14th International Conference on Structural Safety and Reliability - ICOSSAR'25
1-6 June, 2025 - Los Angeles California, USA
R. Ghanem, E. Johnson, S. Masri and A. Olivier (Eds)

ACTIVE LEARNING KRIGING METHODS FOR BAYESIAN UPDATING

IOANNIS A. PRENTZAS AND MICHALIS M. FRAGIADAKIS

Laboratory for Earthquake Engineering, National Technical University of Athens,
Iroon Polythechniou 9, Zografou 15772 Athens, Greece
{gprentzas, mfrag}@mail.ntua.gr and https://www.civil.ntua.gr

Key words: Bayesian updating, Active learning, Kriging, Reliability Analysis, Subset Simulation, Inverse Problems

Abstract. The paper studies the efficiency and the accuracy of active learning Kriging methods for Bayesian updating. Classical Bayesian updating is formulated as a reliability analysis problem using the Bayesian Updating with Structural reliability (BUS) methods, enabling active learning Kriging models to estimate the posterior distributions. The resulting structural reliability problem is handled with the aid of the quantified active learning Kriging (qAK) algorithm that was recently proposed by the authors. The accuracy of the BUS method depends strongly on the constant c, whose calculation using a surrogate Kriging model is a challenging problem. For this purpose, two approaches for calculating the constant c are compared: in the first approach, it is calculated during the active learning process, and in the second, c is calculated prior to the analysis, solving a global optimization problem. When the adaptive approach is chosen, emphasis is placed on strategies that leverage computational geometry algorithms for computing the convex hull of a set of sample points. In case of a non-adaptive scheme, the calculation of c is achieved by solving a global optimization problem, where an acquisition function guides the Design of Experiments of the Kriging model. The adopted methods are evaluated with the aid of a benchmark problem, demonstrating the advantages and disadvantages of each method. The work highlights the synergies between active learning and Bayesian inference, providing valuable insights for future research and practical applications in structural engineering.

1 INTRODUCTION

In structural dynamics, accurately predicting the behavior of a system under dynamic loading is important for ensuring safety and performance. However, uncertainties in material properties, boundary conditions, geometric configurations, and external excitation often lead to discrepancies between the actual dynamic response and the response predicted by analytical or numerical models. Bayesian updating provides a systematic and probabilistic framework to address these uncertainties by incorporating measured data, such as accelerations, displacements, or modal properties, into the existing model. Thus, the model parameters are updated using observed data and Bayesian methods, which enhance the fidelity of structural dynamic models, enabling more reliable response predictions, improved structural health monitoring, and better-informed decision-making in engineering design and maintenance.

Sampling-based approaches for approximating the posterior distribution have gained significant attention, primarily because they do not require prior assumptions about its form, e.g., assuming a

Gaussian distribution. One of the simplest methods, rejection sampling [1], generates samples from the posterior by treating the likelihood function as a filter. However, in high-dimensional settings or when multiple observations are involved, this method often suffers from a very low acceptance rate, leading to inefficient sampling performance. To address these limitations, Markov Chain Monte Carlo (MCMC) methods have been widely explored for Bayesian updating, offering the ability to draw samples directly from the posterior distribution. To improve convergence, Beck and Au [2] introduced the adaptive Metropolis–Hastings (AMH) algorithm, which employs a sequence of intermediate probability density functions to progressively approximate the target posterior. Building upon this, Ching and Chen [3] developed the Transitional MCMC (TMCMC) method, replacing kernel density estimation with a resampling strategy to further enhance the sampling efficiency. Betz *et. al* [4] propose an improved version of the TMCMC method, which reduces the bias in the estimate of the evidence and improves the convergence behavior of posterior estimates.

Recently, the Bayesian Updating with Structural reliability methods (BUS) [5] algorithm was introduced. BUS links Bayesian updating with reliability analysis and thus allows to incorporate of methods like the Subset Simulation (SuS) method and the Importance Sampling (IS) method to achieve enhanced efficiency. In BUS, Bayesian updating is reformulated into a reliability analysis problem by employing a limit-state function, which incorporates a uniformly distributed auxiliary random variable p and a constant c. An input sample that belongs to the failure domain follows the posterior distribution. Furthermore, Betz *et al.* [6] proposed a modified version of BUS in which the constant c is determined adaptively, while a maximum likelihood-free variant of the original BUS was introduced by Betz *et al.* [7]. While these variants provide different ways to estimate c, they involve considerable computational cost. In this respect, there is a need to develop a more efficient approach for determining c and thus create a robust and effective computational framework for BUS, especially when c is not easy to determine.

Surrogate model-based reliability analysis has become increasingly popular, since it efficiently replaces the original limit-state function with a surrogate model and considerably reduces the number of function evaluations. Surrogate models can be combined with BUS to solve the resulting reliability problem. In Bayesian model updating, Kriging has been integrated into the TMCMC [8] method to reduce the computational cost by replacing the full system simulations in likelihood evaluations. In the BUS framework, ANN-based [9] surrogate models of the likelihood function were used to reduce the number of simulations. Recently, Active learning Kriging methods [10, 11, 12, 13] have been applied in the framework of BUS by calculating the constant c in various ways and adaptively selecting training samples.

In this paper, two different strategies for computing the constant c within the BUS framework are considered. In the first approach, c is updated adaptively at every level of the Subset Simulation method. Specifically, the constant c is computed as the maximum surrogate likelihood over the set $\mathbf{X}^0 \cap \mathcal{C}$, where $\mathbf{X}^0$ represents the initial sample set and C is the convex hull of the initial Design of Experiments (DoE). The convex hull of the DoE is used as a filtering criterion to ensure that surrogate predictions are reliable. The value of c is then computed as the maximum surrogate likelihood within the intersection of the subset samples and the convex hull of the training data. This approach helps to maintain both accuracy and stability during the sampling process. The second method treats the computation of c as a global optimization problem. The goal is to directly estimate the maximum of the surrogate log-likelihood function over the entire input space. More specifically, in order to find the maximum, we use the Kriging model and the *Expected Improvement (EI)* function for selecting new support points. Once the c constant is known, the standard BUS approach is subsequently performed.

