



HAL
open science

Bayesian inversion using adaptive Polynomial Chaos Kriging within Subset Simulation

D. Rossat, J. Baroth, Matthieu Briffaut, F. Dufour

► **To cite this version:**

D. Rossat, J. Baroth, Matthieu Briffaut, F. Dufour. Bayesian inversion using adaptive Polynomial Chaos Kriging within Subset Simulation. *Journal of Computational Physics*, 2022, 455, pp.110986. 10.1016/j.jcp.2022.110986 . hal-04496087

HAL Id: hal-04496087

<https://hal.science/hal-04496087>

Submitted on 22 Jul 2024

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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Bayesian inversion using adaptive Polynomial Chaos Kriging within Subset Simulation

D. Rossat^{a,b,*}, J. Baroth^a, M. Briffaut^a, F. Dufour^a

^a*Univ. Grenoble Alpes, CNRS, Grenoble INP, 3SR, F-38000 Grenoble, France*

^b*Electricité de France (EDF-DIPNN-DT), 69007 Lyon, France*

Abstract

In this paper, we propose a Bayesian inversion approach combining adaptive *Polynomial Chaos Kriging* (PCK) surrogate models and a rare event estimation method called *Subset Simulation* (SuS). It is based on the recently introduced *Bayesian Updating with Structural reliability* (BUS) framework that enables to reformulate the classical Bayesian inference into a rare event estimation problem. In this context, the SuS method aims at drawing samples from the posterior distribution as well as estimating the model evidence, which is usually computationally intractable when considering classical MCMC approaches. The proposed approach involves the construction of a PCK surrogate model which provides both global and local approximations of the likelihood function, through the combination of a Polynomial Chaos and Kriging surrogates. Furthermore, we propose an adaptive scheme for enriching the PCK surrogate throughout the SuS sampling procedure, in order to improve its accuracy near informative regions. The applicability and the efficiency of the proposed approach are assessed through its application to three cases studies with increasing complexity. Results highlighted that the proposed approach enables to accurately approximating posteriors with a limited amount of full model calls, even in the case of multi-modal posteriors, which are usually difficult to sample when using classical MCMC algorithms.

Keywords: , Bayesian inversion, Polynomial Chaos, Kriging, Structural Reliability Methods, Subset Simulation, Inverse problems

*Corresponding author

Email address: donatien.rossat@3sr-grenoble.fr (D. Rossat)

1. Introduction

In science and engineering, computational models are widely used for assessing the response of a physical system. Such models can be seen as input-output maps, which predict measurable quantities related to the system response for a given set of input parameters. Inverse problems consist of inferring these parameters given a set of observations of the system response [1].

In most cases, solving inverse problems remains a challenging task due to their *ill-posedness* in the Hadamard sense: indeed, a solution may not exist and/or be not unique, and may not be continuously varying with respect to the observation data. In this context, the Bayesian formulation for inverse problems provides a general framework for quantifying epistemic uncertainties on parameters, and assimilating noisy observation data in order to update these uncertainties. Based on probability theory, it requires the definition of a *prior* distribution which represents the beliefs before collecting observation data, and contributes to the regularization of ill-posed inverse problems [2, 3]. A discrepancy model is also required in order to link model predictions to observations, through the definition of a so-called *likelihood function*. Then, Bayesian inference provides a *posterior* distribution which summarizes all the available knowledge on parameters once the observation data have been collected and analyzed. This posterior distribution may subsequently be used for computing some Quantities of Interest (QoI) related to parameters, such as mean or covariance matrix, or marginals distributions. It is also possible to make new predictions by integrating over the posterior. Furthermore, the so-called *model evidence*, given by the normalization constant of the posterior, may be used for quantifying the plausibility of a given model given the available data, in the framework of model selection [4].

The computation of the posterior constitutes a major task of Bayesian inference. In general, a closed-form expression is not available, notably due to the fact that the model evidence is often a multi-dimensional integral, **the calculation of which is often intractable**. Then, the posterior is usually estimated from samples, and Monte Carlo estimates are often considered for computing posterior QoI. In this context, Markov Chain Monte Carlo (MCMC) techniques constitute a widely used class in the framework of Bayesian computations [5, 6]. They typically consist of generating Markov chains in the parameter space to draw samples from the posterior distribution, by avoiding the computation of the model evidence. In this context, the Metropolis-Hastings (MH) [7, 8] constitutes the cornerstone of MCMC algorithms. A wide range of MCMC variants have been developed, including adaptive variants [9, 10], algorithms based on Hamiltonian dynamics [11], or affine-invariant algorithms [12]. Nevertheless, MCMC algorithms suffer from several drawbacks: firstly, the Markovian nature of such processes induces sample auto-correlation, which may alter the reliability of Monte Carlo estimates of posterior QoI. Secondly, the absence of clear convergence criterion constitutes a major drawback. Instead, numerical heuristic criterion have been proposed in the literature [13, 14, 15]. Lastly, MCMC algorithms require a large amount of likelihood evaluations, which make such algorithms **intractable** when dealing with computationally extensive models.

Then, few alternatives to MCMC-based approaches have been proposed in the literature, such as Variational

36 Bayesian (VB) inference [16] aiming at computing the posterior by solving an optimization problem, Approximate
37 Bayesian Computation (ABC) [17], Laplace approximations [18] or Nested Sampling [19, 20]. Recently, some
38 authors have proposed novel *sampling-free* approaches, in the sense that an analytical form of the posterior is
39 computed instead of drawing samples for estimating it. El Moselhy & Marzouk [21] introduced an approach
40 based on optimal transport maps, which push forward the prior to the posterior. Nagel & Sudret [22] introduced
41 the so-called Spectral Likelihood Expansion (SLE) approach, based on a spectral representation of the likelihood
42 function in a basis of polynomials which are orthonormal with respect to the prior. Then, as a generalization
43 of SLE, Wagner et al. [23] proposed an adaptive approach based on Stochastic Spectral Embedding (SSE) [24],
44 which aims at constructing local spectral expansions of the likelihood function.

45 Another recent novel framework for Bayesian computations has been introduced by Straub & Papaionannou
46 [25], known as Bayesian Updating with Structural reliability methods (BUS). It aims at reformulating the
47 classical Bayesian inference formulation into a rare event estimation problem [26]. Structural Reliability (SR)
48 methods [27, 28] have been developed in order to estimate the probability of rare events, based on physical
49 and engineering models. Then, such methods can be used in the BUS framework, in order to compute the
50 posterior as well as the model evidence, since the latter can be written as the probability of a rare event [26]. In
51 particular, the Subset Simulation (SuS) method [29] constitutes a widely used SR method, enabling to estimate
52 the probability of rare events for a wide range of applications. Such method has been recently implemented in
53 the BUS framework [30], in order to efficiently draw samples from posterior distributions.

54 Nevertheless, such approach requires an important amount of likelihood evaluations, as MCMC sampling
55 techniques. Firstly, some authors proposed to accelerate MCMC calculations by using surrogate models. In this
56 context, surrogates may be constructed offline, and then used for replacing the full forward model throughout
57 MCMC sampling. Such strategy has been considered with Polynomial Chaos (PC) surrogates based on the prior
58 distribution [31, 32, 33], or with Kriging surrogates [34]. Another approach consists in adaptively refining the
59 surrogate model throughout the sampling procedure, as proposed in [35] for PC surrogates or in [36] for Deep
60 Neural Networks (DNN), to name a few. Likewise, in the BUS framework, some authors recently proposed
61 approaches based on surrogate models: Giovanis et al. [37] proposed an approach based on SuS and Artificial
62 Neural Networks (ANN), whereas Wang & Shafieezadeh [38] combined adaptive Kriging and Rejection Sampling
63 (RS).

64 Moreover, SR methods provide a wide range of active learning approaches, enabling to construct surrogate
65 models with adaptive experimental designs, which are iteratively enriched to improve the accuracy of surrogate
66 model predictions of QoI related to rare events [39, 40]. In particular, such approaches tend to limit the number
67 of model calls, which makes their use suitable for computationally expensive models. In this context, *local*
68 surrogate models such as Kriging [41, 38] are often considered, due to the local error measure they provide,
69 which can be used for selecting new enrichment points for experimental designs. Recently, Schöbi et al. [42]
70 proposed a new surrogate modeling technique named Polynomial Chaos Kriging (PCK), which combines a PC
71 for approximating the global behavior of the model, and a Kriging for catching local variations. PCK provides

72 Kriging local features (*e.g.* prediction variance) to a PC surrogate, by embedding a PC into the trend of a
73 Kriging process. Hence, PCK surrogate models are well suited for Kriging-based active learning approaches,
74 which makes it a powerful candidate for rare event estimation, as highlighted in [43].

75 In this paper, a hybrid Bayesian inversion approach based on the BUS framework is proposed, which uses
76 adaptive PCK surrogate models within SuS. **In this context, our main contributions are the following:**

- 77 • We propose an implementation of PCK within SuS in the BUS framework, to accelerate the SuS sampling
78 procedure. The main idea is to construct a PCK surrogate of the log-likelihood function, which can be
79 enriched in an adaptive way throughout the rare event estimation procedure performed with SuS.
- 80 • In order to improve the accuracy of the PCK in posterior high-probability zones with a limited number
81 of full model calls, we propose an active learning scheme for adaptively refining the PCK surrogate in
82 posterior high-probability zones.

83 Firstly, a global approximation of the log-likelihood on the prior support is constructed, by using an Exper-
84 imental Design (ED) of limited size. It is well known that the accuracy of prior-based surrogates near posterior
85 high-probability zones cannot be guaranteed in general [44, 35]. This motivates the introduction of the pro-
86 posed active learning scheme: the ED of the PCK is then iteratively enriched by selecting points in *informative*
87 zones, based on a combination of PCK local features and a clustering technique [45]. The proposed approach
88 enables to efficiently improve the local accuracy of the PCK surrogate with a limited amount of full model calls.
89 More specifically, such adaptive scheme enables to handle multi-modal posteriors, which are difficult to sample
90 when using classical MCMC sampling approaches. The approach fully exploits the versatility of PCK, by firstly
91 capturing the global trend of the posterior and subsequently focusing on informative zones.

92 The paper is organized as follows: the basics of Bayesian inference and the BUS framework are briefly
93 presented in Section 2. Then, Section 3 establishes the main theoretical and practical basics related to PCK
94 surrogate models. The proposed Bayesian inversion approach is subsequently described in Section 4. The
95 performance of the proposed approach is assessed in Section 5, through four cases studies of increasing complexity.
96 Lastly, the paper is summarized and concluded in Section 6.

97 2. Bayesian Updating with Structural reliability methods (BUS)

98 2.1. Bayesian inference

99 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω is a sample space, \mathcal{F} a Borel algebra on Ω and \mathbb{P} a probability
100 measure on (Ω, \mathcal{F}) . Consider a system with uncertain parameters, which are represented by a random vector
101 $\Theta : \Omega \rightarrow \mathcal{D}_\theta \subset \mathbb{R}^m$. The behavior of the system is represented by an computational model $\mathcal{M} : \mathcal{D}_\theta \rightarrow \mathcal{D}_y \subset \mathbb{R}^n$
102 which maps a set of parameters $\theta \in \mathcal{D}_\theta$ to a response $\mathcal{M}(\theta)$. The random vector Θ is assumed to admit a

103 Probability Density Function (PDF) denoted by $\rho(\boldsymbol{\theta})$, which expresses the prior belief in Θ . Then, let suppose
 104 that a set of observations of the system response is available, and gathered in a vector $\mathbf{y} \in \mathcal{D}_{\mathbf{y}}$. The so-called
 105 *likelihood function* $\mathcal{L}(\boldsymbol{\theta}; \mathbf{y})$ is proportional to the probability of observing \mathbf{y} given a realization $\Theta = \boldsymbol{\theta}$ of the
 106 uncertain parameters. In order to link observations to responses of \mathcal{M} , an additive Gaussian discrepancy model
 107 is often considered:

$$\mathbf{y} = \mathcal{M}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \quad (2.1)$$

108 where $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$ is a normal random vector with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$. In this case, the
 109 likelihood function reads:

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) = \phi_n(\mathbf{y} - \mathcal{M}(\boldsymbol{\theta}); \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \quad (2.2)$$

110 where $\phi_n(\cdot; \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$ is the PDF of the distribution $\mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$.

111 Then, applying Bayes' theorem, the *posterior* PDF $\pi(\boldsymbol{\theta}|\mathbf{y})$ of Θ can be written as follows:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\rho(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta}; \mathbf{y})}{\mathcal{Z}} \quad (2.3)$$

112 where \mathcal{Z} is a normalization constant, known as *model evidence*:

$$\mathcal{Z} = \int_{\mathcal{D}_{\boldsymbol{\theta}}} \rho(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta}; \mathbf{y})d\boldsymbol{\theta} \quad (2.4)$$

113 The posterior PDF (2.3) summarizes all available information on uncertain parameters Θ after the data
 114 have been collected. In most cases, the posterior PDF does not admit a closed-form expression. Moreover, the
 115 model evidence is often a multidimensional integral which is computationally expensive to evaluate. In this
 116 context, Markov Chain Monte Carlo (MCMC) methods are widely used for drawing samples from the posterior
 117 distribution, without computing the model evidence [5, 6]. Such methods aim at constructing an ergodic Markov
 118 chain $(\Theta^{(\tau)})_{\tau \in \mathbb{N}}$ whose invariant distribution is the posterior distribution, which writes:

$$\pi(\boldsymbol{\theta}^{(\tau+1)}|\mathbf{y}) = \int_{\mathcal{D}_{\boldsymbol{\theta}}} \pi(\boldsymbol{\theta}^{(\tau)}|\mathbf{y})\mathcal{K}(\boldsymbol{\theta}^{(\tau)}, \boldsymbol{\theta}^{(\tau+1)})d\boldsymbol{\theta}^{(\tau)} \quad (2.5)$$

119 where $\mathcal{K}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is the density of the transitional distribution from the state $\boldsymbol{\theta}^{(\tau)}$ of the chain $(\Theta^{(\tau)})_{\tau \in \mathbb{N}}$ to
 120 the state $\boldsymbol{\theta}^{(\tau+1)}$. A sufficient condition for Eq.(2.5) to apply is the so-called *detailed balance* condition, which
 121 expresses the time-reversibility of the Markov chain:

$$\pi(\boldsymbol{\theta}^{(\tau)}|\mathbf{y})\mathcal{K}(\boldsymbol{\theta}^{(\tau)}, \boldsymbol{\theta}^{(\tau+1)}) = \pi(\boldsymbol{\theta}^{(\tau+1)}|\mathbf{y})\mathcal{K}(\boldsymbol{\theta}^{(\tau+1)}, \boldsymbol{\theta}^{(\tau)}) \quad (2.6)$$

122 This condition is generally verified for most MCMC algorithms, such as the well-known Metropolis-Hastings
 123 algorithm [7, 8], which constitutes the cornerstone of MCMC.

124 2.2. The BUS framework

125 2.2.1. Structural Reliability problems

126 The BUS framework, initially introduced in [25], consists in reformulating the classical Bayesian updating
 127 problem presented in Section 2.1 into a Structural Reliability (SR) problem. SR methods aim at estimating the
 128 probability of rare events [27, 46]. Such events are usually described in terms of a *limit state function* (LSF)
 129 $h : \mathcal{D}_{\mathbf{x}} \rightarrow \mathbb{R}$, whose values are classifying *safety* and *failure* domains D_s, D_f respectively, in the space $\mathcal{D}_{\mathbf{x}}$ of
 130 the random parameters \mathbf{X} of the reliability problem:

$$\begin{aligned} D_s &= \{\mathbf{x} \in \mathcal{D}_{\mathbf{x}} \mid h(\mathbf{x}) > 0\} \\ D_f &= \{\mathbf{x} \in \mathcal{D}_{\mathbf{x}} \mid h(\mathbf{x}) \leq 0\} \end{aligned} \quad (2.7)$$

131 The main goal of SR methods is to estimate the failure probability $\mathbb{P}_f = \mathbb{P}(h(\mathbf{X}) \in D_f)$, which can be written
 132 as the following integral if \mathbf{X} admits a density $\rho(\mathbf{x})$:

$$\mathbb{P}_f = \int_{\mathcal{D}_{\mathbf{x}}} \mathbf{1}_{D_f}(\mathbf{x}) \rho(\mathbf{x}) d\mathbf{x} \quad (2.8)$$

133 where $\mathbf{1}_A$ is the indicator function of $A \subset \mathcal{D}_{\mathbf{x}}$.

134 The integral which defines the probability (2.8) can not be solved analytically in the general case. The
 135 Rejection Sampling (RS) method aims at estimating \mathbb{P}_f by Monte Carlo simulation:

$$\widehat{\mathbb{P}}_f = \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{D_f}(\mathbf{x}^{(n)}) \quad (2.9)$$

136 where $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)})$ are samples drawn from the distribution of the parameters \mathbf{X} . The accuracy of this
 137 estimate may be measured by computing the Coefficient of Variation (CoV) of $\widehat{\mathbb{P}}_f$, which is given by:

$$\mathbf{CoV} [\widehat{\mathbb{P}}_f] = \sqrt{\frac{1 - \widehat{\mathbb{P}}_f}{N \widehat{\mathbb{P}}_f}} \quad (2.10)$$

138 It appears from Eq. (2.10) that such direct Monte Carlo estimation may become inefficient when failure proba-
 139 bilities are small. Indeed, the required number of samples to obtain a target CoV dramatically increases when
 140 the failure probability is decreasing.

141 In this context, several alternative SR methods have been developed in order to provide more efficient
 142 estimations of failure probabilities, such as First Order Reliability Method (FORM) [28], importance sampling
 143 [47], or Subset Simulation (SuS) [29]. One will focus on the latter method in the present paper.

144 *2.2.2. BUS framework formulation*

145 Keeping the notations introduced in Section 2.1, let $p : \Omega \rightarrow \mathcal{D}_p = [0, 1]$ a random variable uniformly
 146 distributed on $[0, 1]$ (*i.e.* $p \sim \mathcal{U}([0, 1])$), which is assumed to be independent from the parameters Θ . The BUS
 147 framework proposes to consider the parameters (Θ, p) which take values in the augmented space $\mathcal{D}_\Theta \times \mathcal{D}_p$, as
 148 well as the failure event $F \subset \mathcal{D}_\Theta \times \mathcal{D}_p$ defined by:

$$F = \{(\boldsymbol{\theta}, p) \in \mathcal{D}_\Theta \times \mathcal{D}_p \mid p \leq c\mathcal{L}(\boldsymbol{\theta}; \mathbf{y})\} \quad (2.11)$$

149 where $c > 0$ is a constant, chosen such that $c\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) \leq 1$, $\forall \boldsymbol{\theta} \in \mathcal{D}_\Theta$. The choice of such constant will be detailed
 150 in further sections.

