such a level of uncertainty, it shows the difficulty involved by the region of interest. Basically, the algorithm does not converge with classical importance level (small values of ε_{pdf}), but may converge with specific values (high values). It is because the response is highly non-linear over a large region, whereas it is more smooth on a small one. In [5], this frame structure has been treated with a LHS around the mean point, and a sparse polynomial chaos of maximum degree 6 was necessary to obtain accurate enough results on the performance function (23). In [35], an accurate result is obtained on the same function with a complete quadratic RS but with a refinement strategy. These two methods are quite different, but it shows this frame structure can be difficult to handle, strongly depending on which size of design space is considered. From a practical point of view, what occurred here (non convergence with small ε_{pdf} and convergence with high ε_{pdf}) can be useful to reveal a difficulty, particularly a limitation in the use of sparse quadratic RS. In such a case, it should be better to use other kinds of meta-models or methods in order to compare results.

4. Conclusion

4.1. Summary

This paper introduces a method based on response surface (RS) for reliability analysis. It consists in an adaptive scheme where the RS is iteratively refined in the region which mainly contribute to the probability of failure, i.e. close to the design point. The methodology, used to build the RS, is a selective procedure of the most important terms with respect to statistical criteria. As several RS are determined, cross-validation technique is used to choose the best one. The region of interest, wherein the RS must be refined, is determined from an importance level on the probability density at the design point. In this manner, it is sized with respect to the result of interest, i.e. the probability of failure. In order to reduce the number of simulations, the algorithm starts from a small initial experimental design and proceeds in three steps:

1. The experimental design is enriched inside the region of interest and the RS is updated (so the region of interest also), while the design point is not precisely determined (*region convergence*);
2. The experimental design is enriched inside the region of interest, fixed at the end of previous step, while RS quality, determined with cross-validation, is not good enough;
3. Finally, bootstrap method is used to calculate variations of RS predictions and leads to the interval $[P_{f_{\text{low}}}, P_{f_{\text{up}}}]$, containing the estimated probability of failure.

This latter validation criterion has the advantage to be very convenient for reliability analysis. The initial experimental design is a Latin Hypercube Sampling (LHS) and enrichment in the region of interest is done in order to keep a quasi-LHS structure.

The algorithm is tested on three examples. The first one is a two dimensional analytical case, which allows to study the method progress. The second example shows how the method can be applied on a real finite element model. Final validation step enables to justify the required value for the RS quality criterion. The method is restarted twice, to reduce the interval around the estimated probability of failure given by bootstrap-based indicators. Provided results are fairly satisfactory in term of precision and number of simulations. The last case is also a finite element model and shows the difficulty involved by the region of interest size. Actually, it is argued this difficulty is a means of detecting problematic cases, where it is absolutely needed to complete the analysis with an other method in order to increase confidence in estimated results.

4.2. Discussion

The method, proposed here, is intended to be applied to industrial problems. Of course, these applications must correspond to main hypothesis involved by the method: the response of interest must not be highly non-linear to be well approximated with a quadratic RS, and the design point must be unique. Such hypothesis have been made because they correspond to most of cases in application considered by authors (and a lot of other ones), which deals with spacecraft structures. Indeed, in this domain, uncertainties, taken into account, are relatively low and the response behavior is often quite smooth. Generally, computational time of a finite element simulation is not very high, but sufficiently to make impracticable direct simulation methods. The number of variables can be moderately large (several tens or hundreds), but large enough to make more complex kind of meta-models difficult to use. Consequently, RS appears to be a good trade-off as a first step towards reliability analysis application in this domain. However, most of methods, based on RS, are still expensive even with a moderately large number of variables, and criteria, used for validation of estimated results, are not really convenient.

The method is focused on several main points. First, reduction of computational time is addressed by using sparse RS. Although in examples presented above, the number of variables is not a real problem, it is not the case in industrial applications. Indeed, even with a moderately large number of variables, a complete RS would need a high number of simulations, as the number of interaction terms is $M(M-1)/2$, if M is the number of variables. The terms selection procedure combined with an adaptive strategy, to refine the experimental design around the design point, really reduce computational time. In the recent contribution of [7], selection of terms to build a sparse polynomial chaos has also shown a large reduction of simulations.