2 THEORETICAL BACKGROUND

The proposed approach is based on the use of the Kriging metamodel [14] for solving the Bayesian updating problem [2] within the framework of BUS [5]. A brief introduction to these algorithms is provided in the lines that follow.

2.1 Bayesian Model Updating

Bayesian model updating provides a probabilistic framework for incorporating new information, such as experimental data or system observations, into an existing model. The goal is to reduce uncertainty in the model parameters by updating their probabilistic description in light of observed evidence. Let $\mathbf{x} \in \mathbb{R}^n$ be the vector of uncertain model parameters, and $\boldsymbol{y} \in \mathbb{R}^m$ represent the set of observed data. The Bayesian updating of the parameter distribution is expressed through Bayes' theorem as follows:

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \frac{L(\boldsymbol{y}|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{c_E} \tag{1}$$

where, $f_{\mathbf{X}}(\mathbf{x})$ is the prior distribution representing prior knowledge of $\mathbf{x}$, $L(\boldsymbol{y}|\mathbf{x})$ is the likelihood function expressing the probability of observing $\mathbf{y}$ given $\mathbf{x}$ and $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\boldsymbol{y})$ is the posterior distribution, updated after incorporating data. Finally, the denominator c_E is known as the model evidence, or the normalization constant, and is equal to:

$$c_E = \int L(\boldsymbol{y}|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{2}$$

In practice, the likelihood function is usually based on the assumption of Gaussian measurement errors. For every observation y_i, the individual likelihood component is given by:

$$L_i(\mathbf{x}|y_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i - g_i(\mathbf{x}))^2}{2\sigma_i^2}\right) \tag{3}$$

where $g_i(\mathbf{x})$ is the model prediction for observation y_i, and σ_i^2 is the variance of the associated measurement error. Assuming independence of observations, the overall likelihood is obtained as:

$$L(\mathbf{x}|\boldsymbol{y}) = \prod_{i=1}^{m} L_i(\mathbf{x}|y_i) \tag{4}$$

Once the likelihood is known, Eq. 1 provides the posterior distribution $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\boldsymbol{y})$. Sampling methods (e.g., Markov Chain Monte Carlo methods) can be employed to explore the posterior distribution for any application desired, i.e., model updating, uncertainty quantification, and prediction.

2.2 The BUS method

The BUS method [5] recasts the Bayesian model updating problem as an equivalent structural reliability problem. The goal is to evaluate the posterior distribution of uncertain model parameters $\mathbf{x}$, given prior information and new observational data $\boldsymbol{y}$. The key idea in BUS is to introduce an auxiliary random variable $p \sim \mathcal{U}(0, 1)$ and define the augmented space $(\mathbf{x}, p)$. The posterior sampling is reformulated as a reliability problem with the following limit-state function:

$$G(\mathbf{x}, p) = p - c \cdot L(\mathbf{x}|\boldsymbol{y}) \tag{5}$$

The failure domain Ω_F is defined as:

$$\Omega_F = \{(\mathbf{x}, p) \mid G(\mathbf{x}, p) \leq 0\} = \{(\mathbf{x}, p) \mid p \leq c \cdot L(\mathbf{x}|\mathbf{y})\} \tag{6}$$

where, c is a scaling constant such that $c \cdot L(\mathbf{x}|\mathbf{y}) \leq 1$ for all $\mathbf{x}$. The probability of failure in this reliability context is the probability of drawing a sample from the failure domain:

$$P_f = \mathbb{P}[G(\mathbf{x}, p) \leq 0] \tag{7}$$

The estimated failure probability P_f associated with Ω_F is thus equal to the acceptance rate of the posterior samples. Since, in general small probabilities are sought, the Subset Simulation (SuS) method can be used within the BUS framework to efficiently generate such samples, as described in Algorithm 1. The samples $\mathbf{x}$ associated with the accepted pairs $(\mathbf{x}, p) \in \Omega_F$ at the final SuS level correspond to posterior realizations from $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\boldsymbol{y})$, allowing for robust Bayesian model updating.

The major challenge in the formulation of BUS is the efficient selection of the multiplier c of Eq. 5 so that the inequality $c \cdot L(\mathbf{x}|\mathbf{y}) \leq 1$ is satisfied. Some researchers prefer to examine the mathematical structure of the likelihood function [5], while others propose an adaptive calculation method [6, 15]. However, whatever approach is chosen, difficulties can arise when a surrogate model within the framework of the BUS is used. Alternative methods for calculating c and thus overcoming this problem are here discussed.

Algorithm 1 Bayesian Updating using SuS [5, 16]

Require: Number of samples per subset K, probability of intermediate subsets p_0, BUS constant c
Ensure: Samples $(\mathbf{x}_k^i, p_k^i)_{k=1,\ldots,K}$, failure probability p_f

Initialization:
Draw K samples $(\mathbf{x}_k^0, p_k^0)$ from the prior distribution of $(\mathbf{X}, P)$
Set $i \leftarrow 0$ and $p_f \leftarrow 1$
while $h_i \geq 0$ **do**
    $i \leftarrow i + 1$
    Sort $(\mathbf{x}_k^{i-1}, p_k^{i-1})_{k=1,\ldots,K}$ with respect to limit-state function values $G_i(\mathbf{x}_k^{i-1}, p_k^{i-1})$
    Set G_i as the p_0-percentile of the set $\left(G_i(\mathbf{x}_k^{i-1}, p_k^{i-1})\right)_{k=1,\ldots,K}$
    if $G_i \leq 0$ **then**
        Set $i \leftarrow 0$
        $p_i \leftarrow \frac{n}{K}$, where $n = \sum_{k=1}^{K} \mathbf{1}\left[G_i(\mathbf{x}_k^{i-1}, p_k^{i-1}) \leq \max(G_i, 0)\right]$
    else
        $p_i \leftarrow p_0$
    Generate the samples $(\mathbf{x}_k^i, p_k^i)_{k=1,\ldots,K}$ by using MCMC; n Markov chains of length K/n are generated from the seeds $(\mathbf{x}_k^{i-1}, p_k^{i-1})$
Set $r = i$ and $p_f = \prod_{i=0}^{r} p_j$