151 It has been demonstrated in [25] that the posterior PDF $\pi(\boldsymbol{\theta}|\mathbf{y})$ given in Eq. (2.3) rewrites as follows:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{\mathbb{P}_f} \int_{\mathcal{D}_p} \mathbf{1}_F(\boldsymbol{\theta}, p) \rho(\boldsymbol{\theta}) dp \quad (2.12)$$

152 where \mathbb{P}_f is the failure probability given by:

$$\mathbb{P}_f = \mathbb{P}((\boldsymbol{\Theta}, p) \in F) = \int_{\mathcal{D}_\Theta} \int_{\mathcal{D}_p} \mathbf{1}_F(\boldsymbol{\theta}, p) \rho(\boldsymbol{\theta}) dp d\boldsymbol{\theta} \quad (2.13)$$

153 Thus, Eqs. (2.12) and (2.13) ensure that sampling from the posterior distribution is tantamount to solving the
 154 reliability problem in the augmented space $\mathcal{D}_\Theta \times \mathcal{D}_p$ with the LSF given by [25]:

$$h(\boldsymbol{\theta}, p) = p - c\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) \quad (2.14)$$

155 Moreover, it is worth noting that the failure probability \mathbb{P}_f is directly linked to the model evidence given by Eq.
 156 (2.4). Indeed, rewriting the integral in Eq. (2.13) yields:

$$\mathbb{P}_f = c \int_{\mathcal{D}_\Theta} \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) \rho(\boldsymbol{\theta}) d\boldsymbol{\theta} = c\mathcal{Z} \quad (2.15)$$

157 *2.3. BUS and subset simulation*

158 *2.3.1. Formulation*

159 Subset Simulation (SuS) [29] is a widely used SR method which enables to efficiently estimate small failure
 160 probabilities in high-dimensional random variable spaces. The main idea of SuS consists in expressing the rare
 161 event F in Eq. (2.11) as an intersection of nested intermediate events $F_r \subset \dots \subset F_1 \subset F_0$, where F_0 is the
 162 certain event (*i.e.* $\mathbb{P}(F_0) = 1$) and $F_r = F$. In this framework, the failure probability in Eq. (2.13) rewrites:

$$\mathbb{P}_f = \mathbb{P}\left(\bigcap_{i=0}^r F_i\right) = \prod_{i=1}^r \mathbb{P}(F_i | F_{i-1}) \quad (2.16)$$

163 Thus, the failure probability \mathbb{P}_f is expressed as a product of larger conditional probabilities, which can be
 164 more efficiently estimated than a single small probability. Following [29] and [30], the intermediate events are
 165 defined by using a sequence of threshold values $0 = h_r < \dots < h_1 < h_0 = +\infty$:

$$F_i = \{(\boldsymbol{\theta}, p) \in \mathcal{D}_{\boldsymbol{\theta}} \times \mathcal{D}_p \mid h_l(\boldsymbol{\theta}, p) \leq h_i\} \quad (2.17)$$

166 where $h_l(\boldsymbol{\theta}, p)$ is a modified form of the LSF function $h(\boldsymbol{\theta}, p)$ given by Eq. (2.14):

$$h_l(\boldsymbol{\theta}, p) = \log(p) - \log(c) - \log(\mathcal{L}(\boldsymbol{\theta})) \quad (2.18)$$

167 where the dependence to data \mathbf{y} of \mathcal{L} will be omitted for the sake of concision.

168 Such LSF formulation has been introduced in [30] in order to ensure smooth transitions between intermediate
 169 events, and therefore to improve numerical stability of SuS within BUS [37].

170 The threshold values $(h_i)_{1 \leq i \leq r}$ are chosen so that the corresponding conditional probabilities $(\mathbb{P}(F_i|F_{i-1}))_{1 \leq i \leq r}$
 171 are equal to a value \mathbb{P}_t in average, usually chosen as $\mathbb{P}_t = 10\%$ [29, 48, 30]. In this way, the estimation of the
 172 probabilities $(\mathbb{P}(F_i|F_{i-1}))_{1 \leq i \leq r}$ through Monte Carlo estimation is sensibly more efficient than the direct estima-
 173 tion of \mathbb{P}_f . The generation of samples splits into two parts: firstly, samples are drawn from the prior distribution
 174 in order to estimate the probability $\mathbb{P}(F_1|F_0)$ through direct Monte Carlo (see Eq. (2.9)). Next, for each
 175 $i \in \{1, \dots, r\}$, samples conditional to intermediate event F_i are generated with MCMC sampling, as described
 176 by Algorithm 1. Finally, samples of the last SuS level correspond to samples of the posterior distribution of
 177 parameters $\boldsymbol{\Theta}$. It is important to underline that the model evidence (see Eq. (2.4)) is available as a by-product
 178 of the SuS procedure, due to the relation given in Eq.(2.15). Then, unlike classical MCMC sampling methods
 179 which provide posterior samples by avoiding to compute the model evidence, the SuS procedure applied in the
 180 BUS framework enables to both draw posterior samples and compute the model evidence.

181 2.3.2. MCMC sampling within SuS

182 In this paper, MCMC sampling (Step 5 of Algorithm 1) will be performed by using the adaptive Conditional
 183 Sampling (aCS) algorithm introduced in [49]. The samples are generated in an underlying standard normal
 184 probability space, also known as U-space, by considering an isoprobabilistic transform $\mathcal{T} : \mathcal{D}_{\boldsymbol{\theta}} \times \mathcal{D}_p \longrightarrow \mathcal{D}_{\mathbf{u}} =$
 185 \mathbb{R}^{m+1} , which maps the random vector $(\boldsymbol{\Theta}, p)$ to a standard normal uncorrelated random vector \mathbf{U} . Such mapping
 186 can be performed by using the Nataf [50, 51] or Rosenblatt [52] transform. Then, considering a Markov chain
 187 state $\mathbf{u} \in \mathcal{D}_{\mathbf{u}}$, a candidate sample $\mathbf{v} \sim \mathcal{N}_{m+1}(\mathbf{u}, \mathbf{I})$ is drawn, by imposing that \mathbf{u} and \mathbf{v} are jointly Gaussian with
 188 component-wise cross-correlation coefficients equal to $\rho \in [0, 1]$. This correlation parameter is adapted on the
 189 fly through MCMC sampling, in order to obtain an optimal Markov chain acceptance rate. For further details,
 190 the reader is referred to [49].

Algorithm 1: Subset Simulation in the BUS framework [29, 30].

Input: Number of samples per subset K , probability of intermediate subsets \mathbb{P}_t , BUS normalizing constant c

Output: Samples $(\boldsymbol{\theta}^{(i,k)}, p^{(i,k)})_{(i,k) \in \{0, \dots, r\} \times \{1, \dots, K\}}$, failure probability \mathbb{P}_f

Initialization;

1. Draw K samples $(\boldsymbol{\theta}^{(0,k)}, p^{(0,k)})_{k \in \{1, \dots, K\}}$ from the prior distribution of $(\boldsymbol{\Theta}, p)$;
2. Set $i = 0$ and $p_0 = 1$;

while $h_i > 0$ **do**

3. Set $i = i + 1$;
4. Sort $(\boldsymbol{\theta}^{(i-1,k)}, p^{(i-1,k)})_{k \in \{1, \dots, K\}}$ with respect to the LSF values $(h_l(\boldsymbol{\theta}^{(i-1,k)}, p^{(i-1,k)}))_{k \in \{1, \dots, K\}}$;
4. Set h_i as the \mathbb{P}_t -percentile of the set $(\boldsymbol{\theta}^{(i-1,k)}, p^{(i-1,k)})_{k \in \{1, \dots, K\}}$ with respect to the LSF values $(h_l(\boldsymbol{\theta}^{(i-1,k)}, p^{(i-1,k)}))_{k \in \{1, \dots, K\}}$;

if $h_i < 0$ **then**

- | Set $h_i = 0$ and $\mathbb{P}_i = \frac{n}{K}$, where $n = \text{Card} \left\{ k \in \{1, \dots, K\} \mid h_l(\boldsymbol{\theta}^{(i-1,k)}, p^{(i-1,k)}) \leq \max(h_i, 0) \right\}$;

else

- | Set $\mathbb{P}_i = \mathbb{P}_t$;

end

5. Generate the samples $(\boldsymbol{\theta}^{(i,k)}, p^{(i,k)})_{k \in \{1, \dots, K\}}$ by using MCMC. n Markov chains of length K/n are generated from the seeds $(\boldsymbol{\theta}^{(i-1,k)}, p^{(i-1,k)})_{k \in \{1, \dots, n\}}$;

end

6. Set $r = i$ and $\mathbb{P}_f = \prod_{j=0}^r \mathbb{P}_j$;

191 2.3.3. Choosing the BUS normalizing constant

192 Lastly, the choice of the BUS normalizing constant c plays a crucial role in the generation of posterior
 193 samples with the SuS procedure. The choice $c^{-1} = \mathcal{L}_{\max} = \max_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\theta}}} \mathcal{L}(\boldsymbol{\theta})$ clearly constitutes the optimal
 194 choice. However, such value is not known in advance in general. Moreover, evaluating the likelihood function
 195 may require an important computational cost, which makes Maximum Likelihood Estimate (MLE) **potentially**
 196 **intractable**. On the one hand, if $c^{-1} \leq \mathcal{L}_{\max}$, samples produced by BUS do not follow the posterior distribution
 197 [25, 53]. On the other hand, if $c^{-1} \geq \mathcal{L}_{\max}$, the efficiency of the BUS procedure may be affected, since the failure
 198 probability in Eq. (2.13) is proportional to c .

199 In this context, adaptive approaches have been proposed [54, 30] in order to learn the constant c during the
 200 SuS procedure. More specifically, such approaches consist in taking c as the largest likelihood within the already
 201 generated samples, and updating it when larger likelihood values are computed.

202 3. Surrogate modeling

203 3.1. Polynomial Chaos

204 3.1.1. Formulation

205 Keeping the notations introduced in Section 2, one supposes that the computational model takes scalar values,
 206 *i.e.* $\mathcal{M} : \mathcal{D}_{\boldsymbol{\theta}} \rightarrow \mathbb{R}$. The following can be easily extended to the case of vector valued models. Furthermore,

207 the random vector $\Theta : \Omega \rightarrow \mathcal{D}_\theta$ is supposed to have independent components, denoted by $(\Theta_1, \dots, \Theta_m)$.
 208 Polynomial Chaos formulations for dependent input random variables can be found in [55, 56]. The marginal
 209 parameter space associated to Θ_i will be denoted by \mathcal{D}_{θ_i} , so that $\mathcal{D}_\theta = \prod_{i=1}^m \mathcal{D}_{\theta_i}$.

210 The PDF ρ_i of each component Θ_i enables to define a Hilbert space, namely the space $\mathbb{L}^2(\rho_i)$ of functions
 211 defined on \mathcal{D}_{θ_i} which are square integrable with respect to ρ_i :

$$\mathbb{L}^2(\rho_i) = \left\{ \phi : \mathcal{D}_{\theta_i} \rightarrow \mathbb{R} \left| \mathbb{E} [\phi(\Theta_i)^2] = \int_{\mathcal{D}_{\theta_i}} \phi(\theta)^2 \rho_i(\theta) d\theta < +\infty \right. \right\} \quad (3.1)$$

212 This space is equipped with the following inner product:

$$\langle \phi | \psi \rangle_{\rho_i} = \int_{\mathcal{D}_{\theta_i}} \phi(\theta_i) \psi(\theta_i) \rho_i(\theta_i) d\theta_i = \mathbb{E} [\phi(\Theta_i) \psi(\Theta_i)] \quad (3.2)$$

213 for all $\phi, \psi \in \mathbb{L}^2(\rho_i)$. Then, it is possible to construct a Hilbert basis of $\mathbb{L}^2(\rho_i)$ formed by univariate polynomials
 214 $(\psi_j^{(i)})_{j \in \mathbb{N}}$ which are orthonormal with respect to ρ_i [57, 58]. Such construction has been originally introduced
 215 for Gaussian random variables by Wiener in [59], and then generalized to other distributions in [58] and [60].
 216 In general, it is possible to construct a Hilbert basis of polynomials for a given probability distribution, under
 217 some conditions precised in [61].

218 Moreover, let $\mathbb{L}^2(\rho)$ be the Hilbert space of functions defined on \mathcal{D}_θ which are square integrable with respect
 219 to ρ . This space is equipped with the inner product defined by $\langle \Phi | \Psi \rangle_\rho = \int_{\mathcal{D}_\theta} \Phi(\theta) \Psi(\theta) \rho(\theta) d\theta = \mathbb{E} [\Phi(\Theta) \Psi(\Theta)]$.
 220 Knowing that the components $(\Theta_1, \dots, \Theta_m)$ are mutually independent, their joint PDF π can be expressed as
 221 the tensor product of the marginals PDF:

$$\rho(\theta) = \prod_{i=1}^m \rho_i(\theta_i) \quad (3.3)$$

222 for all $\theta = (\theta_1, \dots, \theta_m) \in \mathcal{D}_\theta$. This ensures that $\mathbb{L}^2(\pi)$ is isomorphic to the tensorial product of Hilbert spaces
 223 $(\mathbb{L}^2(\rho_i))_{1 \leq i \leq m}$:

$$\mathbb{L}^2(\rho) \cong \bigotimes_{i=1}^m \mathbb{L}^2(\rho_i) \quad (3.4)$$

224 Consequently, the following multivariate polynomials define a hilbertian basis of $\mathbb{L}^2(\rho)$:

$$\Psi_\alpha(\theta) = \prod_{i=1}^m \psi_{\alpha_i}^{(i)}(\theta_i) \quad (3.5)$$

225 where the multi-index notation $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}^m$ has been used. Such polynomials are orthonormal
 226 with respect to π (*i.e.* $\langle \Psi_\alpha | \Psi_\beta \rangle_\rho = \delta_{\alpha\beta}$, where $\delta_{\alpha\beta}$ is the multi-index version of the Kronecker delta), and
 227 any function which belongs to $\mathbb{L}^2(\rho)$ can be expanded in the basis $(\Psi_\alpha)_{\alpha \in \mathbb{N}^m}$. Assuming that $\mathcal{M} \in \mathbb{L}^2(\rho)$, the
 228 so-called Polynomial Chaos Expansion (PCE) [57] of \mathcal{M} reads:

$$\mathcal{M}(\Theta) = \sum_{\alpha \in \mathbb{N}^m} a_\alpha \Psi_\alpha(\Theta) \quad (3.6)$$

229 where $(a_{\alpha})_{\alpha \in \mathbb{N}^m}$ are the PCE coefficients. For each multi-index $\alpha \in \mathbb{N}^m$, the coefficient a_{α} can be interpreted
 230 as the coordinate of the orthogonal projection of $\mathcal{M}(\Theta)$ on the vector space spanned by Ψ_{α} :

$$a_{\alpha} = \langle \mathcal{M}(\Theta) | \Psi_{\alpha} \rangle_{\rho} = \mathbb{E}[\mathcal{M}(\Theta) \Psi_{\alpha}(\Theta)] \quad (3.7)$$

231 3.1.2. Calculation of PC coefficients

232 For computational purposes, the PCE representation (3.6) is truncated by using a finite set $\mathcal{A} \subset \mathbb{N}^m$ of
 233 multi-indices:

$$\mathcal{M}(\Theta) \approx \widehat{\mathcal{M}}^{PC}(\Theta) = \sum_{\alpha \in \mathcal{A}} a_{\alpha} \Psi_{\alpha}(\Theta) \quad (3.8)$$

234 Re-indexing the multi-indices $\{\alpha \in \mathcal{A}\}$ by $\{1, \dots, P\}$ with $P = \text{Card}(\mathcal{A})$ yields the following vector notation:

$$\widehat{\mathcal{M}}^{PC} = \mathbf{a}^{\top} \Psi \quad (3.9)$$

235 where $\mathbf{a} = (a_1, \dots, a_P)$ and $\Psi = (\Psi_1, \dots, \Psi_P)$. The PCE coefficients \mathbf{a} can be computed by using *non-intrusive*
 236 methods, from point-wise model evaluations. This refers to projection methods [62, 63], stochastic collocation
 237 [64] or regression [65, 66, 67]. In the context of regression approaches, Ordinary Least Squares (OLS) are usually
 238 considered [65, 66]. Given an experimental design (ED) $\mathcal{X} = (\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(K)})$ and the corresponding model
 239 outputs $\mathcal{Y} = (\mathcal{M}(\boldsymbol{\theta}^{(1)}), \dots, \mathcal{M}(\boldsymbol{\theta}^{(K)}))$ such approaches aim at solving the following minimization problem:

$$\mathbf{a} = \arg \min_{\mathbf{a}^* \in \mathbb{R}^P} \|\mathcal{Y} - \mathbf{A} \mathbf{a}^*\|^2 \quad (3.10)$$

240 where \mathbf{A} is the $K \times P$ matrix with coefficients $\mathbf{A}_{ij} = \Psi_j(\boldsymbol{\theta}^{(i)})$, and $\|\cdot\|$ is the Euclidean norm of \mathbb{R}^P , *i.e.*
 241 $\|\mathbf{x}\|^2 = \sum_{i=1}^P x_i^2$ for $\mathbf{x} = (x_1, \dots, x_P) \in \mathbb{R}^P$. In most cases, PCE representations are *compressible*, in the sense
 242 that only few terms are significantly contributing to the variance of the model response. Then, compressive
 243 sensing [68] techniques have been recently used in the PCE framework in order to construct sparse PCE with
 244 ED of limited sizes. **One refers to [69] for an recent in-depth review and benchmark on sparse PCE. Finally, once**
 245 **PCE coefficients are computed, the *generalization error* of the truncated PCE defined by $\mathbb{E}[(\mathcal{M}(\Theta) - \widehat{\mathcal{M}}(\Theta))^2]$**
 246 **may be estimated by using cross-validation techniques [70]. In this context, the Leave One Out (LOO) error is**
 247 **commonly used as an estimate of the PCE generalization error (see [67] for further details).**

248 3.2. Polynomial Chaos-based Kriging

249 Polynomial Chaos Kriging (PCK) [42, 43] consists in a Kriging [71] whose trend is given by a PCE:

$$\mathcal{M}(\boldsymbol{\theta}) \approx \widehat{\mathcal{M}}^{PCK}(\boldsymbol{\theta}) = \sum_{\alpha \in \mathcal{A}} a_{\alpha} \Psi_{\alpha}(\boldsymbol{\theta}) + \sigma^2 Z(\boldsymbol{\theta}) \quad (3.11)$$

250 where $\sum_{\alpha \in \mathcal{A}} a_{\alpha} \Psi_{\alpha}(\boldsymbol{\theta})$ is a truncated PCE, σ^2 the PCK variance and Z a zero-mean stationary Gaussian
 251 process with a correlation kernel $R_Z(\boldsymbol{\theta}, \boldsymbol{\theta}'; \boldsymbol{\eta})$.