A second point deals with the area used to refine the experimental design. In most of existing methods, this region is arbitrarily chosen and often becomes smaller and smaller during iterations. The region, used in this paper, is defined with respect to the probability density at the design point. Since the wished result is the probability of failure, this choice appears as more suitable in order to have a better confidence in results, estimated with RS.

Finally, the validation step is a key point. The probability interval $[P_{f_{low}}, P_{f_{up}}]$, obtained by bootstrap-based indicators, is an effective measure to know if the approximation is accurate enough for reliability analysis, more than a RS error estimate. This method of validation goes in same way that recent contributions

to validate adaptive procedure based on kriging [15] or support vector machine [12].

Acknowledgments

This work is supported by Centre National d'Etudes Spatiales (CNES) and Thales Alenia Space (TAS). The authors thank Fabrice Buffe (CNES), Jérôme Buffe and Anne Carpine (TAS).

References

- [1] Akaike, H., 1973. Information theory and an extension of the likelihood ratio principle. In: 2nd International Symposium on Information Theory, Tsahkadsor, Armenian. pp. 267–81.
- [2] Au SK, Beck JL. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics* 2001;16(4):263–77.
- [3] Azaïs, J., Bardet, J., 2005. Le modèle linéaire par l'exemple: régression, analyse de la variance et plans d'expériences illustrés avec R, SAS et Splus. Dunod.
- [4] Bichon B, Eldred M, Swile L, Mahadevan S, McFarland J. Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA journal* 2008;46(10):2459–68.
- [5] Blatman, G., 2009. Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis. Ph.D. thesis, IFMA/Université Blaise Pascal.
- [6] Blatman G, Sudret B. Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *Comptes Rendus Mécanique* 2008;336(6):518–23.
- [7] Blatman G, Sudret B. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics* 2010;25(2):183–97.
- [8] Bourinet J, Deheeger F, Lemaire M. Assessing small failure probabilities by combined subset simulation and support vector machines. *Structural Safety* September 2011;33(6):343–53.
- [9] Bucher C. Asymptotic sampling for high-dimensional reliability analysis. *Probabilistic Engineering Mechanics* 2009;24(4):504–10.
- [10] Bucher CG, Bourgund U. A fast and efficient response surface approach for structural reliability problems. *Structural Safety* 1990;7(1):57–66.
- [11] Das PK, Zheng Y. Cumulative formation of response surface and its use in reliability analysis. *Probabilistic Engineering Mechanics* 2000;15(4):309–15.
- [12] Deheeger, F., Lemaire, M., 2007. Support vector machine for efficient subset simulations: Åsmart method. In: 10th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP10).
- [13] Deng J, Gu D, Li X, Yue ZQ. Structural reliability analysis for implicit performance functions using artificial neural network. *Structural Safety* January 2005;27(1):25–48.
- [14] Dubourg, V., Deheeger, F., Sudret, B., 2011a. Metamodel-based importance sampling for structural reliability analysis. Arxiv preprint arXiv:1105.0562.
- [15] Dubourg V, Sudret B, Bourinet J. Reliability-based design optimization using kriging surrogates and subset simulation. *Structural and Multidisciplinary Optimization* 2011:1–18.
- [16] Dutfoy, A., Lebrun, R., 2007. Le test du maximum fort : une façon efficace de valider la qualité d'un point de conception. In: Congrès Français de Mécanique.
- [17] Efron B. Bootstrap methods: another look at the jackknife. *The annals of statistics* 1979;7(1):1–26.
- [18] Efron B, Tibshirani R. An introduction to the bootstrap. Chapman Hall; 1997.
- [19] Efron B, Tibshirani R. Improvements on cross-validation: The 632 + bootstrap method. *Journal of the American Statistical Association* 1997;92(438):548–60.
- [20] Elhewy AH, Mesbahi E, Pu Y. Reliability analysis of structures using neural network method. *Probabilistic Engineering Mechanics* 2006;21(1):44–53.
- [21] Faravelli L. Response-surface approach for reliability analysis. *Journal of Engineering Mechanics* 1989;115(12):2763–81.
- [22] Gayton N, Bourinet JM, Lemaire M. CQ2RS: a new statistical approach to the response surface method for reliability analysis. *Structural Safety* 2003;25(1):99–121.
- [23] Ghanem RG, Spanos PD. Stochastic finite elements: a spectral approach. Dover Pubns; 2003.
- [24] Hasofer A, Lind N. Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division* 1974;100:111–21.
- [25] Hurtado J. An examination of methods for approximating implicit limit state functions from the viewpoint of statistical learning theory. *Structural Safety* 2004;26(July 3):271–93.
- [26] Kang SC, Koh HM, Choo JF. An efficient response surface method using moving least squares approximation for structural reliability analysis. *Probabilistic Engineering Mechanics* 2010;25(4):365–71.
- [27] Kaymaz I. Application of kriging method to structural reliability problems. *Structural Safety* 2005;27(April 2):133–51.
- [28] Kim SH, Na SW. Response surface method using vector projected sampling points. *Structural Safety* 1997;19(1):3–19.