2.3 Basic Kriging definitions

Let us assume a Design of Experiments, $\mathbf{X}^* = [\mathbf{x}_1^*, \ldots, \mathbf{x}_{N_0}^*]^T$,where N_0 is the initial number of DoE, with $\mathbf{x}_i^* \in \mathbb{R}^n$ $(i = 1, 2, \ldots, N_0)$. The true response function for sample i, can be approximated as follows:

$$g(\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{f}(\mathbf{x}) + Z(\mathbf{x}), \tag{8}$$

where $\boldsymbol{\beta}^T \mathbf{f}(\mathbf{x})$ is the trend of a Gaussian Process (GP), and $\beta^T = [\beta_1, \ldots, \beta_n]^T$ is the regression coefficient of the basis function $\mathbf{f}(\mathbf{x}) = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)]^T$ and n is the number of basis functions. The term $\sigma_{\widehat{g}}^2$ is the variance of $Z(\mathbf{x})$, which denotes a zero-mean stationary Gaussian Process that relates to a covariance matrix:

$$Cov(\mathbf{x}, \mathbf{x}') = \sigma_g^2 R(\mathbf{x} - \mathbf{x}', \boldsymbol{\vartheta}). \tag{9}$$

where $R(\mathbf{x} - \mathbf{x}^{'}, \boldsymbol{\vartheta})$ is the correlation function between points $\mathbf{x}$ and $\mathbf{x}^{'}$ and $\boldsymbol{\vartheta}$ is the vector with the parameters of the correlation function. A widely used autocorrelation function is the *anisotropic squared exponential model*:

$$R(\mathbf{x} - \mathbf{x}', \boldsymbol{\vartheta}) = \exp\left[-\sum_{k=1}^{n} \left(\frac{x_k - x_k'}{\vartheta_k} \right)^2 \right] \tag{10}$$

According to the principle of Kriging, $g(\mathbf{x})$ can be approximated with the PDF of a Gaussian distribution:

$$\widehat{g}(\mathbf{x}) \sim \mathcal{N}(\mu_{\widehat{g}(\mathbf{x})}, \sigma_{\widehat{g}(\mathbf{x})}). \tag{11}$$

3 BUS-BASED ACTIVE LEARNING KRIGING MODEL

As the Bayesian updating problem has been transformed to a reliability problem, the *quantified active learning Kriging-based (qAK)* methodology [17] is employed to solve it. The *qAK method*, recently proposed by the authors, has two variations, a Monte Carlo-based (qAK-MCS) and another based on the Subset Simulation (qAK-SuS) method. Though more complicated in terms of implementation, the latter has computational advantages since the candidate sample for the active learning is smaller and more informative.

The qAK-SuS method is here adopted, focusing on constructing a Kriging surrogate model that is adaptively refined throughout the Subset Simulation procedure. An initial DoE is defined as $\mathbf{X}^* = \left[\mathbf{x}_1^*, \ldots, \mathbf{x}_{N_0}^*\right]^T$, where N_0 is the initial number of DoE. The log-likelihood $\log(L(\mathbf{x}))$ in Eq. 5 can be approximated with the aid of a Kriging model using $\mathbf{X}^*$. Expressing Equation 5 in logarithmic form and incorporating the surrogate model, the limit-state function is reformulated as:

$$\widehat{G}(\mathbf{x}, p) = \log(p) - \log(c) - \log(\widehat{g}(\mathbf{x})) \tag{12}$$

Here, the original likelihood function is replaced by the Kriging surrogate model $\widehat{g}(\mathbf{x})$, which is constructed following an active learning strategy. As it already discussed to define the limit-state function, it is important to evaluate the constant c, which requires knowledge of the maximum value of the likelihood function.

At the first MCS step ($i = 0$) of the SuS method, the limit-state function is replaced by the surrogate limit-state function $\widehat{G}$. As a result, K samples $\{\mathbf{x}_k^0, p_k^0\}, k = 1, \ldots K$, are produced, together with the associated log-likelihood values $\widehat{g}(\mathbf{x}_k^0)$. Note that the superscript i denotes the level/step of the SuS method. At the i-th level of the SuS method, the surrogate limit-state function $\widehat{G}$ is employed

to determine the threshold value $\widehat{G}_i$ (Algorithm 1). Then, a total of $p_0 \times K$ Markov chain seeds, $\left(\mathbf{x}_k^{i-1}, p_k^{i-1}\right)$ are selected sampling from the surrogate limit-state function $\widehat{g}$. These seeds are subsequently used as candidate samples for identifying new training points and refining the Kriging model of the next SuS level.

We propose enriching the DoE at level $i-1$ of the Kriging model by incorporating new points selected via the quantified active learning method [17]. A new surrogate model is then trained using the updated DoE $\mathbf{X}^{*,i-1}$. An adaptive conditional sampling procedure [18] is executed, employing the surrogate limit-state function to create samples $(\mathbf{x}_k^i, p_k^i)$ for the current subset. The computation of the intermediate threshold value $\widehat{G}_i$ at every subset level depends on the method used to estimate the constant c.