252 The interest of such formulation is twofold: the trend provides a *global* approximation of the model response
 253 on the support $\mathcal{D}_{\boldsymbol{\theta}}$ of input parameters, whereas the Gaussian process enables to catch the *local* variability of the
 254 model response. Furthermore, the prediction variance associated to the Gaussian process provides a local error
 255 measure for PCK model predictions, which can be used as an indicator of the sparsity of the ED. Indeed, Kriging
 256 is an exact interpolating method, in the sense that $\mu_{\hat{y}}(\boldsymbol{\theta}) = \mathcal{M}(\boldsymbol{\theta})$ and $\sigma_{\hat{y}}(\boldsymbol{\theta}) = 0$, $\forall \boldsymbol{\theta} \in \mathcal{X}$, **where $\mu_{\hat{y}}(\boldsymbol{\theta})$ and**
 257 **$\sigma_{\hat{y}}(\boldsymbol{\theta})$ are respectively the Kriging prediction mean and standard deviation at the point $\boldsymbol{\theta}$.** Thus, the prediction
 258 variance is low near ED points and becomes important in unexplored zones of the parameter space. Such local
 259 indicator has been exploited in the framework of adaptive surrogate modeling [39, 43], in order to enrich EDs.
 260 Note also that a global error measure consisting in a formulation of LOO error is available for Kriging surrogates
 261 [72].

262 Moreover, it is worth noting that global surrogates such as PCE do not dispose of local error measures as
 263 simple and direct as Kriging prediction variance, but rather global error measures, such as the LOO error [67].
 264 The PCK formulation then provides a simple way to assess the local error of PCE predictions. It is also worth
 265 noting that recent work of Sudret & Marelli [73] provided a local error estimate for PCE surrogates, based on
 266 bootstrap resampling.

267 The setup of a PCK surrogate model requires the calculation of the PC coefficients $\mathbf{a} = (a_{\alpha})_{\alpha \in \mathcal{T}}$ as well
 268 as Kriging variance σ^2 and correlation hyper-parameters $\boldsymbol{\eta}$. In this context, Schöbi et al [42] propose two
 269 computational approaches, namely Sequential PCK (S-PCK) and Optimal PCK (O-PCK). The first approach
 270 consist in constructing a PCE surrogate by using the LAR algorithm [67], and then using this PCE as the trend
 271 of a universal Kriging model, whose parameters are calibrated **through Maximum Likelihood Estimation (MLE)**
 272 **[74]**. The second approach aims at iteratively constructing a PCK by adding polynomials one-by-one to the
 273 trend, and selecting the surrogate which minimizes a LOO error estimate. For further details, the reader is
 274 referred to [42].

275 4. Bayesian updating using adaptive PCK within subset simulation

276 4.1. General framework

277 The proposed Bayesian updating methodology aims at constructing a PCK surrogate which is enriched in
 278 an adaptive way throughout the SuS procedure described by Algorithm 1. Keeping the notations introduced
 279 in the previous sections, an initial ED $\mathcal{X}^{(0)} = (\boldsymbol{\theta}^1, \dots, \boldsymbol{\theta}^{K_0})$ of size K_0 is defined. Then, the log-likelihood
 280 $L(\boldsymbol{\theta}) = \log(\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}))$ in Eq. (2.18) is approximated by a PCK surrogate by using the ED $\mathcal{X}^{(0)}$:

$$L(\boldsymbol{\theta}) \approx \hat{L}(\boldsymbol{\theta}) = \sum_{\alpha \in \mathcal{T}} a_{\alpha} \Psi_{\alpha}(\boldsymbol{\theta}) + \sigma^2 Z(\boldsymbol{\theta}) \quad (4.1)$$

281 where the polynomials $(\Psi_{\alpha})_{\alpha \in \mathcal{T}}$ are chosen to be orthonormal with respect to the prior density $\rho(\theta)$.

282 This surrogate model aims at approximating the global behavior of the log-likelihood on the support of the
 283 prior distribution. The PCE trend is built by using the LAR-based procedure introduced in [67], in order to
 284 provide sparse PCE which can be trained with an ED of small size (typically with $K_0 \sim 10^1 - 10^2$). Substituting
 285 the full log-likelihood by its surrogate in the modified LSF expression given by Eq. (2.18) yields a surrogate
 286 LSF, given by:

$$\widehat{h}(\theta, p) = \log(p) - \log(c) - \widehat{L}(\theta) \quad (4.2)$$

287 Firstly, given a prescribed number N of samples per subset, the Monte Carlo Simulation (MCS) step of SuS is
 288 performed, by using the surrogate LSF instead of the full one. In this way, samples $\{(\theta^{(0,k)}, p^{(0,k)})\}_{1 \leq k \leq K}$ are
 289 generated, and the corresponding log-likelihood values $\{\widehat{L}(\theta^{(0,k)})\}_{1 \leq k \leq K}$ are computed. Then, a critical step
 290 is the calculation of an initial value of the BUS normalizing constant c . In the adaptive BUS-SuS approach
 291 introduced in [30], this value is set as the largest log-likelihood of the generated samples. In the framework
 292 of the proposed approach, a modification of this approach is required: indeed, samples drawn through Monte
 293 Carlo sampling which are too far away from the ED $\mathcal{X}^{(0)}$ may lead to inaccurate PCK predictions. This implies
 294 to potentially sensibly underestimate the optimal value $c_{opt} = \max \mathcal{L}$, or conversely, to strongly overestimate it
 295 and altering the efficiency of the SuS procedure, as mentioned in Section 2.3.3. Consequently, we propose to set
 296 the initial value of c as the maximum likelihood on a subset of $\mathcal{S}^{(0)} = \{\theta^{(0,k)}\}_{1 \leq k \leq K}$, formed by points which
 297 are not too far away from $\mathcal{X}^{(0)}$. At this point, the question is to define a criterion for filtering points whose
 298 predictions are deemed to be not accurate.

299 A quite natural choice proposed in [37] or [75] is based on the convex hull of the points of the ED. The convex
 300 hull of $\mathcal{X}^{(0)}$, denoted by $\text{Conv}(\mathcal{X}^{(0)})$ is the smallest convex set which contains the points of $\mathcal{X}^{(0)}$. If the ED
 301 points are sufficiently dense, such set may be seen as the "representativeness domain" of the trained surrogate
 302 model. Then, the initial constant c may be set as the maximum surrogate likelihood on this set:

$$c = \max_{\theta \in \mathcal{K}^{(0)}} \exp(\widehat{L}(\theta)) \quad (4.3)$$

303 where $\mathcal{K}^{(0)} = \mathcal{S}^{(0)} \cap \text{Conv}(\mathcal{X}^{(0)})$. The Quickhull algorithm [76] will be used for computing the convex hull, as in
 304 [75, 37].

305 Next, at SuS level $i \geq 1$, the surrogate LSF \widehat{h}_i is used for setting the threshold value h_i (see Algorithm 1).
 306 Subsequently, n Markov chain seeds $\{(\theta^{(i-1,k)}, p^{(i-1,k)})\}_{1 \leq k \leq n}$ are selected according to surrogate LSF values.
 307 We propose to enrich the ED $\mathcal{X}^{(i-1)}$ of the surrogate log-likelihood \widehat{L} , by adding $K_t \geq 1$ new points, based on a
 308 point enrichment criterion, which will be described in Section 4.2. A new PCK is then trained on the enriched
 309 ED $\mathcal{X}^{(i)}$. The adaptive Conditional Sampling [49] procedure is then launched by using the surrogate LSF, in
 310 order to generate the samples $\{(\theta^{(i,k)}, p^{(i,k)})\}_{1 \leq k \leq K}$ of the current subset. Then, the value of the constant c is
 311 updated at the end of the SuS stage i . For the same reasons underlined for the MCS step, this constant is the
 312 maximum likelihood on the set defined by $\mathcal{K}^{(i)} = \text{Conv}(\mathcal{X}^{(i)}) \cap \{\theta^{(i,k)}\}_{1 \leq k \leq K}$, *i.e.* as $c^* = \max(c, c_i)$, where

313 $c_i = \max_{\boldsymbol{\theta} \in \mathcal{K}^{(i)}} \exp(\widehat{L}(\boldsymbol{\theta}))$. Such update is *a priori* modifying the definition of the current subset F_i (see Eq.
314 (2.17)) since the latter explicitly depends on the scaling constant. However, as derived in [30], updating the
315 scaling constant c does not affect the distribution of the samples of F_i if its corresponding threshold value h_i is
316 adjusted as follows:

$$h_i^* = h_i + \log \frac{c}{c^*} \quad (4.4)$$

317 Moreover, we also follow the procedure of [30] aiming at decreasing the dependence of the samples of each subset.
318 Knowing that for a fixed $\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\theta}}$, any $p \in \mathcal{D}_p$ which satisfies $\log(p) \leq \log(c) - L(\boldsymbol{\theta}) + h_i$ lies in the subset F_i ,
319 the p component of a sample $(\boldsymbol{\theta}, p) \in F_i$ may be resampled by drawing $p^* \sim \mathcal{U}([0, \min(1, c \exp(h_i) \mathcal{L}(\boldsymbol{\theta}))])$. This
320 enables to increase the performance of the BUS procedure, since the rejection of a sample during the MCMC
321 sampling phase implies to duplicate an existing sample.

322 Then, the SuS procedure is continued until reaching the last subset, *i.e.* until $h_i \leq 0$.

323 4.2. Point enrichment step

324 In the framework of active learning structural reliability methods [77, 39, 78], *learning functions* are used in
325 order to measure the information gain related to quantities of interest provided by adding a new point to the
326 ED. In this context, a wide range of functions has been developed, including the *Expected Feasibility Function*
327 (EFF) [77], the so-called *U-function* [39], the *Fraction of Bootstrap Replicates* (FBR) [73] or the *constrained*
328 *min-max function* [78].

329 The *U-function* introduced in [39] is particularly well suited for Kriging-based approaches, by focusing on
330 the accuracy on the sign of the surrogate LSF: sample points with non-positive LSF values are classified as
331 failure points, which traduces in the BUS framework by points which are distributed according to the posterior
332 distribution. Therefore, the potentially important uncertainties on predictions at points which are close to the
333 Limit State Surface (LSS) given by $\partial \mathcal{D}_j = \{(\boldsymbol{\theta}, p) \in \mathcal{D}_{\boldsymbol{\theta}} \times \mathcal{D}_p \mid h_i(\boldsymbol{\theta}, p) = 0\}$ can cause them to change sign, which
334 may lead to misclassification. The *U-function* aims at selecting points in zones where points are announced to
335 be close to the LSS, and/or where Kriging prediction variance is important [39]. This function is given by:

$$U(\boldsymbol{\theta}, p) = \frac{|\mu_{\widehat{h}}(\boldsymbol{\theta}, p)|}{\sigma_{\widehat{h}}(\boldsymbol{\theta}, p)} \quad (4.5)$$

336 where $\mu_{\widehat{h}}(\boldsymbol{\theta}, p)$ and $\sigma_{\widehat{h}}(\boldsymbol{\theta}, p)$ are, in turn, the prediction mean and variance of the surrogate LSF given by Eq.
337 (4.2). Note that, for a fixed value of c and $p \in \mathcal{D}_p$, $\widehat{h}(\boldsymbol{\theta}, p)$ is a Gaussian random variable:

$$\widehat{h}(\boldsymbol{\theta}, p) \sim \mathcal{N}\left(\log(p) - \log(c) - \mu_{\widehat{L}}(\boldsymbol{\theta}), \sigma_{\widehat{L}}^2(\boldsymbol{\theta})\right) \quad (4.6)$$

338 where $\mu_{\widehat{L}}(\boldsymbol{\theta})$ and $\sigma_{\widehat{L}}(\boldsymbol{\theta})$ are the prediction mean and variance of the surrogate log-likelihood (4.1). Therefore,

339 in the framework of the proposed approach, the U -function (4.5) rewrites:

$$U(\boldsymbol{\theta}, p) = \frac{|\log(p) - \log(c) - \mu_{\hat{L}}(\boldsymbol{\theta})|}{\sigma_{\hat{L}}(\boldsymbol{\theta})} \quad (4.7)$$

340 The *misclassification probability* of a point $(\boldsymbol{\theta}, p) \in \mathcal{D}_{\boldsymbol{\theta}} \times \mathcal{D}_p$ is defined by the probability of having $\mu_{\hat{L}}(\boldsymbol{\theta}, p) > 0$
 341 and $h(\boldsymbol{\theta}, p) \leq 0$ or *vice-versa*. This probability may be expressed as follows [79, 43]:

$$\mathbb{P}_m(\boldsymbol{\theta}, p) = \Phi(-U(\boldsymbol{\theta}, p)) = \Phi\left(-\frac{|\log(p) - \log(c) - \mu_{\hat{L}}(\boldsymbol{\theta})|}{\sigma_{\hat{L}}(\boldsymbol{\theta})}\right) \quad (4.8)$$

342 Such probability is bounded between 0 and 0.5. The case $\mathbb{P}_m(\boldsymbol{\theta}, p) \rightarrow 0$ corresponds to points whose prediction
 343 variance is low, and/or points which are far away from the LSS. Conversely, the case $\mathbb{P}_m(\boldsymbol{\theta}, p) \rightarrow 0.5$ corresponds
 344 to “dangerous” points, which are close to the LSS and/or whose prediction variance is important.

345 In the case where only one enrichment point is needed, the latter may be selected as the point which
 346 maximizes the misclassification probability (4.8). In the case of multiple point enrichment, clustering techniques
 347 [45] applied in the framework of adaptive surrogate modeling for SR analysis [43, 41, 80] enable to efficiently
 348 selecting new points. In this paper, the weighted k -means algorithm [45] is considered.

349 Recall that at SuS stage $i \geq 1$, we dispose of n Markov chain seeds $\left\{(\boldsymbol{\theta}^{(i-1,k)}, p^{(i-1,k)})\right\}_{1 \leq k \leq n}$. As proposed
 350 in [43, 41], the weights considered for clustering are set as the corresponding misclassification probabilities
 351 $\left\{\mathbb{P}_m(\boldsymbol{\theta}^{(i-1,k)}, p^{(i-1,k)})\right\}_{1 \leq k \leq n}$. However, the surrogate log-likelihood $\hat{L}(\boldsymbol{\theta})$ is defined on the physical parameter
 352 space $\mathcal{D}_{\boldsymbol{\theta}}$, and then does not depend on values of $p \in \mathcal{D}_p$. Hence, performing a clustering in the augmented
 353 space $\mathcal{D}_{\boldsymbol{\theta}} \times \mathcal{D}_p$ may lead to redundant enrichment points, since the obtained clusters may lie on a line of
 354 the form $\{\boldsymbol{\theta}\} \times \mathcal{D}_p$. Consequently, we propose to perform weighted k -means clustering on the candidate set
 355 given by $\left\{\boldsymbol{\theta}^{(i-1,k)}\right\}_{1 \leq k \leq n}$ by using the same misclassification probabilities $\left\{\mathbb{P}_m(\boldsymbol{\theta}^{(i-1,k)}, p^{(i-1,k)})\right\}_{1 \leq k \leq n}$ as
 356 weights. This is tantamount to projecting the seeds onto the parameter space $\mathcal{D}_{\boldsymbol{\theta}}$ by using the canonical
 357 projection $\mathcal{D}_{\boldsymbol{\theta}} \times \mathcal{D}_p \rightarrow \mathcal{D}_{\boldsymbol{\theta}}$. In this way, the centroids $\left\{\boldsymbol{\theta}_*^{(j)}\right\}_{1 \leq j \leq K_t}$ of the K_t obtained clusters are selected
 358 in order to enrich the ED of the surrogate log-likelihood $\hat{L}(\boldsymbol{\theta})$. The updated ED of $\hat{L}(\boldsymbol{\theta})$ is then given by
 359 $\mathcal{X}^{(i)} = \mathcal{X}^{(i-1)} \cup \left\{\boldsymbol{\theta}_*^{(j)}\right\}_{1 \leq j \leq K_t}$.

360 4.3. Stopping criterion

361 The stopping criterion of the surrogate model learning phase may be defined based on the variability of
 362 estimated failure probabilities [74], the stability of the estimated failure domain [43, 78] or learning functions
 363 [39, 73]. In this paper, we use the stopping criterion based on the U -function introduced in [39]. At SuS stage
 364 i , this criterion writes:

$$\min_{(\boldsymbol{\theta}, p) \in \mathcal{S}_0^{(i)}} U(\boldsymbol{\theta}, p) \geq \epsilon_U \quad (4.9)$$

365 where $\mathcal{S}_0^{(i)}$ is the set formed by the n Markov chain seeds selected for the i -th subset (see Algorithm 1), and ϵ_U is
 366 threshold value. Following [39], we choose the value $\epsilon_U = 2$, which has provided satisfactory accuracy in several
 367 application examples. Choosing this value is tantamount to impose a maximal misclassification probability
 368 of $\Phi(-2) \approx 0.023$. Notwithstanding its widespread use, it should be underlined that such value may be too
 369 conservative in some cases, for complex LSF functions and/or for high dimensionalities [41]. However, the
 370 search for an optimal stopping criterion is out of the scope of this paper, and the reader is referred to [41] for
 371 further details about recent improvement strategies.