- [29] Lascaux P, Théodor R. 1986. Analyse numérique matricielle appliquée à l'art de l'ingénieur, Tomes 1 et 2. Masson.
- [30] Lemaire M. 2006. Reliability of Structures. ISTE.
- [31] Mallows CL. Some comments on Cp. *Technometrics* 2000;42(1):87–94.
- [32] Miller, A.J., 2002. Subset selection in regression. CRC Pr I Llc.
- [33] Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005;21(15):3301–7.
- [34] Naess A, Leira B, Batsevych O. System reliability analysis by enhanced monte carlo simulation. *Structural Safety* 2009;31(September (5)):349–55.
- [35] Nguyen XS, Sellier A, Duprat F, Pons G. Adaptive response surface method based on a double weighted regression technique. *Probabilistic Engineering Mechanics* 2009;24(2):135–43.
- [36] Papadrakakis M, Lagaros ND. Reliability-based structural optimization using neural networks and monte carlo simulation. *Computer methods in applied mechanics and engineering* 2002;191(32):3491–507.
- [37] Picheny V, Ginsbourger D, Roustant O, Haftka R, Kim NH. Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design* 2010;132(7).
- [38] Pradlwarter HJ, Pellissetti MF, Schenk CA, Schuëller GI, Kreis A, Fransen S, Calvi A, Klein M. Realistic and efficient reliability estimation for aerospace structures. *Computer Methods in Applied Mechanics and Engineering* 2005;194(12–16):1597–617.
- [39] Rajashekhar MR, Ellingwood BR. A new look at the response surface approach for reliability analysis. *Structural Safety* 1993;12(3):205–20.
- [40] Sacks J, Welch W, Mitchell T, Wynn H. Design and analysis of computer experiments. *Statistical science* 1989;4(4):409–23.
- [41] Schwarz G. Estimating the dimension of a model. *The annals of statistics* 1978;6(2):461–4.
- [42] Sichani M, Nielsen S, Bucher C. Applications of asymptotic sampling on high dimensional structural dynamic problems. *Structural Safety* 2011;33(July (4–5)):305–16.
- [43] Sichani M, Nielsen S, Bucher C. Efficient estimation of first passage probability of high-dimensional nonlinear systems. *Probabilistic Engineering Mechanics* 2011;26(October (4)):539–49.
- [44] Simpson TW, Lin DKJ, Chen W. Sampling strategies for computer experiments: Design and analysis. *International Journal of Reliability and Applications* 2001;2(3):209–40.
- [45] Simpson TW, Poplinski JD, Koch PN, Allen JK. Metamodels for Computer-based Engineering Design: Survey and recommendations. *Engineering with Computers* 2001;17(2):129–50.
- [46] Wang GG. Adaptive response surface method using inherited latin hypercube design points. *Journal of Mechanical Design* 2003;125(2):210–20.
- [47] Wei D, Rahman S. Structural reliability analysis by univariate decomposition and numerical integration. *Probabilistic Engineering Mechanics* 2007;22(1): 27–38.