When the active learning strategy is adopted, the *misclassification* function [19] is employed to identify new training points. This function enhances the classification accuracy near the limit-state surface where the surrogate limit-state function is close to zero. Within the BUS framework, points with non-positive limit-state function values correspond to samples from the posterior distribution. The *misclassification function* has been designed to identify regions where predictions that are close to the limit-state surface and/or where the Kriging model exhibits high variance. For fixed c and p values, the limit-state function is normally distributed as:

$$\widehat{G}(\mathbf{x}, p) \sim \mathcal{N}(\log(p) - \log(c) - \mu_{\widehat{g}}(\mathbf{x}), \sigma_{\widehat{g}}^2(\mathbf{x})) \tag{13}$$

where $\mu_{\widehat{g}}(\mathbf{x})$ and $\sigma_{\widehat{g}}(\mathbf{x})$ are the mean value and the standard deviation of the surrogate of the log-likelihood function. Accordingly, the U-function becomes:

$$U(\mathbf{x}, p) = \frac{\left|\log(p) - \log(c) - \mu_{\widehat{g}}(\mathbf{x})\right|}{\sigma_{\widehat{g}}(\mathbf{x})} \tag{14}$$

The misclassification probability for a point $(\mathbf{x}, p)$ is the probability that the predicted sign of the limit-state function is incorrect and is given by:

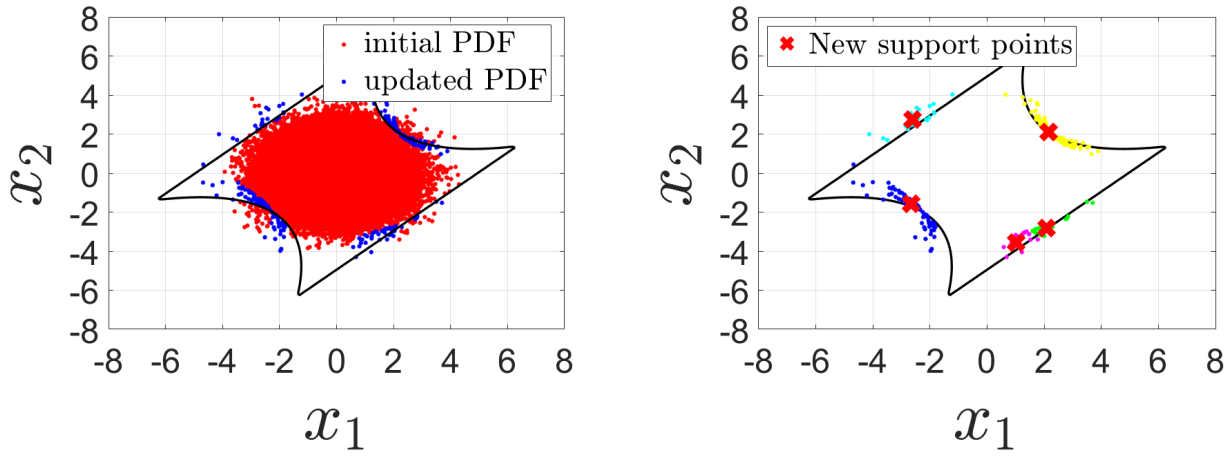
$$P_m(\mathbf{x}, p) = \Phi(-U(\mathbf{x}, p)) \tag{15}$$

This probability lies in the range $[0, 0.5]$. Values near zero indicate and/or low prediction variance while values approaching 0.5 suggest proximity to the limit-state surface and/or high prediction variance. The selection of new support points requires the calculation of the *misclassification* function $U(\mathbf{x})$. For sample $\mathbf{x}_k^i$, the PDF from which the candidate sample is created is updated as follows:

$$\widehat{f}_{\mathbf{X}}(\mathbf{x}_k^i) = f_{\mathbf{X}}(\mathbf{x}_k^i) \cdot H(w(\mathbf{x}_k^i) - w_{lim}) \tag{16}$$

where H is the Heaviside function, w are weights and w_{lim} is a threshold, usually set equal to 0.1. According to Eq. 16, the updated PDF is formed by accepting points whose weights are greater than w_{lim}. The weights are equal to the *misclassification* function, i.e., $w(\mathbf{x}_k^i) = U(\mathbf{x}_k^i, p)$. The new PDF $\widehat{f}_{\mathbf{X}}$ receives large values close to $G(\mathbf{x}_k^i) = 0$ and decays away from G. Therefore, Eq. 16 shifts the PDF to the areas that mostly affect the failure probability, P_f.

Multiple new points are subsequently selected using the K-means [20, 21] clustering algorithm. The misclassification probabilities serve as weights for the clustering, and the points x_k^i with weights $w(x_k^i) > w_{\text{lim}}$ are discarded. It is important to emphasize that the clustering is performed on the input space of the Kriging model (i.e., the training input space), rather than on the space of limit-state function evaluations. Following this approach, the most promising candidate samples from the Markov chain seeds of each subset are then selected to update the metamodel population $\mathbf{X}^{*,i}$.



(a) (b)

Figure 1: Example of a four-branch limit-state function: (a) Samples from the initial and the updated PDF. (b) Data clustering and selection of new support points (new points are shown with a red x). [17]

Figure 1a shows the candidate sample from both the initial and the updated PDF. More specifically, in Figure 1a, the red points were sampled according to the initial distribution of each random variable, while the blue points were obtained with the updated PDF. The blue points, which are close to the limit-state surface, are fewer than the red points since points with weights less than w_{lim} have been rejected. Thus, it can be seen that points with large weight values are near the limit-state surface, while points with weight less than w_{lim} were discarded. As a stopping criterion of the active learning procedure, the probabilistic-based criterion previously were proposed by the authors [17, 22] was adopted here.