372 4.4. Summary of the proposed method

373 The proposed approach, which will be named BUS-PCK in the following sections, is summarized hereafter:

- 374 1. An initial ED $\mathcal{X}^{(0)}$ of size K_0 is selected, in order to build a PCK surrogate $\widehat{L} \approx L$ of the log-likelihood
 375 (Eq. (4.1)). This induces the definition of a surrogate LSF \widehat{h} (see Eq. (4.2)).
- 376 2. Level 0 of SuS procedure (MCS): K samples are drawn from the prior distribution, and the corresponding
 377 surrogate LSF values are computed.
- 378 3. Initialize BUS normalizing constant c from the generated sample according to Eq. (4.3). This implies the
 379 computation of the convex hull of MCS samples with the Quickhull algorithm [76].
- 380 4. At SuS level $i \geq 1$:
 - 381 a. The threshold value h_i of the current subset F_i is computed as the \mathbb{P}_t -percentile of the surrogate LSF
 382 values on the samples of level $i - 1$ (with $\mathbb{P}_t = 10\%$, following [29, 30]). The intermediate probability
 383 $\mathbb{P}_i = \mathbb{P}(F_i|F_{i-1})$ is then computed (see Algorithm 1).
 - 384 b. Learning phase of the PCK surrogate (see Section 4.2): K_t points are added to the ED $\mathcal{X}^{(i-1)}$ at
 385 previous SuS level, by using the seeds (*i.e.* the $\mathbb{P}_i K$ samples which are lying in F_i) as candidate
 386 points. If the stopping condition of the learning phase is met (see Section 4.3), the PCK surrogate
 387 model is deemed to be sufficiently accurate near posterior high-probability zones.
 - 388 c. Generation of conditional samples: the population of the next SuS subset is generated by using the
 389 aCS MCMC algorithm [49], from the previously identified seeds.
 - 390 d. Update the BUS normalizing constant c and the threshold value h_i according to Section 4.1.
 - 391 e. Convergence of SuS procedure: if $h_i \leq 0$, the SuS procedure is stopped, and the failure probability \mathbb{P}_f
 392 is computed as described in Algorithm 1. The samples of the last SuS level correspond to posterior
 393 samples.
- 394 5. Set $r = i$ and compute the failure probability $\mathbb{P}_f = \prod_{j=1}^r$. Posterior samples are given by the projection
 395 on \mathcal{D}_θ of the K samples which lie in the subset $F = F_r$. The model evidence is computed as $\mathcal{Z} = \mathbb{P}_f/c$.

396 The flowchart of the proposed approach is shown in Fig. 1.

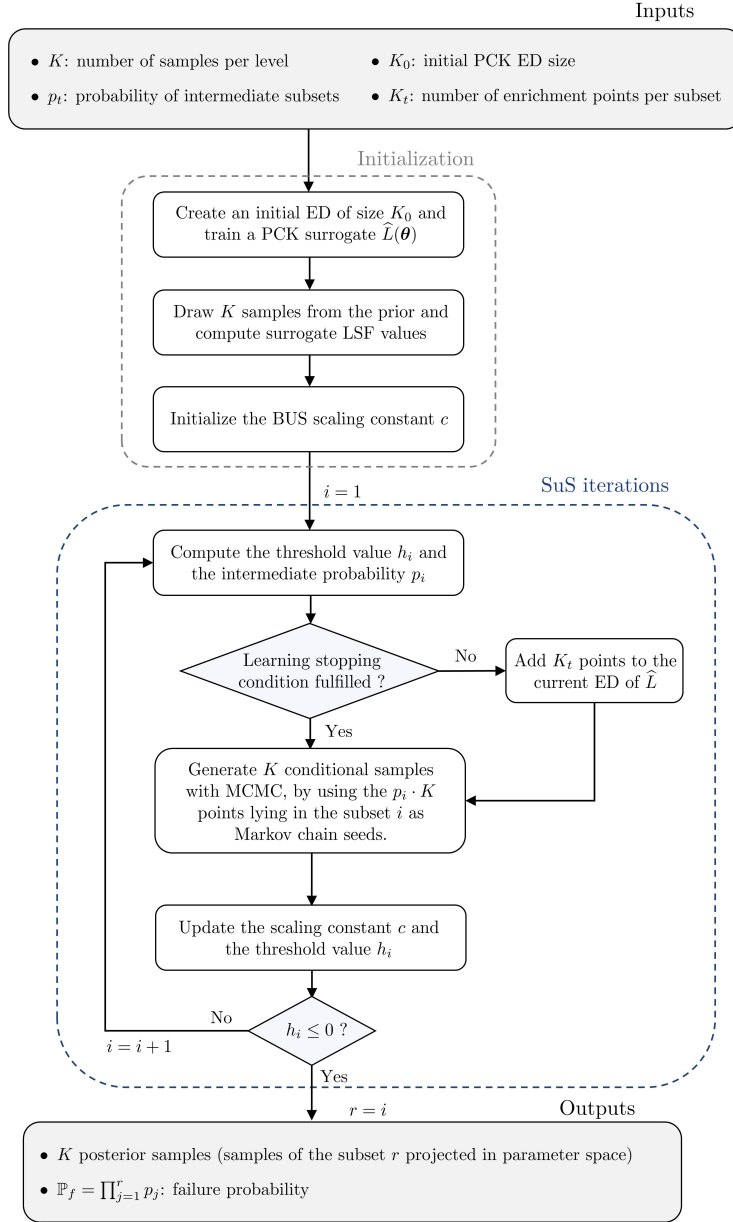


Figure 1: Flowchart of the proposed approach.

397 5. Case studies

398 In this section, three case studies of increasing complexity are presented for assessing the performance of
 399 the proposed Bayesian updating approach, which will named BUS-PCK in the following. The PCK surrogates
 400 will be constructed by using the Optimal PC-Kriging (O-PCK) approach described in [42]. The PCE trends
 401 $\sum_{\alpha \in \mathcal{T}} a_{\alpha} \Psi_{\alpha}$ will be constructed by using the classical isotropic truncation sets given by:

$$\mathcal{T}_q = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^m \mid \|\boldsymbol{\alpha}\|_1 = \sum_{i=1}^m \alpha_i \leq q \right\} \quad (5.1)$$

402 where $q \leq 10$ is the PCE degree. Sparse PCE are constructed with these truncation sets by using the LAR-
 403 based approach introduced in [67]. Moreover, we consider the Matérn kernel with shape parameter $\nu = \frac{5}{2}$ for all
 404 Kriging Gaussian processes. This kernel is given by:

$$R(\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|; \boldsymbol{\eta}) = \prod_{i=1}^m \frac{1}{2^{\nu-1} \Gamma(\nu)} \left(\sqrt{2\nu} \frac{|\theta_i - \theta'_i|}{\eta_i} \right)^{\nu} \mathcal{K}_{\nu} \left(\sqrt{2\nu} \frac{|\theta_i - \theta'_i|}{\eta_i} \right) \quad (5.2)$$

405 where Γ is the Euler's Gamma function and \mathcal{K}_{ν} is the modified Bessel function of the second kind. In this
 406 context, we use the software UQLab [81] which provides an implementation of PCK.

407 The PCK approximation $\widehat{L}(\boldsymbol{\theta})$ of the log-likelihood $L(\boldsymbol{\theta})$ induces the definition of a surrogate posterior PDF:

$$\widehat{\pi}(\boldsymbol{\theta}) = \frac{\rho(\boldsymbol{\theta}) \exp(\widehat{L}(\boldsymbol{\theta}))}{\widehat{\mathcal{Z}}} \quad (5.3)$$

408 where $\widehat{\mathcal{Z}}$ is the surrogate model evidence:

$$\widehat{\mathcal{Z}} = \int_{\mathcal{D}_{\boldsymbol{\theta}}} \rho(\boldsymbol{\theta}) \exp(\widehat{L}(\boldsymbol{\theta})) d\boldsymbol{\theta} \quad (5.4)$$

410 In order to assess the accuracy of the approximation of the posterior PDF, measures taken from information
 411 theory are often considered. Indeed, the *Kullback-Leibler Divergence* (KLD) [82] is usually used for measuring
 412 the similarity between the surrogate and the true posterior PDF [32], and is defined by:

$$\mathcal{D}_{KL}(\widehat{\pi}||\pi) = \int_{\mathcal{D}_{\boldsymbol{\theta}}} \widehat{\pi}(\boldsymbol{\theta}) \log \left(\frac{\widehat{\pi}(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} \right) d\boldsymbol{\theta} \quad (5.5)$$

413 This measure does not define a distance in the mathematical sense, especially due to the fact that it is not
 414 symmetric, *i.e.* $\mathcal{D}_{KL}(\widehat{\pi}||\pi) \neq \mathcal{D}_{KL}(\pi||\widehat{\pi})$. A symmetrized and somewhat regularized version of KLD is given by
 415 the *Jensen-Shannon Divergence* (JSD), defined by [83]:

$$\mathcal{D}_{JS}(\widehat{\pi}||\pi) = \frac{1}{2} (\mathcal{D}_{KL}(\widehat{\pi}||\tilde{\pi}) + \mathcal{D}_{KL}(\pi||\tilde{\pi})) \quad (5.6)$$

416 where $\tilde{\pi} = \frac{1}{2} (\widehat{\pi} + \pi)$. In this paper, we will use the *average Jensen-Shannon Divergence* (aJSD) introduced in
 417 [23], given by:

$$\overline{\mathcal{D}_{JS}}(\widehat{\pi}||\pi) = \frac{1}{m} \sum_{i=1}^m \mathcal{D}_{JS}(\widehat{\pi}_i||\pi_i) \quad (5.7)$$

418 where $\widehat{\pi}_i$ and π_i are the i -th surrogate and true posterior marginal PDF. This measure enables to summarize
 419 the accuracy of the approximation of the posterior PDF in a single scalar, which is computationally affordable
 420 since it is based on an average of univariate integrals.

421 In the following sections, one will consider $K = 5000$ samples per subset, based on [26, 30]. Indeed, in the
 422 framework of the numerical examples presented hereafter, such sample size has been observed to be sufficient for
 423 exploring the posterior distribution with the SuS procedure [26, 30]. In the general case, when combining the SuS

424 procedure with surrogate models, it might be possible to consider an “overkill” sample size (*i.e.* $K \sim 10^5 - 10^6$)
 425 [84], in order to ensure that all posterior high-probability regions are discovered. The intermediate probability
 426 \mathbb{P}_t is chosen as $\mathbb{P}_t = 10\%$, based on [30]. The SuS procedure described in Section 2.3 is first applied by using
 427 the full model, and the corresponding results will be considered as a reference. Regarding to the proposed
 428 approach, varying initial ED sizes K_0 and ratios $\kappa = K_t/K_0$ will be considered. Moreover, 50 replications of the
 429 calculations will be produced.

430 5.1. One-dimensional analytical function

431 5.1.1. Problem description

432 The first case study focuses on a one-dimensional analytical function, which presents a combination of a
 433 global trend and a localized behavior:

$$\mathcal{M}(\theta) = 1 + \cos\left(\frac{\theta}{2}\right) + 3 \exp(-4(\theta - 2)^2) \quad (5.8)$$

434 The first two terms may be well approximated by a low-order PCE, whereas the third term induces a peak which
 435 requires a sensibly high PCE degree for providing an accurate approximation. Hence, such behavior would make
 436 classical PCE-based Bayesian inversions fail, since global approximations are not sufficient here to accurately
 437 mimic the model near the high-likelihood zone (around $\theta = 2$). One considers a random variable Θ with prior
 438 $\mathcal{N}(0, 1)$, an observation data $y = \mathcal{M}(2) = 4.54$ and an additive Gaussian discrepancy model with noise variance
 439 $\sigma = 0.4$. This leads to the following likelihood function:

$$\mathcal{L}(\theta; y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\mathcal{M}(\theta) - y)^2\right) \quad (5.9)$$

440 5.1.2. Posterior distribution

441 First, the proposed methodology is illustrated in Fig.2, for the case $K_0 = 20$ and $\kappa = 0.2$. At level 0 of SuS
 442 procedure, a PCK surrogate model of the log-likelihood is built on the initial ED $\mathcal{X}^{(0)}$ of size K_0 . The points
 443 of $\mathcal{X}^{(0)}$ are sensibly dense around the prior mean $\theta = 0$, whereas the log-likelihood peak is located near $\theta \approx 2$.
 444 As shown in Fig.2a, the predictions of the PCK surrogate for $\theta \geq 2$ are sensibly uncertain and overestimate the
 445 log-likelihood, which leads to indicate a wrong location of the posterior high-probability zone. Then, at SuS level
 446 1, the point enrichment step described in Section 4.2 enables to select $K_t = \kappa K_0$ new points $\mathcal{X}_{new} = \{\theta_*^{(i)}\}_{1 \leq i \leq 4}$
 447 to add to the ED, based on the misclassification probability given in Eq.(4.8). These points locate in the vicinity
 448 of the log-likelihood peak, near $\theta \approx 2$, as shown by the white diamonds in Fig. 2. Then, the PCK surrogate
 449 model is updated, by using the ED $\mathcal{X}^{(1)} = \mathcal{X}^{(0)} \cup \mathcal{X}_{new}$. The predictions of the updated surrogate model are
 450 sensibly more accurate near the log-likelihood peak, as underlined by Fig.2b.

451 Then, the convergence of JSD with respect to the effective number of model calls denoted by $\mathbb{E}[N_c]$ is shown
 452 in Fig.3. For a single run, the number of model calls is given by $N_c = K_0(1 + n_S \kappa)$, where n_S is the number

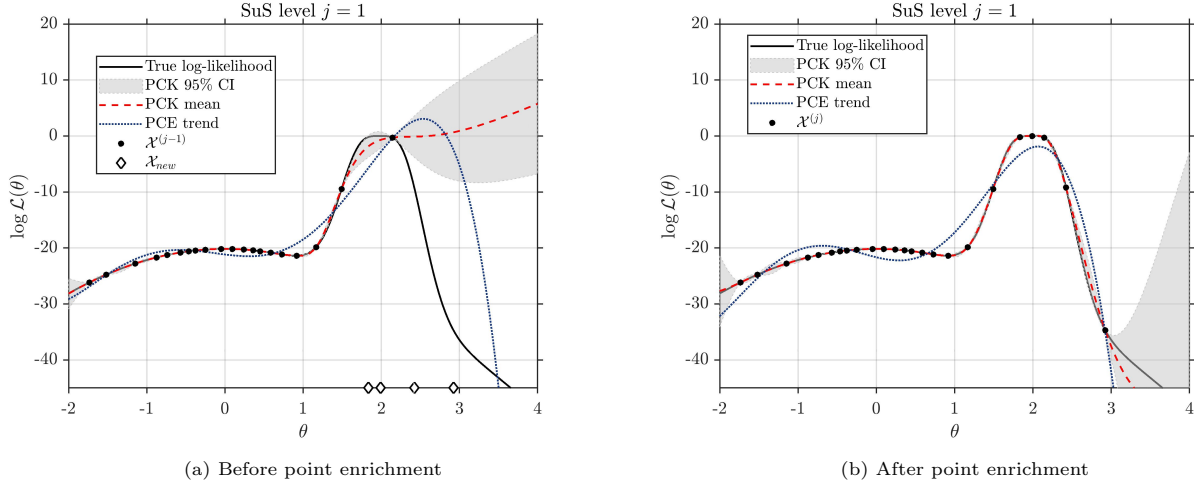


Figure 2: 1D analytical function: illustration of the proposed method at SuS level 1

453 of subsets required for stopping the SuS procedure. For moderate ED sizes $K_0 \approx 50$, the proposed approach
 454 is more efficient than its non-adaptive counterpart ($\kappa = 0$), since the JSD for $\kappa \geq 0.1$ is at least one order of
 455 magnitude smaller than that for $\kappa = 0$. In the case of larger EDs, JSD values for $\kappa \in \{0, 0.1, 0.2\}$ are closer,
 456 which can be explained by the fact that more informative points (*i.e.* with high likelihood) are drawn during
 457 the initial sampling step aiming at constructing the initial ED of the surrogate model. For a fixed amount of
 458 model calls, the JSD decreases when κ increases, though. This underlines the efficiency of the point enrichment
 459 step, which enables to better focus on high-likelihood zones for enriching the surrogate model.

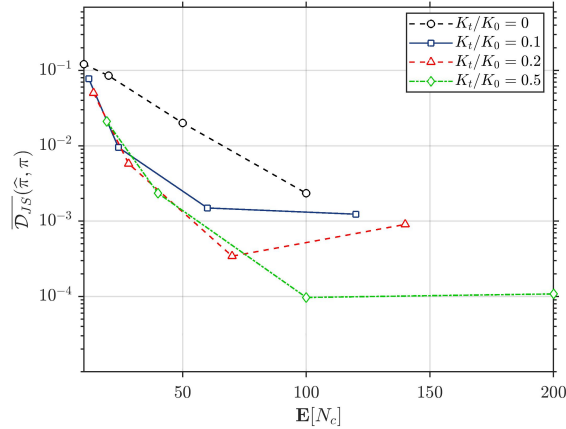


Figure 3: 1D analytical function: convergence of JSD with respect to the average number of model calls $\mathbb{E}[N_c]$.

460 The posterior PDF obtained with the reference calculations and the proposed approach are shown in Fig. 4.
 461 Visual inspection of PDF confirms the analysis of the convergence of JSD, since the reference posterior PDF is
 462 well approximated by its surrogate counterpart as the experimental size increases. Lastly, several QoI obtained
 463 from calculations, namely the effective number of model calls $\mathbb{E}[N_c]$, the posterior mean $\mathbb{E}[\theta|y]$, the standard
 464 deviation $\sqrt{\mathbf{Var}[\theta|y]}$, and the model evidence \mathcal{Z} are summarized in Tab. 1. The corresponding relative errors
 465 $\epsilon = |\chi - \chi_{ref}|/\chi_{ref}$ are given in brackets, in %. The results emphasize the efficiency of the BUS-PCK approach,
 466 since a satisfactory approximation of the posterior PDF is provided for $\mathbb{E}[N_c] = 100$ model calls, which represents

467 a reduction of 99.3% of the full cost of the SuS procedure described in Section 2.3, namely $1.5 \cdot 10^4$ full model
 468 calls.