4 CALCULATION OF THE BUS CONSTANT "c"

4.1 Definition of the constant

An essential step is to estimate the constant c, which controls both the accuracy and the computational efficiency of the BUS method. From Section 2.2, it is evident that the smallest admissible value of c^{-1}, denoted c_{min}^{-1} is found as:

$$c_{\text{min}}^{-1} = \max_{\mathbf{x}\in\mathbf{X}} L(\mathbf{x}) \tag{17}$$

In general, c is not known a priori, hence choosing a suitable likelihood multiplier that satisfies the inequality $c \cdot L(\mathbf{x}) \leq 1$ for all $\mathbf{x}$ is numerically challenging. Selecting a value larger than c_{min}^{-1} ensures an unbiased posterior distribution, but it reduces the efficiency of the sample generation process. On the other hand, choosing a value less than c_{min}^{-1} introduces bias in the posterior samples. Hence, an appropriate choice of the scaling constant c is very important, as it affects the definition of the failure event Ω_F of Eq. 6. Two strategies for calculating the constant c in the framework of active learning Kriging were examined here: (i) the *adaptive* method that modifies the limit-state function during SuS, (ii) the *non-adaptive* method, where c is first calculated and then the SuS method is applied to solve the resulting reliability problem. Both approaches have computational merits and are accurate.

4.2 The adaptive method

The constant is calculated at every SuS subset, also updating and improving the Kriging model. The constant c is calculated following the same procedure as in adaptive BUS [6] with some modifications that account for the replacement of the original limit-state function with a surrogate model. More specifically, the Quickhull method [23] is adopted for computing the convex hull of a finite set of points in order to define the sample from which the constant will be calculated. This idea was first proposed by Giovanis *et al.* [9] and Angelikopoulos *et al.* [8]. Moreover, Rossat *et. al* [13] combined the Quickhull algorithm with polynomial chaos Kriging within the framework of BUS for solving inverse problems, while the original active learning method [19] was used. In this paper, a more efficient active learning algorithm [17] is used, while the calculation of c is further investigated.

The initial training points $\mathbf{X}^{*,0}$ are created using LHS sampling, while the MCS sample is produced by sampling from a standard normal distribution. The MCS samples can be located far from the DoE $\mathbf{X}^{*,0}$, which may result in inaccurate predictions of the surrogate model. This approach may either lead to an underestimation of the optimal value , or to a substantial overestimation, which can adversely affect the efficiency of the SuS procedure. To mitigate the issue of incorrect calculation of the constant c, we define the initial value of c based on a subset $\mathbf{x}_k^i$, $k = 1, \ldots, K$, that includes only the points lying close to $\mathbf{X}^{*,i}$. Specifically, we filter the samples using the DoE convex hull. The convex hull, denoted as $\text{Conv}(\mathbf{X}^{*,i})$, represents the smallest convex set that contains all the points and can be viewed as the representative domain of the trained surrogate model. Note that this approach requires that the DoE is sufficiently dense. The constant c is thus computed as:

$$c^{-1} = \max_{\mathbf{x} \in \mathcal{K}^{(0)}} \exp\left(\widehat{g}(\mathbf{x})\right) \tag{18}$$

where $\mathcal{K}^i = \mathbf{x}_k^i \cap \text{Conv}(\mathbf{X}^{*,i})$, $k = 1, \ldots K$.

At the SuS level i, the surrogate limit-state function $\widehat{G}$ is used to determine the threshold value $\widehat{G}_i$ (see Algorithm 1). A new set of $p_0 \times K$ Markov chain seeds $(\mathbf{x}_k^{i-1}, p_k^{i-1})$ is selected based on the surrogate limit-state function. To improve the accuracy of the surrogate, the DoE $\mathbf{X}^{*,i-1}$ is enriched by adding multiple new samples according to the *qAK* strategy described in Section 3. A new Kriging model is then trained on the updated DoE $\mathbf{X}^{*,i}$.

The surrogate limit-state function is then used to perform the adaptive conditional sampling procedure [18], which generates the next set of samples. At the end of every SuS level, the constant c is updated as:

$$c^* = \max\left(c, c_i\right), \quad \text{where} \quad c_i^{-1} = \max_{\mathbf{x} \in \mathcal{K}^i} \exp\left(\widehat{g}(\mathbf{x})\right) \tag{19}$$

and $\mathcal{K}^i = \text{Conv}(\mathbf{X}^{*,i}) \cap \left\{{}^{(i)}\mathbf{X}\right\}$. Following the recalculation of c, the threshold is adjusted to maintain consistency in the subset definition.

$$\widehat{G}_i^* = \widehat{G}_i + \log\left(\frac{c}{c^*}\right) \tag{20}$$

The SuS method is terminated when $\widehat{G}_i > 0$.

4.3 The non-adaptive method

In the non-adaptive method, the value of the scaling constant c can be determined before the construction of the Kriging-based limit-state function $\widehat{G}(\mathbf{x}, p)$. This approach differs from the adaptive

strategy, where c would be updated iteratively during the process. Determining an optimal value for c in advance requires an accurate estimate of the maximum value of the likelihood function, which is not known a priori and must therefore be estimated.

In this context, the Kriging model and the *Expected Improvement* (EI) acquisition function are used to select new training points for DoE. The EI criterion is a widely used acquisition function in global optimization, particularly within the Efficient Global Optimization (EGO) framework [24]. In this context, we adapt the EI criterion in order to maximize the log-likelihood function.