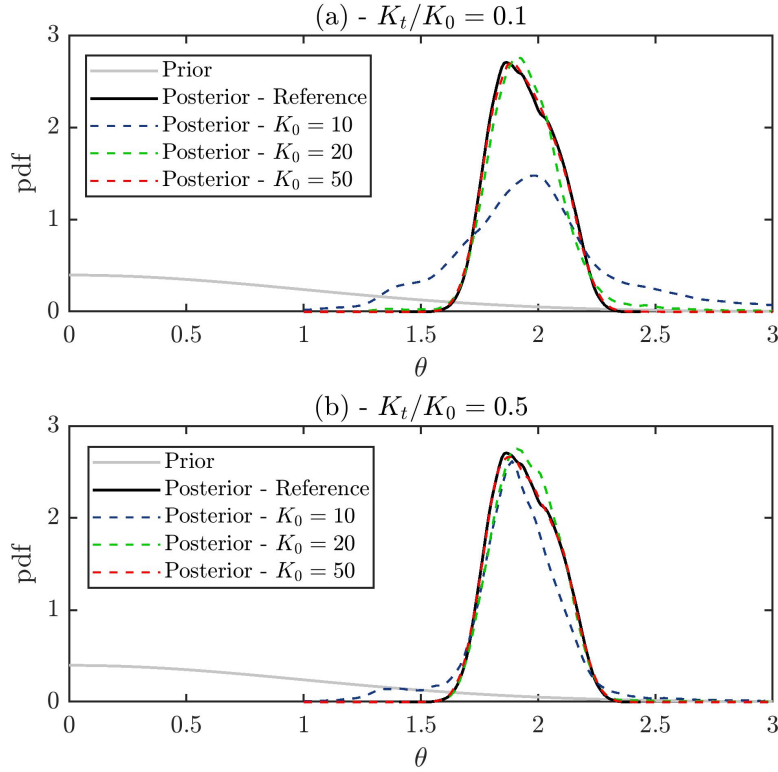


Figure 4: 1D analytical function: prior and posterior PDF: (a) $K_t/K_0 = 0.1$; (b) $K_t/K_0 = 0.5$.

	K_0	K_t/K_0	$\mathbb{E}[N_c]$	$\mathbb{E}[\theta y]$	$\sqrt{\mathbf{Var}[\theta y]}$	\mathcal{Z}
BUS-PCK	20	0.1	24	1.938 (0.2)	0.213 (59.3)	0.028 (16.1)
	20	0.5	40	1.954 (0.6)	0.147 (10.1)	0.037 (53.3)
	50	0.1	60	1.944 (0.1)	0.168 (25.4)	0.025 (4.2)
	50	0.5	100	1.942 (0.0)	0.138 (3.4)	0.025 (3.4)
	100	0.1	120	1.941 (0.0)	0.148 (10.5)	0.024 (0.1)
	100	0.5	200	1.942 (0.0)	0.134 (0.4)	0.024 (0.6)
Reference			15000	1.942	0.134	0.024

Table 1: 1D analytical function: posterior QoI, with relative errors in %.

469 5.2. Two degrees-of-freedom shear building

470 5.2.1. Problem description

471 The second case study was originally introduced in [85], and then studied in [30, 37, 25, 38], and involves
 472 a two degrees-of-freedom shear building, shown in Fig. 5. The problem consists in estimating the posterior
 473 inter-story stiffnesses of the structure, based on measurements of the eigen-frequencies of the structure.

474 The inter-story stiffnesses are set as $k_i = \Theta_i k_n$ for $i \in \{1, 2\}$, where $k_n = 29.7 \cdot 10^6 \text{ N.m}^{-1}$ and Θ_1, Θ_2

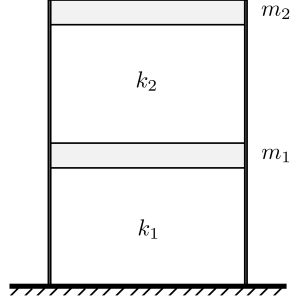


Figure 5: Two degrees-of-freedom shear building

475 are random variables endowed with lognormal priors with modes 1.3 and 0.8, and standard deviations equal to
 476 1. The masses of the two stories are supposed to be deterministic, and are set as $m_1 = 16.531 \cdot 10^3$ kg and
 477 $m_2 = 16.131 \cdot 10^3$ kg. Two measured eigen-frequencies $\tilde{\mathbf{f}} = (\tilde{f}_1, \tilde{f}_2)$ will be used for Bayesian updating, with
 478 $\tilde{f}_1 = 3.13$ Hz and $\tilde{f}_2 = 9.83$ Hz. The corresponding likelihood function writes:

$$\mathcal{L}(\boldsymbol{\theta}; \tilde{\mathbf{f}}) = \exp\left(-\frac{J(\boldsymbol{\theta})}{2\sigma^2}\right) \quad (5.10)$$

479 where $\sigma = 1/16$ and $J(\boldsymbol{\theta})$ is the modal measure-of-fit function given by [85]:

$$J(\boldsymbol{\theta}) = \sum_{i=1}^2 \mu_i^2 \left(\frac{f_i(\boldsymbol{\theta})^2}{\tilde{f}_i^2} - 1 \right)^2 \quad (5.11)$$

480 with $\mu_1 = \mu_2 = 1$, and $(f_i(\boldsymbol{\theta}))_{i \in \{1,2\}}$ are the eigen-frequencies obtained by solving the equation of motion for
 481 un-damped free vibration applied to the structure. This equation is given by:

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{0} \quad (5.12)$$

482 where $\mathbf{M} = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}$ is the mass matrix and $\mathbf{K} = \begin{pmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{pmatrix}$ is the stiffness matrix, and $\mathbf{u} = (u_1, u_2)^\top$
 483 the displacement vector.

484 The posterior distribution of the parameters $\boldsymbol{\Theta} = (\Theta_1, \Theta_2)$ is bi-modal with significantly distant modes
 485 [85, 37], which makes the problem difficult to solve when using classical MCMC sampling approaches.

486 5.2.2. Posterior distribution

487 The proposed approach is applied by considering $K_0 \in \{20, 50, 100, 200\}$ and $\kappa \in \{0, 0.1, 0.2, 0.5\}$.

488 Fig. 6 illustrates the proposed approach in the framework of the previously presented mechanical problem,
 489 with $K_0 = 50$ and $K_t = 10$. Firstly, an initial ED formed by samples drawn from the prior is defined, and the
 490 surrogate log-likelihood is subsequently trained. The first step of SuS procedure enables to identify $\mathbb{P}_t K = 500$
 491 seeds, to generate the population of the first subset. As shown in Fig. 6a, most of these seeds lie outside

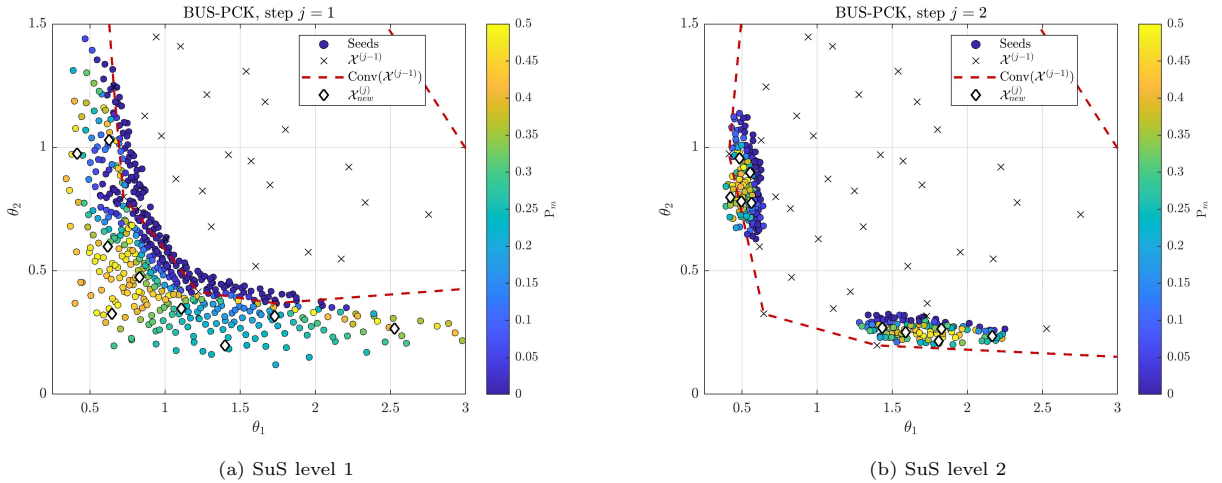


Figure 6: Two DOF structure: misclassification probability of seeds and enrichment of the ED of the surrogate log-likelihood at SuS levels 1 & 2.

492 the convex hull of the current ED. Moreover, the misclassification probability (see Eq. 4.8) of such seeds is
 493 sensibly important, which indicates that log-likelihood predictions are classifying posterior points outside of the
 494 ED convex hull, and/or predictions are significantly uncertain in such zones. The weighted k -means clustering
 495 proposed in Section 4.2 enables to select enrichment points (white diamonds in Fig. 6a) which are quite uniformly
 496 spaced outside from the convex hull of the current ED. The surrogate model is then updated, by using the new
 497 ED. As shown in Fig. 6b, the ED convex hull has been enlarged, and the seeds of SuS level 2 are mostly lying
 498 in it, and form two clusters. Then, the point enrichment step proposes K_t new candidate points which are
 499 located in these clusters. In this way, the accuracy of predictions provided by the surrogate model is improved
 500 specifically in these zones, which correspond to the posterior support.

501 Next, the convergence of aJSD (5.7) with respect to $\mathbb{E}[N_c]$ is shown in Fig. 7. The proposed methodology is
 502 significantly more efficient when $\kappa > 0$, since aJSD values for $\kappa > 0$ are about two orders of magnitude smaller
 503 than those for $\kappa = 0$, for a given ED size (see Fig. 7b). In the non-adaptive case ($\kappa = 0$), the aJSD is converging
 504 much more slower as the ED size increases. This is due to the fact that the posterior support is located in
 505 low-probability zones of the prior (see Fig. 6). Then, without adaptive point enrichment, a very large ED size
 506 K_0 is required to draw a sufficient amount of informative points. In this context, the proposed point enrichment
 507 method based on clustering is well suited to the problem, since it allows to enrich the surrogate log-likelihood
 508 near the posterior high-probability zones, even in the case of a non-connected posterior support.

509 Then, the contours of the posterior joint PDF obtained with reference calculations and the proposed approach
 510 for $K_0 = 200$ and $\kappa = 0.2$ are shown in Fig. 8. The shape of the joint posterior is well approximated by the
 511 proposed approach, with a limited amount of model calls ($\sim 10^2$). Furthermore, the location of posterior modes
 512 is well reproduced. QoI related to left and right clusters of the identified posterior are summarized in Tab. 2 &
 513 3. An average number of model calls of $\mathbb{E}[N_c] = 320$ is sufficient to provide an accurate approximation of the
 514 posterior PDF, which represents a reduction of about 98.4% of the total cost of $2 \cdot 10^4$ model calls required by
 515 using the full likelihood function.

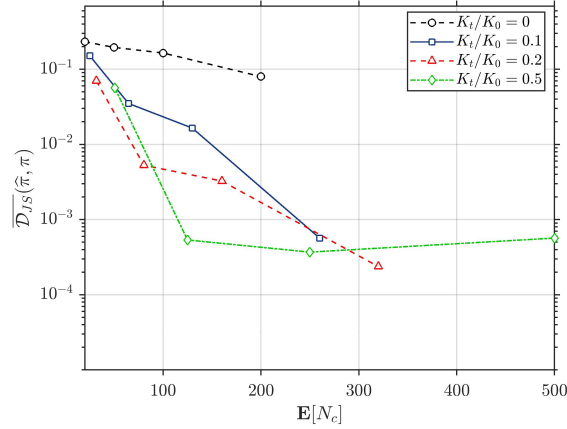


Figure 7: 2 DOF structure: convergence of aJSD with respect to $\mathbb{E}[N_c]$.

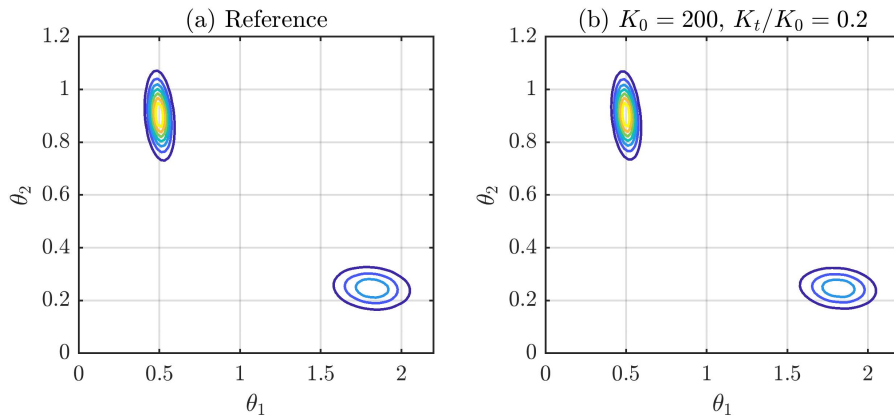


Figure 8: Contours of the posterior joint PDF: (a) Reference; (b) $K_0 = 200$ and $K_t/K_0 = 0.2$.

	K_0	K_t/K_0	$\mathbb{E}[N_c]$	$\mathbb{E}[\theta_1 \tilde{\mathbf{f}}]$	$\mathbb{E}[\theta_2 \tilde{\mathbf{f}}]$	$\sqrt{\text{Var}[\theta_1 \tilde{\mathbf{f}}]}$	$\sqrt{\text{Var}[\theta_2 \tilde{\mathbf{f}}]}$	$\mathcal{Z} (\times 10^{-3})$
BUS-PCK	50	0.2	80.4	0.502 (0.0)	0.893 (0.7)	0.044 (18.8)	0.086 (23.9)	1.484 (1.3)
	50	0.5	125	0.502 (0.0)	0.898 (0.1)	0.039 (3.5)	0.073 (3.9)	1.474 (1.9)
	100	0.2	160	0.505 (0.6)	0.887 (1.4)	0.045 (21.3)	0.085 (22.0)	1.618 (7.6)
	100	0.5	250	0.502 (0.0)	0.899 (0.0)	0.039 (3.0)	0.072 (2.9)	1.485 (1.2)
	200	0.2	320	0.501 (0.2)	0.900 (0.1)	0.038 (2.1)	0.070 (0.9)	1.509 (0.4)
	200	0.5	500	0.502 (0.0)	0.899 (0.0)	0.038 (0.6)	0.070 (0.3)	1.483 (1.3)
Reference			20000	0.502	0.899	0.037	0.070	1.503

Table 2: 2 DOF structure: posterior QoI related to the left cluster, with relative errors in %.

516

517 5.3. Multimodal mixture of Gaussians

518 5.3.1. Problem description

519 The third case study is introduced in [86], and involves a mixture of Gaussian distributions. A uniform prior

	K_0	K_t/K_0	$\mathbb{E}[N_c]$	$\mathbb{E}[\theta_1 \tilde{\mathbf{f}}]$	$\mathbb{E}[\theta_2 \tilde{\mathbf{f}}]$	$\sqrt{\text{Var}[\theta_1 \tilde{\mathbf{f}}]}$	$\sqrt{\text{Var}[\theta_2 \tilde{\mathbf{f}}]}$	$\mathcal{Z} (\times 10^{-3})$
BUS-PCK	50	0.2	80.4	1.811 (0.2)	0.245 (0.8)	0.178 (25.7)	0.022 (17.9)	1.484 (1.3)
	50	0.5	125	1.812 (0.1)	0.246 (0.2)	0.144 (1.8)	0.020 (5.9)	1.474 (1.9)
	100	0.2	160	1.801 (0.7)	0.246 (0.3)	0.154 (8.5)	0.021 (11.9)	1.618 (7.6)
	100	0.5	250	1.813 (0.1)	0.246 (0.2)	0.144 (1.3)	0.019 (3.6)	1.485 (1.2)
	200	0.2	320	1.814 (0.0)	0.247 (0.0)	0.143 (0.9)	0.019 (0.5)	1.509 (0.4)
	200	0.5	500	1.814 (0.0)	0.247 (0.0)	0.144 (1.3)	0.019 (2.1)	1.483 (1.3)
Reference			20000	1.814	0.247	0.142	0.019	1.503

Table 3: 2 DOF structure: posterior QoI related to the right cluster, with relative errors in %.

520 over the square $[0, a] \times [0, a]$ is considered, whereas the likelihood function is defined by:

$$\mathcal{L}(\boldsymbol{\theta}; \{\boldsymbol{\mu}_j\}) = \sum_{j=1}^n w_j \varphi_2(\boldsymbol{\theta}; \boldsymbol{\mu}_j, \sigma) \quad (5.13)$$

521 where the $\{\boldsymbol{\mu}_j\}$ are drawn from the prior $\mathcal{U}([0, a] \times [0, a])$, $\{w_j\}$ are the weights of the mixture, and $\varphi_2(\cdot; \boldsymbol{\mu}_j, \sigma)$
522 is the PDF of the bivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_j, \sigma^2 \mathbf{I})$. We set $a = 10$, $\sigma = 0.1$, $n = 10$ and $w_j = 1/n$ for
523 all $j \in \{1, \dots, n\}$. The corresponding posterior PDF $\pi(\boldsymbol{\theta}; \{\boldsymbol{\mu}_j\}) \propto \mathbf{1}_{[0, a]^2}(\boldsymbol{\theta}) \mathcal{L}(\boldsymbol{\theta}; \{\boldsymbol{\mu}_j\})$ is multimodal, with 10
modes. This is illustrated in Fig. 9, where contours of the likelihood function are presented.

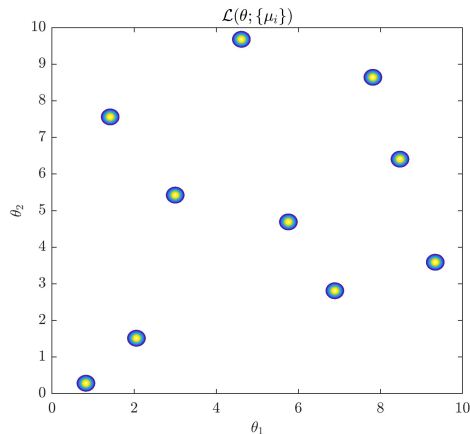


Figure 9: Multimodal Gaussian mixture: contours of likelihood function

524

525 5.3.2. Posterior distribution

526 The proposed approach is applied by considering $K_0 \in \{50, 100, 150, 250\}$ and $\kappa \in \{0, 0.1, 0.2\}$. Fig. 10
527 presents a comparison between posterior samples drawn with the SuS procedure (see Section 2.3) and samples
528 drawn with the proposed approach, for $K_0 \in \{100, 150, 250\}$ and $\kappa = 0.2$. Firstly, as depicted by Fig. 10a, the
529 posterior samples obtained with the aBUS-SuS algorithm separate into 10 significantly spaced clusters. This
530 underlines the complexity of the posterior PDF, which would be difficult to sample by using classical MCMC

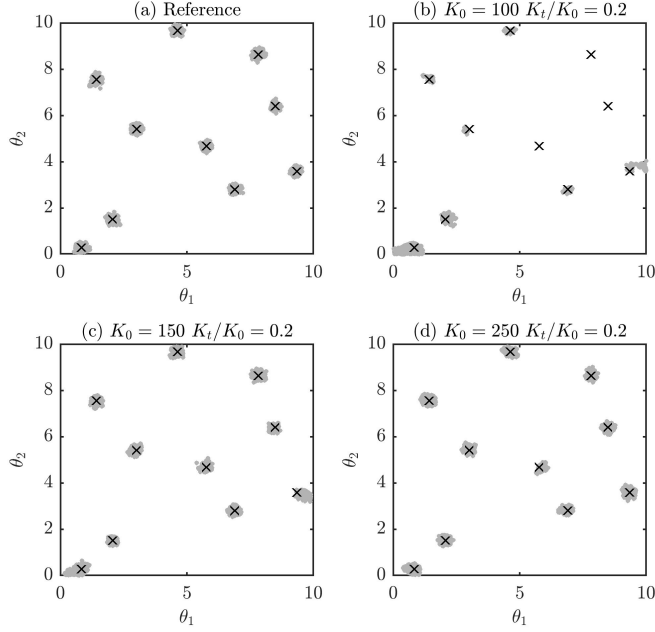


Figure 10: Multimodal Gaussian mixture: reference (a) and BUS-PCK-based (b-d) posterior samples (sample size: $5 \cdot 10^3$). Black crosses (\times) represent the modes $\{\mu_j\}$ of the posterior distribution.

531 approaches, for which the generated Markov chains may be trapped near few modes [86].

532 Next, for $K_0 = 100$ and $\kappa = 0.2$, the proposed approach completely misses 3 modes (see Fig. 10b). When
 533 $K_0 = 150$ and $\kappa = 0.2$, the proposed approach correctly identifies the 10 modes of the posterior (see Fig. 10c),
 534 even though the spread of the clusters near $(0, 0)$ and $(10, 3)$ is roughly approximated. Lastly, the 10 modes of
 535 the posterior are well reproduced when $K_0 = 250$ and $\kappa = 0.2$ (see Fig. 10d).

536 The convergence of aJSD (5.7) with respect to $\mathbb{E}[N_c]$ is shown in Fig. 11. First, when no surrogate model
 537 adaptation is performed (*i.e.* $\kappa = 0$), the aJSD stagnates around a value of 10^{-1} . This may be explained by the
 538 highly localized behavior of the likelihood function, which seems to be difficult to approximate with a fixed ED
 539 built from the uniform prior $\mathcal{U}([0, a] \times [0, a])$. For a given number of full likelihood calls, significantly lower aJSD
 540 values are obtained when $\kappa > 0$. In particular, when $\mathbb{E}[N_c] \geq 200$, the aJSD with $\kappa = 0.2$ is approximately two
 541 orders of magnitude smaller than aJSD in the case $\kappa = 0$. This underlines the efficiency of the proposed adaptive
 542 point enrichment scheme, as mentioned in the example presented in Section 5.2. The proposed approach takes
 543 advantage of the local features provided by the Gaussian process of the PCK surrogate for better catching the
 544 likelihood behavior near the modes $\{\mu_j\}$ of the posterior.

545 Finally, QoI related to the posterior distribution are summarized in Tab. 4. For the sake of brevity, only the
 546 components of the posterior mean $\mathbb{E}[\theta|\{\mu_j\}]$ and the standard deviation $\sqrt{\text{Var}[\theta|\{\mu_j\}]}$ are considered as QoI,
 547 rather than considering the statistics of each of the 10 clusters of the posterior. The proposed approach provides
 548 correct estimates of $\mathbb{E}[\theta_i|\{\mu_j\}]$ and $\sqrt{\text{Var}[\theta_i|\{\mu_j\}]}$ from $K_0 = 100$ and $\kappa = 0.2$. Moreover, fair estimates of
 549 the model evidence \mathcal{Z} are obtained with $K_0 = 150$ and $\kappa = 0.2$. Finally, it is worth noting that the proposed

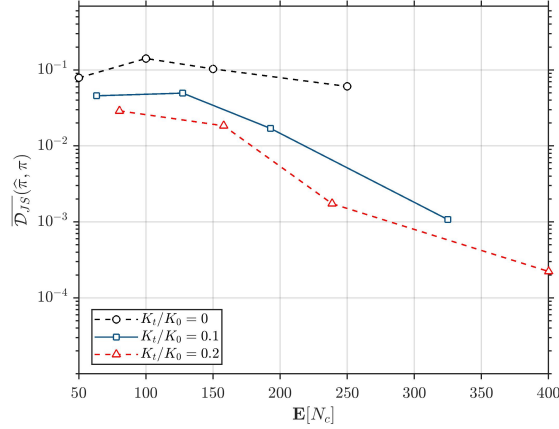


Figure 11: Multimodal Gaussian mixture: convergence of aJSD with respect to $\mathbb{E}[N_c]$

550 method provides a quite good approximation of the posterior with a number of full likelihood evaluations of
 551 about $\mathbb{E}[N_c] \approx 240$ (see Tab. 4), which corresponds to a reduction of about 98.8% of the total cost of $2 \cdot 10^4$ full
 552 likelihood evaluations required by a direct application of the SuS procedure.

	K_0	K_t/K_0	$\mathbb{E}[N_c]$	$\mathbb{E}[\theta_1 \{\mu_j\}]$	$\mathbb{E}[\theta_2 \{\mu_j\}]$	$\sqrt{\text{Var}[\theta_1 \{\mu_j\}]}$	$\sqrt{\text{Var}[\theta_2 \{\mu_j\}]}$	\mathcal{Z}
BUS-PCK	100	0.1	127.4	4.323 (14.6)	4.309 (13.8)	2.730 (7.2)	2.960 (2.2)	0.063 (525.4)
	100	0.2	158	4.584 (9.5)	4.778 (4.4)	2.825 (3.9)	2.986 (3.1)	0.046 (352.0)
	150	0.1	192.9	4.590 (9.4)	4.794 (4.1)	2.857 (2.9)	2.949 (1.8)	0.024 (138.0)
	150	0.2	238.8	5.069 (0.1)	4.921 (1.5)	2.954 (0.4)	2.852 (1.5)	0.012 (16.3)
	250	0.1	325	5.053 (0.2)	4.830 (3.4)	2.901 (1.4)	2.843 (1.8)	0.011 (4.4)
	250	0.2	400	5.058 (0.1)	4.953 (0.9)	2.927 (0.5)	2.885 (0.4)	0.010 (2.7)
Reference			20000	5.065	4.998	2.941	2.896	0.010

Table 4: Multimodal Gaussian mixture: posterior QoI, with relative errors in %.

553 5.4. Diffusion problem

554 5.4.1. Problem description

555 The last case study consists in a 1D diffusion problem introduced in [26], as the steady state version of the
 556 problem studied in [33]. The problem is described by the following diffusion equation on the domain $\mathcal{D} = [0, 1]$:

$$\nabla (a(x)\nabla u(x)) + b(x) = 0 \quad (5.14)$$

557 where the diffusivity $a(x)$ is represented by a random field, and the source term $b(x)$ is given by:

$$b(x) = \sum_{i=1}^N \frac{s_i}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(l_i - x)^2}{2\sigma_i^2}\right) \quad (5.15)$$

558 This term corresponds to N localized sources with locations $(l_i)_{1 \leq i \leq N}$ and strengths $(s_i)_{1 \leq i \leq N}$ and widths
 559 $(\sigma_i)_{1 \leq i \leq N}$. One considers $N = 3$ sources at locations $l_1 = 0.25$, $l_2 = 0.5$ and $l_3 = 0.75$. The strengths

560 are assumed to be identical, and are set as $s_i = 10$, as well as the widths, which are fixed to $\sigma_i^2 = 10^{-3}$.
 561 Furthermore, Dirichlet boundary conditions $u(0) = u(1) = 0$ are adopted for solving Eq. (5.14). The diffusion
 562 equation (5.14) is solved by using linear finite elements, on a uniform grid with spacing $\Delta x = \frac{1}{48}$.

563 The problem consists in inferring the diffusivity $a(x)$ from noisy measurements of the field $u(x)$, provided by
 564 a set of $n = 11$ sensors uniformly spaced in \mathcal{D} (excluding the endpoints). The prior of the log-diffusivity $\log a$ is
 565 supposed to be a Gaussian stationary random field, with mean $\mu_{\log a} = 0.1$ and standard deviation $\sigma_{\log a} = 0.2$,
 566 and an exponential correlation kernel:

$$R_{\log a}(x, x') = \exp\left(-\frac{|x - x'|}{\ell_c}\right) \quad (5.16)$$

567 with $\ell_c = 0.3$.

568 The log-diffusivity random field is represented by a truncated Karhunen-Loève (KL) [87, 57] expansion:

$$\log a(x) \approx \mu_{\log a} + \sum_{i=1}^m \sqrt{\lambda_i} \zeta_i \phi_i(x) \quad (5.17)$$

569 where $m = 10$ [26], $(\zeta_i)_{1 \leq i \leq m}$ are the m first KL eigenmodes, and $(\lambda_i, \phi_i)_{1 \leq i \leq m}$ are the eigenvalues and eigen-
 570 functions associated to the covariance kernel $\Sigma_{\log a}(x, x') = \sigma_{\log a}^2 R_{\log a}(x, x')$. The latter are satisfying the
 571 Fredholm integral equations of the second kind:

$$\int_{\mathcal{D}} \Sigma_{\log a}(x, x') \phi_i(x') dx' = \lambda_i \phi_i(x) \quad (5.18)$$

572 which can be solved analytically in the case of exponential kernels [57].

573 Then, synthetic data is generated from a realization of the log-diffusivity random field, by computing the
 574 corresponding field $u(x)$ at sensors locations and adding to it a Gaussian noise term with zero mean and standard
 575 deviation $\sigma_u = 0.1$. In order to avoid to commit an *inverse crime* [1], this synthetic data is generated by using
 576 a much finer grid for solving Eq. (5.14). The $m = 10$ first KL eigenmodes $(\zeta_i)_{1 \leq i \leq m}$ are subsequently inferred
 577 from the generated data.

578 5.4.2. Posterior distribution

579 The proposed approach is applied by considering $K_0 \in \{50, 100, 250, 500\}$ and $\kappa \in \{0, 0.1, 0.2\}$. The conver-
 580 gence of aJSD (5.7) with respect to $\mathbb{E}[N_c]$ is shown in Fig. 12. The performance of the approach depends less
 581 on the adaptation step of the surrogate model, since the aJSD presents a similar behavior for $\kappa = 0$ and $\kappa > 0$.
 582 However, for moderate to large ED sizes ($\approx 250 - 500$), the adaptive point enrichment provides a slightly smaller
 583 aJSD than in the non-adaptive case ($5 \cdot 10^{-3}$ when $\kappa = 0$ and 10^{-3} when $\kappa > 0$). Such behavior may be explained
 584 by the fact that most of the posterior mass is enclosed in high-probability zones of the prior. In this case, the
 585 adaptive point enrichment targets zones in which samples are frequently drawn when sampling from the prior.

586 This clearly underlines the difference of performance between the present diffusion problem and the mechanical
 587 problem presented in Section 5.2, which presented a bi-modal with support located in low-probability zones of
 588 the prior.

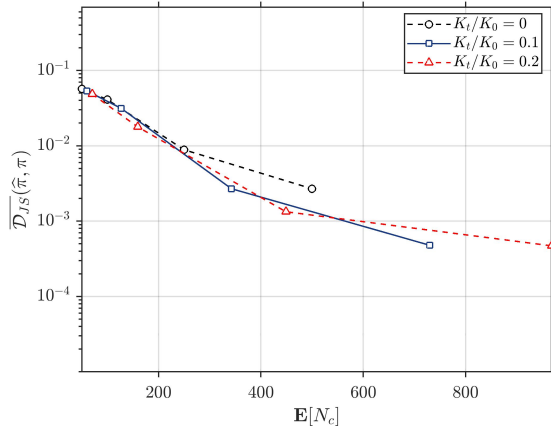


Figure 12: Diffusion problem: convergence of aJSD with respect to $\mathbb{E}[N_c]$.

589 Then, it is worth noting that the *active dimensionality* of the problem is limited to 6, since the posterior
 590 distribution of the four last KL eigenmodes is sensibly close to their prior distribution. This has been often
 591 encountered in the framework of UQ benchmark problems related to diffusion with random diffusivity fields [23].
 592 Consequently, the following analysis will be then focused on the six first KL eigenmodes.

593 The posterior univariate and bivariate PDF of the six first KL eigenmodes are shown in Fig.13. The reference
 594 PDF and those obtained for $K_0 = 250$ and $\kappa = 0.1$ are compared. The shape of the posterior distribution is well
 595 approximated with the proposed approach, with a quite limited amount of model calls (≈ 350). In particular,
 596 the correlation structure is well reproduced, notably regarding to the couples (ζ_1, ζ_3) and (ζ_2, ζ_4) which exhibit
 597 quite strong linear correlations. Posterior mean and standard deviation values are summarized in Tab. 5 & 6.
 598 It is also worthwhile noting that the following approach provides a fair estimate of the model evidence with a
 599 quite moderate amount of model calls ($\approx 350 - 500$), as underlined by Tab. 7 which summarizes values of the
 600 model evidence. As mentioned earlier, a quite important advantage of the proposed approach is that it enables
 601 to compute the model evidence as a by-product of the SuS procedure (see Algorithm 1). Classical MCMC
 602 approaches do not enable to estimate the model evidence once the sampling procedure is performed.

$\mathbb{E}[\zeta_i \mathbf{d}]$	K_0	K_t/K_0	$\mathbb{E}[N_c]$	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6
BUS-PCK	100	0.1	127	-0.55 (23.4)	-0.13 (68.1)	0.71 (29.6)	-0.57 (41.2)	0.05 (116.0)	-0.54 (46.0)
	100	0.2	159.6	-0.59 (17.2)	-0.26 (33.0)	0.76 (24.4)	-0.72 (24.7)	-0.01 (95.9)	-0.80 (20.1)
	250	0.1	342.5	-0.70 (2.9)	-0.41 (3.0)	0.96 (4.3)	-0.89 (7.8)	-0.22 (27.0)	-0.96 (5.1)
	250	0.2	449	-0.70 (2.6)	-0.40 (2.3)	0.97 (3.8)	-0.93 (3.6)	-0.23 (22.1)	-0.96 (5.2)
	500	0.1	730	-0.71 (1.1)	-0.37 (6.5)	0.99 (1.7)	-0.91 (5.2)	-0.28 (5.0)	-1.02 (1.2)
	500	0.2	968	-0.72 (0.1)	-0.38 (4.8)	1.01 (0.3)	-0.92 (4.3)	-0.30 (2.8)	-1.03 (2.6)
Reference			27600	-0.72	-0.39	1.00	-0.96	-0.30	-1.01

Table 5: Diffusion problem: posterior means of few KL eigenmodes.

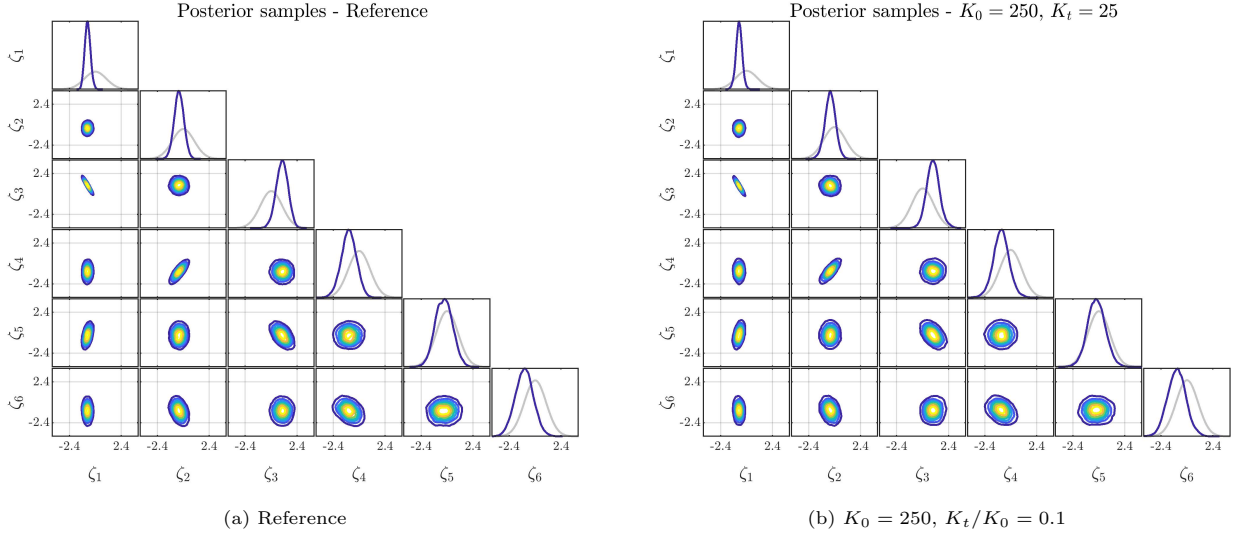


Figure 13: Diffusion problem: posterior univariate and bivariate marginal PDF of the six first KL eigenmodes

$\sqrt{\text{Var}[\zeta_i \mathbf{d}]}$	K_0	K_t/K_0	$\mathbb{E}[N_c]$	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6
BUS-PCK	100	0.1	127	0.45 (69.5)	0.71 (59.0)	0.87 (59.1)	0.98 (38.9)	1.13 (43.2)	1.07 (28.0)
	100	0.2	159.6	0.38 (44.5)	0.62 (40.3)	0.77 (39.6)	0.86 (22.8)	1.00 (26.6)	1.04 (24.1)
	250	0.1	342.5	0.29 (8.9)	0.49 (10.7)	0.60 (9.1)	0.74 (4.8)	0.82 (3.9)	0.84 (0.7)
	250	0.2	449	0.28 (5.5)	0.48 (7.1)	0.58 (6.3)	0.72 (2.7)	0.81 (3.3)	0.85 (1.3)
	500	0.1	730	0.26 (0.2)	0.45 (1.4)	0.55 (0.2)	0.69 (1.3)	0.79 (0.2)	0.85 (1.8)
	500	0.2	968	0.26 (0.4)	0.45 (0.4)	0.55 (0.3)	0.70 (1.0)	0.77 (2.1)	0.83 (0.4)
Reference			27600	0.26	0.44	0.55	0.70	0.79	0.84

Table 6: Diffusion problem: posterior standard deviations of few KL eigenmodes.