The EI function evaluates the expected amount of improvement over the best current prediction, thereby guiding the optimization process towards regions of the input space that are likely to yield better solutions. It balances exploration and exploitation by considering both the predicted mean and the associated uncertainty of the surrogate model. In practice, the EI function seeks the sample point that maximizes the expected improvement relative to the current known minimum of the objective function, promoting efficient convergence to the global optimum. Mathematically, it is defined as:

$$\mathrm{EI} = \mathbb{E}\left[\max((\widehat{g}_{\min}) - (\widehat{g}(\mathbf{x})), 0)\right] \tag{21}$$

which can be expanded as:

$$\mathrm{EI} = \int_{-\infty}^{\widehat{g}_{\min}} (\widehat{g}_{\min} - \widehat{g}(\mathbf{x})) \frac{1}{\sqrt{2\pi}\sigma_{\widehat{g}}(\mathbf{x})} \exp\left(-\frac{(\widehat{g}(\mathbf{x}) - \mu_{\widehat{g}}(\mathbf{x}))^2}{2\sigma_{\widehat{g}}^2(\mathbf{x})}\right) d\widehat{g}(\mathbf{x}) \tag{22}$$

The closed-form expression of the EI function is given by:

$$\mathrm{EI} = (\widehat{g}_{\min} - \mu_{\widehat{g}}(\mathbf{x}))\, \Phi\left(\frac{\widehat{g}_{\min} - \mu_{\widehat{g}}(\mathbf{x})}{\sigma_{\widehat{g}}(\mathbf{x})}\right) + \sigma_{\widehat{g}}(\mathbf{x})\, \phi\left(\frac{\widehat{g}_{\min} - \mu_{\widehat{g}}(\mathbf{x})}{\sigma_{\widehat{g}}(\mathbf{x})}\right) \tag{23}$$

where:

- $\widehat{g}_{\min} = \min(\widehat{g}(\mathbf{x}))$ is the current minimum prediction,
- $\mu_{(\widehat{g})}(\mathbf{x})$ and $\sigma_{(\widehat{g})}(\mathbf{x})$ are the mean and standard deviation of the Kriging model,
- $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function (CDF) and the probability density function (PDF) of the standard normal distribution, respectively.

The new training point is then selected by maximizing the EI criterion:

$$\mathbf{x}_{new} = \arg\max\ \mathrm{EI}(\mathbf{x}) \tag{24}$$

The procedure is outlined in Algorithm 2. A Kriging model of $L(\mathbf{x}|\boldsymbol{y})$ is first constructed within the parameter space $\mathbf{X}^*$. This surrogate model is subsequently employed to estimate the maximum of the likelihood, allowing for a more efficient and precise determination of the constant c. Once c is calculated, the limit-state function can be constructed, enabling the application of the SuS method in combination with the qAK method to determine the posterior distributions of the model's parameters.

Algorithm 2 EGO Framework [24]

Require: The size N_0 of the initial training dataset $\mathbf{X}^*$, the size N_{cand} of the candidate sample $\mathbf{X}$ and the maximum number of iterations N_{iter}

1: Initial training set $\{\mathbf{X}^*, L(\{\mathbf{x}_k^*\}_{k=1}^{N_0})\}\}$
2: **for** $t = 0 \rightarrow N_{\text{iter}}$ **do**
3: Construct a Kriging model with $\{\mathbf{X}_{(t)}^*, L(\{\mathbf{x}_k^*\}_{k=1}^{N_t})\}$
4: $\mathbf{x}_{\text{new}}^* = \arg\max_{x \in \mathbf{X}} EI(\mathbf{x})$
5: calculate $L(\mathbf{x}_{new})$
6: Update training dataset
7: **return** Calculate $\min(\exp(\widehat{g}(\mathbf{x})))$ recorded after optimization

5 NUMERICAL EXAMPLE: TWO-DOF SHEAR BUILDING

In this example, a two-degree-of-freedom system is examined. The masses are considered deterministic, with values $m_1 = 16.531 \times 10^3$ kg and $m_2 = 16.131 \times 10^3$ kg. The corresponding stiffnesses are $K_1 = X_1 k_n$ and $K_2 = X_2 k_n$, where $k_n = 29.7 \times 10^6$ N/m represents the nominal stiffness, and X_1 and X_2 are correction factors, and here are the parameters to be updated. More details can be found in Straub and Papaioannou [5] and Betz *et al.* [6].

Observations of the first two eigenfrequencies, f_1 and f_2, are adopted to update the prior distribution of the parameter vector $\mathbf{X} = [X_1, X_2]^T$ to its posterior distribution. The likelihood function is formulated as:

$$L(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{J(\mathbf{x})}{2\sigma_\varepsilon^2}\right), \tag{25}$$

where the measure-of-fit function $J(\mathbf{x})$ is defined as:

$$J(\mathbf{x}) = \sum_{j=1}^{2} \lambda_j^2 \left(\frac{f_j^2(\mathbf{x}) - \tilde{f}_j^2}{\tilde{f}_j^2}\right)^2 \tag{26}$$

$f_j(\mathbf{x})$ denotes the j-th eigenfrequency predicted by the model based on the parameter vector $\mathbf{x}$, while $\tilde{f}_j$ represents the corresponding observed eigenfrequency. The BUS failure function, thus becomes:

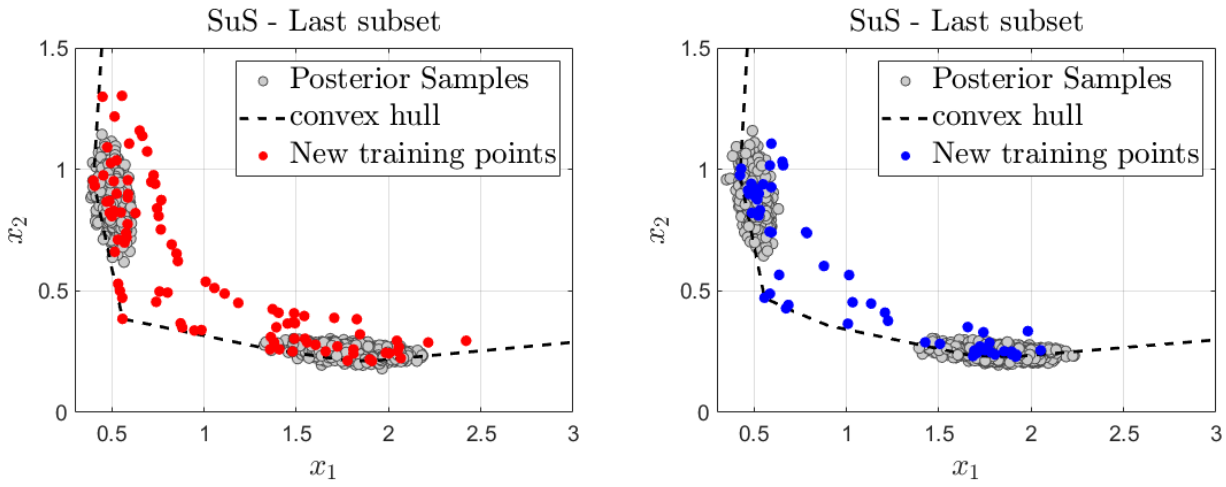
$$G(\mathbf{x}) = p - c \cdot \exp\left(-\frac{J(\mathbf{x})}{2\sigma_\varepsilon^2}\right) \tag{27}$$

The weighting factors are $\lambda_1 = \lambda_2 = 1$, and the standard deviation of the measurement error is $\sigma_\varepsilon = 1/16$. The measured eigenfrequencies are $\widehat{f}_1 = 3.13$Hz, $\widehat{f}_2 = 9.83$ Hz. The prior distributions of X_1 and X_2 are independent lognormal distributions with modes of 1.3 and 0.8, respectively, and standard deviations $\sigma_{X_1} = \sigma_{X_2} = 1$.

We compare the two active learning-based methods with the original techniques. Especially, the non-adaptive and the adaptive methods are compared with BUS [5] and aBUS [6], respectively. The comparison is made in terms of accuracy in the calculation of the constant c, the posterior distributions, and of the computational cost, i.e., the number of function evaluations and the size of the candidate sample for selecting new support points. In Table 1 the results for each method are shown. Moreover, the mean value and the standard deviation of the two parameters are presented in Table 1.

Table 1: Comparison of adaptive and non-adaptive methods with original methods (without Kriging). Reported are the estimated constant c, mean and standard deviation of updated parameters, number of function evaluations (N_{call}), and accepted samples (N_{acc})

Variable	Methods	c	μ	σ	N_{call}	N_{acc}
X_1	BUS	1.000	1.1200	0.6600	15000	5000
	non-adaptive method	1.003	1.1246	0.6589	215	5000
	aBUS	1.0006	1.0960	0.6590	15000	5000
	adaptive method	1.0001	1.1076	0.6690	143	5000
X_2	BUS		0.5800	0.3200		
	non-adaptive method		0.5840	0.3230		
	aBUS		0.6049	0.3270		
	adaptive method		0.6151	0.3310		

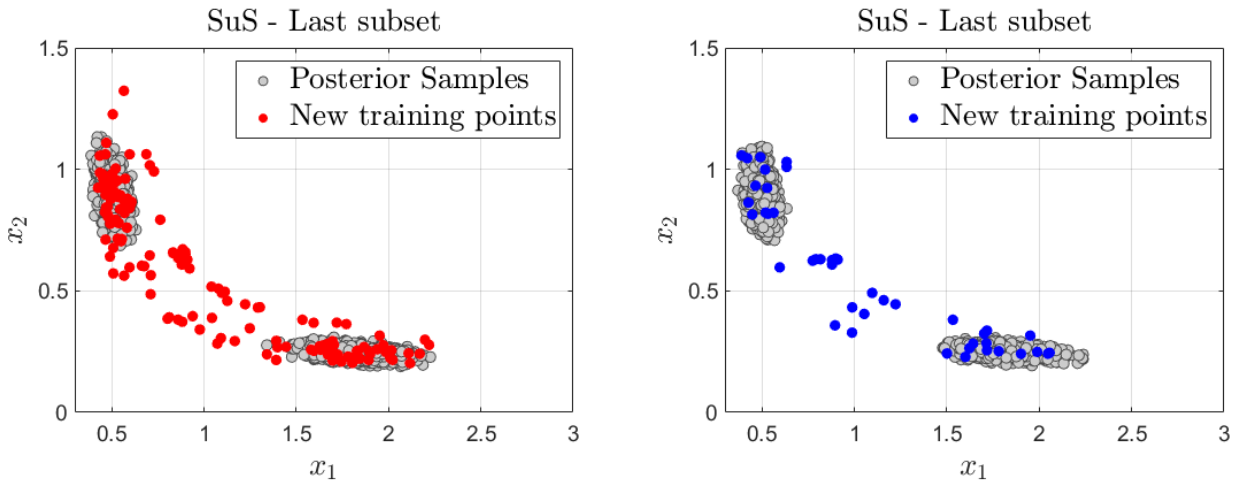


(a) (b)

Figure 2: Two-dof shear building: Posterior samples and enrichment of the DoE using the adaptive method 4.2 in combination with (a) the classical AK method [19] and (b) the qAK method [17].

Figure 2 illustrates the final subset of SuS for a two-degree-of-freedom shear building, comparing two adaptive enrichment strategies with distinct stopping criteria. Both plots display posterior samples (gray circles) and a convex hull (dashed black line), with new training points added using the adaptive method described in Section 4.2. In Figure 2a, red points represent training samples from the classical AK method [19], while in Figure 2b, blue points correspond to the qAK method [17]. Differences are shown in the number of new support points due to the different stopping criteria. The AK method typically uses a fixed threshold-based stopping criterion, which is very conservative. The qAK method employs a probabilistic-based stopping criterion, which is more accurate, leading to fewer new training points achieving enough accuracy.

Figure 3 shows the distribution of posterior samples and new training points at the SuS stage for the two-degree-of-freedom shear building model under the non-adaptive method. Figure 4.3a uses



(a) (b)

Figure 3: Two-dof shear building: Posterior samples and enrichment of the DoE using the non-adaptive method 4.2 in combination with (a) the classical AK method [19] and (b) the qAK method [17].

the classical AK method, resulting in a widespread distribution of red new training points, while fig 4.3b employs the qAK method, showing more localized blue new training points. Both subfigures illustrate how the training points enrich the design of experiments (DoE), with clear differences in exploration strategies between the two methods. In this case, the constant c was calculated using EGO. The training points of the Kriging model at the EGO stage are used as the initial training points at the SuS method.