	K_0	K_t/K_0	$\mathbb{E}[N_c]$	\mathcal{Z}
BUS-PCK	100	0.1	127	87.09 (254.1)
	100	0.2	159.6	29.41 (19.6)
	250	0.1	342.5	22.73 (7.6)
	250	0.2	449	23.36 (5.0)
	500	0.1	730	23.75 (3.4)
	500	0.2	968	24.77 (0.7)
Reference			27600	24.59

Table 7: Diffusion problem: model evidence values.

603 Next, a comparison between **the true log-diffusivity field**, the reference posterior log-diffusivity field and that
604 obtained with the proposed approach for $K_0 \in \{250, 500\}$ and $\kappa = 0.2$ is given in Fig. 14. As a validation of
605 the reference results, one observes that the posterior log-diffusivity field obtained from SuS procedure matches
606 well with the realization used for generating the synthetic data. Furthermore, the proposed approach provides
607 a fair approximation of the posterior log-diffusivity when $K_0 = 250$ and $\kappa = 0.2$, with a slight overestimation of
608 the variance, though (see Fig.14a). Then, the posterior log-diffusivity field is well approximated when $K_0 = 500$
609 and $\kappa = 0.2$ (see Fig. 14b).

610 Then, contours of the reference posterior covariance kernel of the log-diffusivity field and its approximations

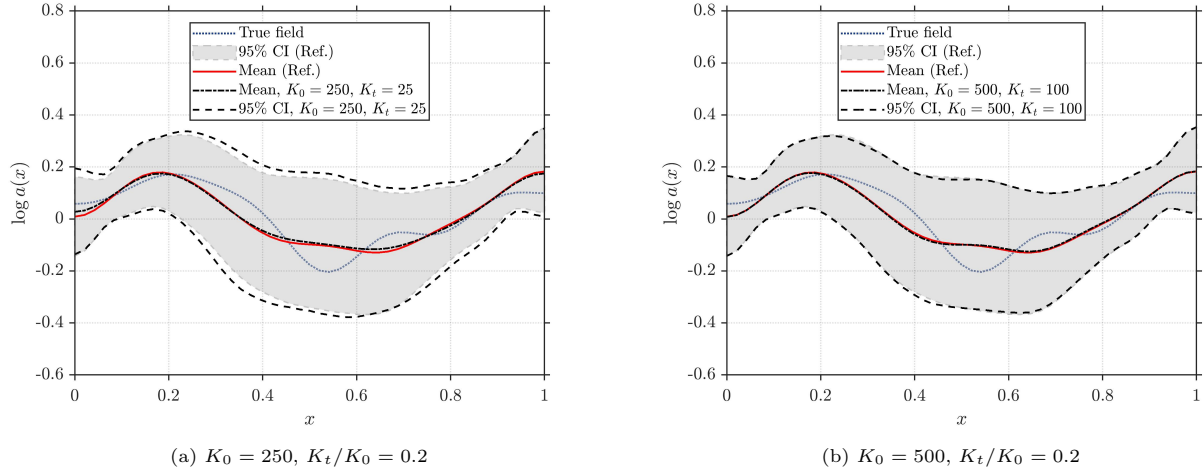


Figure 14: Diffusion problem: posterior log-diffusivity field.

611 provided by the proposed approach for $K_0 \in \{250, 500\}$ and $\kappa = 0.2$ are shown in Fig. 15. Approximately
 612 450 model calls are sufficient to provide a quite accurate approximation of the covariance kernel, in its positive
 613 correlation zones (see Fig.15a). The case $K_0 = 500$ and $\kappa = 0.2$ presents a sensibly accurate approximation of
 614 the posterior covariance kernel, even in the zones which exhibit negative correlations (see Fig.15b).

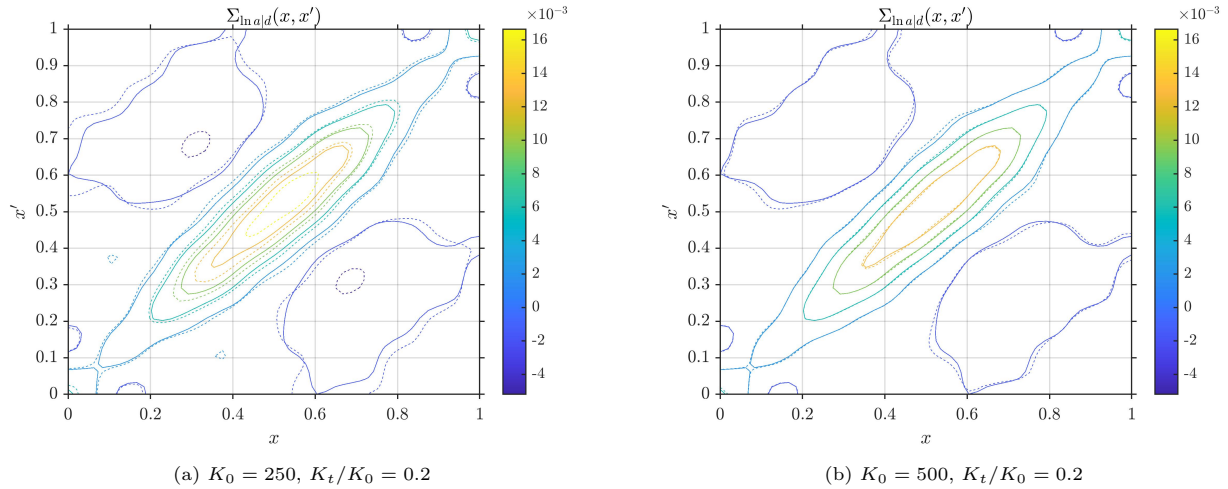


Figure 15: Diffusion problem: posterior covariance kernel of the log-diffusivity field. Continuous lines refer to reference results, whereas dashed lines refer to ones provided by the proposed approach.

615 Lastly, it is worth noting that a satisfactory approximation of the posterior PDF is obtained with an average
 616 number of model calls of $\mathbb{E}[N_c] = 730$, which represents a reduction of about 97.4% of the total cost of $2.76 \cdot 10^4$
 617 required by an inversion using the full likelihood (see Tab. 5 & 6).

618 6. Conclusion

619 In this paper, a Bayesian inversion method using adaptive surrogate models has been presented, based on
 620 the BUS framework [25, 26], which aims at reformulating the classical Bayesian inference into the estimation of

621 a rare event. In this framework, the method combines a well-known Structural Reliability (SR) method known
622 as Subset Simulation (SuS) [29, 30] and adaptive Polynomial Chaos Kriging (PCK) [42, 43], which provides
623 powerful surrogate models able to approximate both global and local behaviors of computational models. In
624 this context, a PCK surrogate model is firstly built in order to catch the global trend of the log-likelihood on
625 the prior support. This surrogate model is subsequently enriched throughout the SuS sampling procedure by
626 selecting points in informative regions, based on principles taken from active learning SR methods [39, 78, 41].
627 The Kriging part of the surrogate model then enables to better catch local variations of log-likelihood, as well
628 as provide an local error measure which can be used for enriching the surrogate model.

629 The proposed approach has been applied to three cases studies of increasing complexity, involving a one-
630 dimensional analytical function, a two degrees-of-freedom structure and a diffusion problem with moderate
631 active dimensionality. It appears from numerical investigations that the proposed approach provides accurate
632 approximations of the posterior distributions, even in the case of likelihood functions with localized behavior, or
633 multi-modal distributions, which are usually difficult to sample for many existing Bayesian inversion algorithms.
634 Moreover, such approximations have been obtained for a quite limited amount of model calls (*i.e.* $\approx 50 - 500$),
635 which makes the proposed approach suitable for computationally demanding models. For each test case studied,
636 the proposed approach has provided a reduction of at least 97.5% of the total number of model calls required by
637 aBUS-SuS inversions using the full likelihood function. Furthermore, classical MCMC-based Bayesian inversion
638 approaches do not enable to assess the performance of several models against observed data, since the compu-
639 tation of the model evidence is avoided [6]. A particular advantage provided by the BUS framework is that the
640 proposed approach enables to estimate the model evidence, as a by-product of the rare event estimation through
641 the SuS procedure. Therefore, the proposed method would be suitable for model selection [88].

642 Then, the proposed approach presents several shortcomings, which may be addressed in further works.
643 Firstly, the updating step of the BUS scaling constant is based on the computation of the convex hull of
644 the current experimental design of the PCK surrogate model, in order to filter points whose surrogate model
645 predictions are too uncertain. This is done for limiting the risk of making a wrong estimation of the maximal
646 likelihood value needed for updating the BUS scaling constant. As mentioned in [75], the Quickhull algorithm
647 [76] used for computing convex hulls significantly suffers from the so-called *curse of dimensionality*. Indeed,
648 for dimensions larger than ≈ 15 , the number of facets of the computed convex hull grows fast with dimension,
649 which considerably slows the overall procedure. Then, it would be necessary to define an alternative filter for
650 updating the BUS scaling constant, based on the distance to the current points of experimental design, and the
651 uncertainties associated to predictions of likelihood values. Moreover, the optimization problem involved in the
652 training of Kriging parameters may be difficult to solve in high dimensions [75, 38]. This is in line with results
653 from a recent survey and benchmark on surrogate-based active learning SR methods [84], which suggest that
654 PCK surrogate models are well suited for parameter dimensions $d < 20$, whereas PCE surrogates perform well
655 in higher dimensions $d \geq 20$. Thus, the proposed approach may be adapted to PCE for tackling problems with
656 higher dimensions, provided a local error estimate adapted to PCE is used. In this context, the Fraction of
657 Bootstrap Replicates (FBR) introduced in [73] could be used in the framework of the proposed adaptive point

658 enrichment scheme. Such guidelines might be considered in further work for tackling problems with higher
659 dimensions. Lastly, it appears from test cases results that the stopping criterion of the point enrichment step
660 has been rarely reached during calculations. This may be due to the fact that the adopted threshold value for
661 the misclassification probability (see Section 4.3) is too conservative, as underlined in [41]. Therefore, a more
662 precise stopping criterion may be defined in order to avoid unnecessary model calls throughout the Bayesian
663 updating scheme.

664 References

- 665 [1] J. Kaipio, E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, 2005. doi:10.1007/
666 b138659.
667 URL <http://dx.doi.org/10.1007/b138659>
- 668 [2] J. Idier, *Bayesian Approach to Inverse Problems*, ISTE, 2008. doi:10.1002/9780470611197.
669 URL <http://dx.doi.org/10.1002/9780470611197>
- 670 [3] A. M. Stuart, Inverse problems : a Bayesian perspective, *Acta Numer.* 19 (2010) 451–559. doi:10.1017/
671 s0962492910000061.
- 672 [4] G. Claeskens, N. L. Hjort, *Model Selection and Model Averaging*, Cambridge Series in Statistical and
673 Probabilistic Mathematics, Cambridge University Press, 2008. doi:10.1017/CB09780511790485.
- 674 [5] C. P. Robert, G. Casella, *Monte Carlo Statistical Methods*, 2nd Edition, Springer Series in Statistics, New
675 York, 2004. doi:10.1007/978-1-4757-4145-2.
- 676 [6] S. Brooks, A. Gelman, G. Jones, X.-L. Meng, *Handbook of Markov Chain Monte Carlo*, Chapman &
677 Hall/CRC, Boca Raton, Florida, USA, 2011. doi:10.1201/b10905.
- 678 [7] N. Metropolis, A. W. Rosenbluth, A. H. Teller, Equations of state calculations by fast computing machines,
679 *J. Chem. Phys.* 21 (1953) 1087–1092. doi:10.2172/4390578.
- 680 [8] W. Hastings, Monte Carlo sampling methods using Markov chains and their application, *Biometrika* 57
681 (1970) 97–109. doi:10.1093/biomet/57.1.97.
- 682 [9] H. Haario, M. Laine, A. Mira, E. Saksman, *DRAM: Efficient adaptive MCMC*, *Statistics and Computing*
683 16 (4) (2006) 339–354. doi:10.1007/s11222-006-9438-0.
684 URL <https://doi.org/10.1007/s11222-006-9438-0>
- 685 [10] C. Andrieu, J. Thoms, *A tutorial on adaptive MCMC*, *Statistics and Computing* 18 (4) (2008) 343–373.
686 doi:10.1007/s11222-008-9110-y.
687 URL <https://doi.org/10.1007/s11222-008-9110-y>
- 688 [11] R. M. Neal, *MCMC Using Hamiltonian Dynamics*, Chapman & Hall - CRC Press, Boca Raton London,
689 2011, Ch. 5, p. 13–162.

- 690 [12] J. Goodman, J. Weare, [Ensemble samplers with affine invariance](#), *Commun. Appl. Math. Comput. Sci.* 5 (1)
691 (2010) 65–80. doi:10.2140/camcos.2010.5.65.
692 URL <https://doi.org/10.2140/camcos.2010.5.65>
- 693 [13] M. Cowles, B. Carlin, Markov chain Monte Carlo convergence diagnostics : a comparative review, *J. Amer.*
694 *Stat. Assoc.* 91 (1996) 883–904. doi:10.1080/01621459.1996.10476956.
- 695 [14] S. E. Adlouni, A. Favre, B. Bobée, Comparison of methodologies to assess the convergence of Markov chain
696 Monte Carlo methods, *Computational Statistics & Data Analysis* 50 (2006) 2685–2701. doi:10.1016/j.
697 *csda.2005.04.018*.
- 698 [15] A. Gelman, B. Rubin, Inference from iterative simulation using multiple sequences, *Statistical Science* 7
699 (1992) 457–472. doi:10.1214/ss/1177011136.
- 700 [16] C. W. Fox, S. J. Roberts, [A tutorial on variational Bayesian inference](#), *Artificial Intelligence Review* 38 (2)
701 (2012) 85–95. doi:10.1007/s10462-011-9236-8.
702 URL <https://doi.org/10.1007/s10462-011-9236-8>
- 703 [17] J.-M. Marin, P. Pudlo, C. P. Robert, R. J. Ryder, [Approximate Bayesian computational methods](#), *Statistics*
704 *and Computing* 22 (6) (2012) 1167–1180. doi:10.1007/s11222-011-9288-2.
705 URL <https://doi.org/10.1007/s11222-011-9288-2>
- 706 [18] L. Tierney, R. E. Kass, J. B. Kadane, [Approximate Marginal Densities of Nonlinear Functions](#), *Biometrika*
707 76 (3) (1989) 425–433. doi:10.2307/2336109.
708 URL <http://www.jstor.org/stable/2336109>
- 709 [19] J. Skilling, Nested Sampling for General Bayesian Computation, *Bayesian Analysis* 1 (4) (2006) 833–860.
710 doi:10.1214/06-ba127.
- 711 [20] A. H. Elsheikh, M. F. Wheeler, I. Hoteit, [Hybrid nested sampling algorithm for Bayesian model selection](#)
712 [applied to inverse subsurface flow problems](#), *Journal of Computational Physics* 258 (2014) 319–337. doi:
713 <https://doi.org/10.1016/j.jcp.2013.10.001>.
714 URL <https://www.sciencedirect.com/science/article/pii/S0021999113006669>
- 715 [21] T. El Moselhy, Y. Marzouk, Bayesian inference with optimal maps, *Journal of Computational Physics*
716 231 (23) (2012) 7815–7850. doi:10.1016/j.jcp.2012.07.022.
- 717 [22] J. B. Nagel, B. Sudret, Spectral likelihood expansions for Bayesian inference, *Journal of Computational*
718 *Physics* 309 (2016) 267–294. doi:10.1016/j.jcp.2015.12.047.
- 719 [23] P.-R. Wagner, S. Marelli, B. Sudret, [Bayesian model inversion using stochastic spectral embedding](#), *Journal*
720 *of Computational Physics* 436 (2021) 110141. doi:<https://doi.org/10.1016/j.jcp.2021.110141>.
721 URL <https://www.sciencedirect.com/science/article/pii/S0021999121000334>

- 722 [24] S. Marelli, P.-R. Wagner, C. Lataniotis, B. Sudret, Stochastic Spectral Embedding, International Journal
723 for Uncertainty Quantification 11 (2) (2021) 25–47. doi:[10.1615/Int.J.UncertaintyQuantification.](https://doi.org/10.1615/Int.J.UncertaintyQuantification.2020034395)
724 [2020034395](https://doi.org/10.1615/Int.J.UncertaintyQuantification.2020034395).
- 725 [25] D. Straub, I. Papaioannou, Bayesian Updating with Structural Reliability Methods, Journal of Engineering
726 Mechanics 141 (3) (2015) 04014134. doi:[10.1061/\(ASCE\)EM.1943-7889.0000839](https://doi.org/10.1061/(ASCE)EM.1943-7889.0000839).
- 727 [26] D. Straub, I. Papaioannou, W. Betz, Bayesian analysis of rare events, Journal of Computational Physics
728 314 (2016) 538–556. doi:<https://doi.org/10.1016/j.jcp.2016.03.018>.
729 URL <https://www.sciencedirect.com/science/article/pii/S0021999116001704>
- 730 [27] R. Rackwitz, B. Flessler, Structural reliability under combined random load sequences, Computers & Struc-
731 tures 9 (5) (1978) 489–494. doi:[https://doi.org/10.1016/0045-7949\(78\)90046-9](https://doi.org/10.1016/0045-7949(78)90046-9).
732 URL <https://www.sciencedirect.com/science/article/pii/0045794978900469>
- 733 [28] R. E. Melchers, A. T. Beck, Structural reliability analysis and prediction, John Wiley & Sons Ltd, Hoboken,
734 NJ, 2018. doi:[10.1002/9781119266105](https://doi.org/10.1002/9781119266105).
- 735 [29] S.-K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Prob-
736 abilistic Engineering Mechanics 16 (4) (2001) 263–277. doi:[https://doi.org/10.1016/S0266-8920\(01\)](https://doi.org/10.1016/S0266-8920(01)00019-4)
737 [00019-4](https://doi.org/10.1016/S0266-8920(01)00019-4).
738 URL <https://www.sciencedirect.com/science/article/pii/S0266892001000194>
- 739 [30] W. Betz, I. Papaioannou, J. L. Beck, D. Straub, Bayesian inference with Subset Simulation: Strategies and
740 improvements, Computer Methods in Applied Mechanics and Engineering 331 (2018) 72–93. doi:<https://doi.org/10.1016/j.cma.2017.11.021>.
741 URL <https://www.sciencedirect.com/science/article/pii/S0045782517307211>
- 743 [31] Y. M. Marzouk, H. N. Najm, L. A. Rahn, Stochastic spectral methods for efficient Bayesian solution of
744 inverse problems, Journal of Computational Physics 224 (2) (2007) 560–586. doi:[10.1016/j.jcp.2006.](https://doi.org/10.1016/j.jcp.2006.10.010)
745 [10.010](https://doi.org/10.1016/j.jcp.2006.10.010).
- 746 [32] Y. M. Marzouk, D. B. Xiu, A stochastic collocation approach to Bayesian inference in inverse problems,
747 Commun. Comput. Phys. 6 (2009) 826–847. doi:[10.4208/cicp.2009.v6.p826](https://doi.org/10.4208/cicp.2009.v6.p826).
- 748 [33] Y. M. Marzouk, H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of Bayesian
749 inference in inverse problems, Journal of Computational Physics 228 (6) (2009) 1862–1902. doi:[10.1016/](https://doi.org/10.1016/j.jcp.2008.11.024)
750 [j.jcp.2008.11.024](https://doi.org/10.1016/j.jcp.2008.11.024).
- 751 [34] A. M. Stuart, A. L. Teckentrup, Posterior consistency for Gaussian process approximations of Bayesian
752 posterior distributions, Mathematics of Computation 87 (310) (2017) 721–753. doi:[10.1090/mcom/3244](https://doi.org/10.1090/mcom/3244).
753 URL <http://dx.doi.org/10.1090/mcom/3244>