6 CONCLUSIONS

This paper evaluates the performance of active-learning Kriging techniques for Bayesian updating by reformulating the inference task as a reliability analysis within the Bayesian Updating with Structural reliability framework. We employ our recently developed quantified active-learning Kriging (qAK) algorithm to solve the resulting reliability problems and recover the posterior distributions. A crucial challenge is the accurate estimation of the constant c via a surrogate Kriging model; to address this, we compare two strategies for its determination: an adaptive BUS approach that computes c during active learning, and a standard BUS approach that fixes c a prior. The adaptive scheme leverages computational-geometry tools—specifically, convex-hull computations—to guide enrichment, whereas the non-adaptive method estimates c by solving a global optimization problem driven by an acquisition function in the Design of Experiments. We benchmark these methods on a numerical example, highlighting their respective advantages and limitations, and demonstrate the powerful synergy between active learning and Bayesian inference for structural-engineering applications.

7 ACKNOWLEDGMENT

The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “2nd Call for H.F.R.I. Research Projects to support Faculty Members & Researchers” (Project Number: 04140).

REFERENCES

[1] Gilks, W.R. and Wild, P. Adaptive rejection sampling for Gibbs sampling. *J. R. Stat. Soc. Ser. C. Appl. Stat.* (1992) **4**:337-348.

[2] Beck, J.L. and Au, S.K. Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation. *J. Eng. Mech.* (2002) **128**:380-391

[3] Ching, J. and Chen, Y.-C. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *J. Eng. Mech.* (2007) **832**:816–832.

[4] Betz W., Papaioannou I., Straub D. Transitional Markov chain Monte Carlo: Observations and improvements. *J. Eng. Mech.* (2016) **142**:04016016.

[5] Straub, D. and Papaioannou, I. Bayesian Updating with Structural Reliability Methods. *J. Eng. Mech.* (2015) **141**:04014134.

[6] Betz, W., Papaioannou, I., Beck, J.L. and Straub, D. Bayesian inference with Subset Simulation: Strategies and improvements. *Comput. Methods Appl. Mech. Eng.* (2018) **331**:72-93.

[7] Betz, W., Beck, J.L., Papaioannou, I. and Straub, D. Bayesian inference with reliability methods without knowing the maximum of the likelihood function. averaging. *Probab. Eng. Mech.* (2018) **53**:14–22.

[8] Angelikopoulos, P., Papadimitriou, C. and Koumoutsakos, P. X-TMCMC: adaptive Kriging for Bayesian inverse modeling. *Comput. Methods Appl. Mech. Eng.* (2015) **289**:409-428

[9] Giovanis, D.G. Papaioannou, I., Straub, D. and Papadopoulos, V. Bayesian updating with subset simulation using artificial neural networks. *Comput. Methods Appl. Mech. Eng.* (2017) **319**:124-145

[10] Song C., Wang Z., Shafieezadeh A. and Xiao R. BUAK-AIS: Efficient Bayesian Updating with Active learning Kriging-based Adaptive Importance Sampling. *Comput. Methods Appl. Mech. Eng.* (2022) **391**:114578

[11] Jiang X. and Lu Z. Adaptive Kriging-based Bayesian updating of model and reliability. *Str. Saf.* (2023) **104**:102362.

[12] Feng K., Lu Z., Wang J., He P., and Dai Y. Efficient reliability updating methods based on Bayesian inference and sequential learning Kriging. *Str. Saf.* (2023) **104**:102366.

[13] Rossat D., Baroth J., Briffaut M. and Dufour F. Bayesian inversion using adaptive Polynomial Chaos Kriging within Subset Simulation. *J. Comput. Phys.* (2022) **455**:110986.

[14] Rasmussen, C.E. Gaussian Processes for Machine Learning. *MIT Press* Cambridge 2006.

[15] Jerez, D.J., An effective implementation of reliability methods for Bayesian model updating of structural dynamic models with multiple uncertain parameters *Reliab. Eng. Syst. Saf.* (2022) **225**:108634

[16] Au, S.K. and Beck, J.L. Estimation of small failure probabilities in high dimensions by subset simulation. *Probab. Eng. Mech.* (2001) **16**:2633-277

[17] Prentzas, I. and Fragiadakis, M. Quantified active learning Kriging model for structural reliability analysis. *Probab. Eng. Mech.* (2024) **78**:103699.

[18] Papaioannou, I., Betz, W., Zwirglmaier, K. and Straub, D. MCMC algorithms for subset simulation *Probab. Eng. Mech.* (2015) **41**:89-103

[19] Echard, B., Gayton, N. and Lemaire, M. AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Struct. Saf.* (2011) **33**:145-154

[20] MacQueen, J. Bayesian Some methods for classification and analysis of multivariate observations. *In:Le Cam LM, Neyman J, editors. Proceedings of the Berkeley symposium on mathematical statistics and probability* (1979) **1**:281-297

[21] Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* (1982) **28**:129-137

[22] Prentzas, I. and Fragiadakis, M. A probabilistic based adaptive refinement for reliability assessment using Kriging metamodel. *9th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Athens Greece, 12-14 June* (2023).

[23] Barber, C. Bradford and Dobkin, David P. and Huhdanpaa, Hannu The quickhull algorithm for convex hulls. *Assoc. Comput. Mach.* (1996) **22**:469-483

[24] Jones, D.R., Schonlau, M. and Welch, W.J. Efficient Global Optimization of Expensive Black-Box Functions. *J. Glob. Optim.* (1998) **13**:455-492