- 754 [35] L. Yan, T. Zhou, [Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse prob-](#)
755 [lems](#), *Journal of Computational Physics* 381 (2019) 110–128. doi:[https://doi.org/10.1016/j.jcp.2018.](https://doi.org/10.1016/j.jcp.2018.12.025)
756 [12.025](https://doi.org/10.1016/j.jcp.2018.12.025).
757 URL <https://www.sciencedirect.com/science/article/pii/S0021999119300063>
- 758 [36] L. Yan, T. Zhou, [An Adaptive Surrogate Modeling Based on Deep Neural Networks for Large-Scale Bayesian](#)
759 [Inverse Problems](#), *Communications in Computational Physics* 28 (5) (2020) 2180–2205. doi:[https://doi.](https://doi.org/10.4208/cicp.0A-2020-0186)
760 [org/10.4208/cicp.0A-2020-0186](https://doi.org/10.4208/cicp.0A-2020-0186).
761 URL http://global-sci.org/intro/article_detail/cicp/18409.html
- 762 [37] D. G. Giovanis, I. Papaioannou, D. Straub, V. Papadopoulos, [Bayesian updating with subset simulation](#)
763 [using artificial neural networks](#), *Computer Methods in Applied Mechanics and Engineering* 319 (2017) 124–
764 145. doi:<https://doi.org/10.1016/j.cma.2017.02.025>.
765 URL <https://www.sciencedirect.com/science/article/pii/S0045782516311914>
- 766 [38] Z. Wang, A. Shafieezadeh, [Highly efficient Bayesian updating using metamodels: An adaptive Kriging-based](#)
767 [approach](#), *Structural Safety* 84 (2020) 101915. doi:<https://doi.org/10.1016/j.strusafe.2019.101915>.
768 URL <https://www.sciencedirect.com/science/article/pii/S0167473019303790>
- 769 [39] B. Echard, N. Gayton, M. Lemaire, [AK-MCS: An active learning reliability method combining Kriging](#)
770 [and Monte Carlo Simulation](#), *Structural Safety* 33 (2) (2011) 145–154. doi:[https://doi.org/10.1016/j.](https://doi.org/10.1016/j.strusafe.2011.01.002)
771 [strusafe.2011.01.002](https://doi.org/10.1016/j.strusafe.2011.01.002).
772 URL <https://www.sciencedirect.com/science/article/pii/S0167473011000038>
- 773 [40] R. Teixeira, M. Nogal, A. O'Connor, [Adaptive approaches in metamodel-based reliability analysis: A review](#),
774 *Structural Safety* 89 (2021) 102019. doi:<https://doi.org/10.1016/j.strusafe.2020.102019>.
775 URL <https://www.sciencedirect.com/science/article/pii/S0167473020300989>
- 776 [41] F. Cui, M. Ghosn, [Implementation of machine learning techniques into the Subset Simulation method](#),
777 *Structural Safety* 79 (2019) 12–25. doi:<https://doi.org/10.1016/j.strusafe.2019.02.002>.
778 URL <https://www.sciencedirect.com/science/article/pii/S0167473018301759>
- 779 [42] R. Schöbi, B. Sudret, J. Wiart, [Polynomial chaos based Kriging](#), *International Journal for Uncertainty*
780 [Quantification 5 \(2\) \(2015\) 171–193. doi:\[10.1615/int.j.uncertaintyquantification.2015012467\]\(https://doi.org/10.1615/int.j.uncertaintyquantification.2015012467\).
781 URL <http://dx.doi.org/10.1615/Int.J.UncertaintyQuantification.2015012467>](#)
- 782 [43] R. Schöbi, B. Sudret, S. Marelli, [Rare Event Estimation Using Polynomial-Chaos Kriging](#), *ASCE-ASME*
783 [Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering](#) 3 (2). doi:[10.1061/](https://doi.org/10.1061/ajrua6.0000870)
784 [ajrua6.0000870](https://doi.org/10.1061/ajrua6.0000870).
785 URL <http://dx.doi.org/10.1061/AJRU6.0000870>
- 786 [44] F. Lu, M. Morzfeld, X. Tu, A. J. Chorin, [Limitations of polynomial chaos expansions in the bayesian](#)
787 [solution of inverse problems](#), *Journal of Computational Physics* 282 (2015) 138–147. doi:[10.1016/j.jcp.](https://doi.org/10.1016/j.jcp.2014.11.010)
788 [2014.11.010](https://doi.org/10.1016/j.jcp.2014.11.010).

- 789 [45] M. J. Zaki, W. Meira, Jr, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge
790 University Press, 2014. doi:10.1017/CB09780511810114.
- 791 [46] O. Ditlevsen, H. Madsen, Structural reliability methods, J. Wiley and Sons, Chichester, 1996.
- 792 [47] M. Lemaire, Structural reliability, Wiley, 2009.
- 793 [48] K. M. Zuev, J. L. Beck, S.-K. Au, L. S. Katafygiotis, Bayesian post-processor and other enhancements of
794 Subset Simulation for estimating failure probabilities in high dimensions, Computers & Structures 92-93
795 (2012) 283–296. doi:https://doi.org/10.1016/j.compstruc.2011.10.017.
796 URL https://www.sciencedirect.com/science/article/pii/S0045794911002720
- 797 [49] I. Papaioannou, W. Betz, K. Zwirgmaier, D. Straub, MCMC algorithms for Subset Simulation, Probabilistic
798 Engineering Mechanics 41 (2015) 89–103. doi:https://doi.org/10.1016/j.probengmech.2015.06.006.
799 URL https://www.sciencedirect.com/science/article/pii/S0266892015300205
- 800 [50] A. Nataf, Détermination des distributions dont les marges sont données, C. R. Acad. Sci. 225 (1962) 42–43.
- 801 [51] R. Lebrun, A. Dutfoy, An innovating analysis of the Nataf transformation from the copula viewpoint, Prob-
802 abilistic Engineering Mechanics 24 (3) (2009) 312–320. doi:https://doi.org/10.1016/j.probengmech.
803 2008.08.001.
804 URL https://www.sciencedirect.com/science/article/pii/S0266892008000660
- 805 [52] M. Rosenblatt, Remarks on a Multivariate Transformation, The Annals of Mathematical Statistics 23 (3)
806 (1952) 470 – 472. doi:10.1214/aoms/1177729394.
807 URL https://doi.org/10.1214/aoms/1177729394
- 808 [53] W. Betz, J. L. Beck, I. Papaioannou, D. Straub, Bayesian inference with reliability methods without knowing
809 the maximum of the likelihood function, Probabilistic Engineering Mechanics 53 (2018) 14–22. doi:https:
810 //doi.org/10.1016/j.probengmech.2018.03.004.
811 URL https://www.sciencedirect.com/science/article/pii/S026689201730190X
- 812 [54] F. DiazDelaO, A. Garbuno-Inigo, S. Au, I. Yoshida, Bayesian updating and model class selection with
813 Subset Simulation, Computer Methods in Applied Mechanics and Engineering 317 (2017) 1102–1121. doi:
814 https://doi.org/10.1016/j.cma.2017.01.006.
815 URL https://www.sciencedirect.com/science/article/pii/S0045782516308283
- 816 [55] G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic
817 finite element analysis, Prob. Eng. Mech. 25 (2010) 183–197. doi:10.1016/j.probengmech.2009.10.003.
- 818 [56] J. D. Jakeman, F. Franzelin, A. Narayan, M. Eldred, D. Plfüger, Polynomial chaos expansions for dependent
819 random variables, Computer Methods in Applied Mechanics and Engineering 351 (2019) 643 – 666. doi:
820 https://doi.org/10.1016/j.cma.2019.03.049.
821 URL http://www.sciencedirect.com/science/article/pii/S0045782519301884

- 822 [57] R. Ghanem, P. Spanos, Stochastic finite elements - A spectral approach, Springer Verlag, 1991. doi:
823 [10.1007/978-1-4612-3094-6](https://doi.org/10.1007/978-1-4612-3094-6).
- 824 [58] D. Xiu, G. E. Karniadakis, The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations,
825 SIAM Journal on Scientific Computing 24 (2002) 619–644. doi:[10.1137/S1064827501387826](https://doi.org/10.1137/S1064827501387826).
- 826 [59] N. Wiener, The homogeneous chaos, Am. J. Math. 60 (1938) 897–936. doi:[10.2307/2371268](https://doi.org/10.2307/2371268).
- 827 [60] C. Soize, R. Ghanem, Physical Systems with Random Uncertainties: Chaos Representations with Ar-
828 bitrary Probability Measure, SIAM Journal on Scientific Computing 26 (2) (2004) 395–410. doi:
829 [10.1137/s1064827503424505](https://doi.org/10.1137/s1064827503424505).
830 URL <http://dx.doi.org/10.1137/S1064827503424505>
- 831 [61] O. G. Ernst, A. Mugler, H.-J. Starkloff, E. Ullmann, On the convergence of generalized polynomial chaos
832 expansions, ESAIM: Mathematical Modelling and Numerical Analysis 46 (2) (2012) 317–339. doi:[10.1051/m2an/2011045](https://doi.org/10.1051/m2an/2011045).
833
834 URL <https://www.esaim-m2an.org/articles/m2an/abs/2012/02/m2an110045/m2an110045.html>
- 835 [62] D. Ghiocel, R. Ghanem, Stochastic finite element analysis of seismic soil-structure interaction, Journal of
836 Engineering Mechanics 128 (2002) 66–77. doi:[10.1061/\(asce\)0733-9399\(2002\)128:1\(66\)](https://doi.org/10.1061/(asce)0733-9399(2002)128:1(66)).
- 837 [63] O. P. Le Maître, M. T. Reagan, H. N. Najm, R. G. Ghanem, O. M. Knio, A Stochastic Projection Method
838 for Fluid Flow: II. Random Process, Journal of Computational Physics 181 (1) (2002) 9 – 44. doi:<https://doi.org/10.1006/jcph.2002.7104>.
839
840 URL <http://www.sciencedirect.com/science/article/pii/S0021999102971044>
- 841 [64] D. Xiu, Efficient Collocational Approach for Parametric Uncertainty Analysis, Communications in Compu-
842 tational Physics 2 (2) (2007) 293–309.
- 843 [65] S.-K. Choi, R. V. Grandhi, R. A. Canfield, C. L. Pettit, Polynomial Chaos Expansion with Latin Hypercube
844 Sampling for Estimating Response Variability, AIAA Journal 42 (6) (2004) 1191–1198. arXiv:<https://doi.org/10.2514/1.2220>, doi:[10.2514/1.2220](https://doi.org/10.2514/1.2220).
845
846 URL <https://doi.org/10.2514/1.2220>
- 847 [66] M. Berveiller, B. Sudret, M. Lemaire, Stochastic finite element: a non intrusive approach by regression,
848 European Journal of Computational Mechanics 15 (1-3) (2006) 81–92. arXiv:<https://doi.org/10.3166/remn.15.81-92>, doi:[10.3166/remn.15.81-92](https://doi.org/10.3166/remn.15.81-92).
849
850 URL <https://doi.org/10.3166/remn.15.81-92>
- 851 [67] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on Least Angle Regression, J.
852 Comput. Phys. 230 (2011) 2345–2367. doi:[10.1016/j.jcp.2010.12.021](https://doi.org/10.1016/j.jcp.2010.12.021).
- 853 [68] D. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (4) (2006) 1289–1306. doi:
854 [10.1109/TIT.2006.871582](https://doi.org/10.1109/TIT.2006.871582).

- 855 [69] N. Lüthen, S. Marelli, B. Sudret, [Sparse Polynomial Chaos Expansions: Literature Survey and Benchmark](#),
856 SIAM/ASA Journal on Uncertainty Quantification 9 (2) (2021) 593–649. doi:10.1137/20m1315774.
857 URL <http://dx.doi.org/10.1137/20M1315774>
- 858 [70] S. Arlot, A. Celisse, A survey of cross-validation procedures for model selection, Statistics Surveys 4 (2010)
859 40–79. doi:10.1214/09-ss054.
- 860 [71] T. J. Santner, B. J. Williams, W. I. Notz, [The Design and Analysis of Computer Experiments](#), Springer
861 New York, 2003. doi:10.1007/978-1-4757-3799-8.
862 URL <http://dx.doi.org/10.1007/978-1-4757-3799-8>
- 863 [72] O. Dubrule, [Cross validation of kriging in a unique neighborhood](#), Journal of the International Association
864 for Mathematical Geology 15 (6) (1983) 687–699. doi:10.1007/BF01033232.
865 URL <https://doi.org/10.1007/BF01033232>
- 866 [73] S. Marelli, B. Sudret, [An active-learning algorithm that combines sparse polynomial chaos expansions and
867 bootstrap for structural reliability analysis](#), Structural Safety 75 (2018) 67–74. doi:https://doi.org/10.
868 1016/j.strusafe.2018.06.003.
869 URL <https://www.sciencedirect.com/science/article/pii/S0167473017302977>
- 870 [74] V. Dubourg, [Adaptive surrogate models for reliability analysis and reliability-based design optimization](#),
871 Ph.D. thesis, Universtisté Clermont-Ferrand 2 (2011).
872 URL <http://www.theses.fr/2011CLF22184>
- 873 [75] P. Angelikopoulos, C. Papadimitriou, P. Koumoutsakos, [X-TMCMC: Adaptive kriging for Bayesian inverse
874 modeling](#), Computer Methods in Applied Mechanics and Engineering 289 (2015) 409–428. doi:https:
875 //doi.org/10.1016/j.cma.2015.01.015.
876 URL <https://www.sciencedirect.com/science/article/pii/S0045782515000353>
- 877 [76] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, [The Quickhull Algorithm for Convex Hulls](#), ACM Trans. Math.
878 Softw. 22 (4) (1996) 469–483. doi:10.1145/235815.235821.
879 URL <https://doi.org/10.1145/235815.235821>
- 880 [77] B. J. Bichon, M. S. Eldred, L. P. Swiler, S. Mahadevan, J. M. McFarland, [Efficient Global Reliability
881 Analysis for Nonlinear Implicit Performance Functions](#), AIAA Journal 46 (10) (2008) 2459–2468. arXiv:
882 <https://doi.org/10.2514/1.34321>, doi:10.2514/1.34321.
883 URL <https://doi.org/10.2514/1.34321>
- 884 [78] M. Moustapha, B. Sudret, [Surrogate-assisted reliability-based design optimization: a survey and a unified
885 modular framework](#), Structural and Multidisciplinary Optimization 60 (5) (2019) 2157–2176. doi:10.1007/
886 s00158-019-02290-y.
887 URL <https://doi.org/10.1007/s00158-019-02290-y>

- 888 [79] J. Bect, D. Ginsbourger, L. Li, V. Picheny, E. Vazquez, [Sequential design of computer experiments for](#)
889 [the estimation of a probability of failure](#), *Statistics and Computing* 22 (3) (2011) 773–793. doi:10.1007/
890 [s11222-011-9241-4](#).
891 URL <http://dx.doi.org/10.1007/s11222-011-9241-4>
- 892 [80] R. Teixeira, M. Nogal, A. O'Connor, B. Martinez-Pastor, [Reliability assessment with density scanned](#)
893 [adaptive Kriging](#), *Reliability Engineering & System Safety* 199 (2020) 106908. doi:[https://doi.org/10.](https://doi.org/10.1016/j.ress.2020.106908)
894 [1016/j.ress.2020.106908](https://doi.org/10.1016/j.ress.2020.106908).
895 URL <https://www.sciencedirect.com/science/article/pii/S0951832019307434>
- 896 [81] S. Marelli, B. Sudret, UQLab: A framework for uncertainty quantification in Matlab, in: Proc. 2nd Int.
897 Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom, 2014,
898 pp. 2554–2563.
- 899 [82] S. Kullback, R. Leibler, On information and sufficiency, *Annals of Mathematical Statistics* 22 (1951) 79–86.
900 doi:10.1214/aoms/1177729694.
- 901 [83] J. Lin, Divergence measures based on the Shannon entropy, *IEEE Transactions on Information Theory*
902 37 (1) (1991) 145–151. doi:10.1109/18.61115.
- 903 [84] M. Moustapha, S. Marelli, B. Sudret, [A generalized framework for active learning reliability: survey and](#)
904 [benchmark](#), arXiv:2106.01713 [stat]ArXiv: 2106.01713.
905 URL <http://arxiv.org/abs/2106.01713>
- 906 [85] J. L. Beck, S.-K. Au, Bayesian Updating of Structural Models and Reliability using Markov Chain
907 Monte Carlo Simulation, *Journal of Engineering Mechanics* 128 (4) (2002) 380–391. doi:10.1061/(ASCE)
908 [0733-9399\(2002\)128:4\(380\)](#).
- 909 [86] J. L. Beck, K. M. Zuev, Asymptotically independent Markov sampling: a new Markov Chain Monte Carlo
910 scheme for Bayesian inference, *International Journal for Uncertainty Quantification* 3 (5) (2013) 445–474.
911 doi:10.1615/Int.J.UncertaintyQuantification.2012004713.
- 912 [87] M. Loève, *Probability theory*, Springer Verlag, 1977. doi:10.1007/978-1-4757-6288-4.
- 913 [88] J. L. Beck, K.-V. Yuen, Model Selection Using Response Measurements: Bayesian Probabilistic Approach,
914 *Journal of Engineering Mechanics* 130 (2) (2004) 192–203. doi:10.1061/(ASCE)0733-9399(2004)130:
915 [2\(192\)](#).