Bayesian inference with reliability methods without knowing the maximum of the likelihood function

Wolfgang Betz[a,*], James L. Beck[b], Iason Papaioannou[a], Daniel Straub[a]

[a] *Engineering Risk Analysis Group, Technische Universität München, 80333 München, Germany*
[b] *Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125*

Abstract

In the BUS (Bayesian Updating with Structural reliability methods) approach, the uncertain parameter space is augmented by a uniform random variable and the Bayesian inference problem is interpreted as a structural reliability problem. A posterior sample is given by an augmented vector sample within the failure domain of the structural reliability problem where the realization of the uniform random variable is smaller than the likelihood function scaled by a constant c. The constant c must be selected such that $1/c$ is larger or equal than the maximum of the likelihood function, which, however, is typically unknown a-priori. For BUS combined with sampling based reliability methods, choosing c too small has a negative impact on the computational efficiency. To overcome the problem of selecting c, we propose a post-processing step for BUS that returns an unbiased estimate for the evidence and samples from the posterior distribution, even if $1/c$ is selected smaller than the maximum of the likelihood function. The applicability of the proposed post-processing step is demonstrated by means of rejection sampling. However, it can be combined with any structural reliability method applied within the BUS framework.

Keywords: Bayesian updating, Bayesian model class selection, rejection sampling, structural reliability

1. Introduction

The Bayesian learning process is formalized in *Bayes' theorem*:

$$p(\boldsymbol{\theta}|\mathbf{d}) = {c_{\mathrm{E}}}^{-1} \cdot L(\boldsymbol{\theta}|\mathbf{d}) \cdot p(\boldsymbol{\theta}) \tag{1}$$

The prior probabilistic model $p(\boldsymbol{\theta})$ represents the current state of knowledge about the vector of uncertain model parameters $\boldsymbol{\theta} \in \mathbb{R}^M$. Information (data) $\mathbf{d}$ that becomes available in form of *measurements or observations* is expressed in the likelihood function $L(\boldsymbol{\theta}|\mathbf{d}) = p(\mathbf{d}|\boldsymbol{\theta})$. The posterior probabilistic model $p(\boldsymbol{\theta}|\mathbf{d})$ describes the combined knowledge from the prior distribution and the likelihood function. Consequently, the posterior uncertainty about $\boldsymbol{\theta}$ is reduced compared to its prior uncertainty. The constant c_{E} is a normalizing scalar that is called the *evidence* (marginal likelihood); it measures the plausibility of the assumed *stochastic model class* [1] and is defined as:

$$c_{\mathrm{E}} = \int_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\mathbf{d}) \cdot p(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \tag{2}$$

The evidence c_{E} is required for Bayesian model class selection and model averaging [2, 3].

Except for some special cases, the posterior distribution usually cannot be derived analytically, and posterior samples have to be generated numerically. However, it is typically challenging to compute the evidence c_{E}, because of the multi-dimensional integral in Eq. (2). The BUS approach [4] transforms the Bayesian inference problem to a structural reliability problem. In structural reliability, probabilities of rare events are estimated [5, 6, 7]. By interpreting the Bayesian updating problem as a rare event estimation, existing structural reliability methods can be used to perform the Bayesian analysis (see e.g., [8]). An estimate for the evidence c_{E} is obtained as a by-product of BUS; in BUS, the evidence is directly related to the rare event probability.

In BUS, prior to the analysis, a constant c has to be selected: On the one hand, c^{-1} should not be smaller than the maximum value that the likelihood function can take; i.e., $c^{-1} \geq L_{\max}$. On the other hand, selecting c^{-1} conservatively large decreases the efficiency of the method. Therefore, an appropriate choice of c^{-1} is crucial. However, in many cases the maximum of the likelihood function is not known in advance.

We propose a post-processing step for BUS that returns an unbiased estimate for the evidence and samples from the posterior distribution, even if $1/c$ is selected smaller than the maximum of the likelihood function. Like the original BUS approach, the proposed extension of BUS can be combined with any structural reliability method. The applicability of the proposed post-processing step is demonstrated by means of the standard Monte Carlo method, which corresponds to the rejection sampling approach when combined with BUS.

The structure of the paper is as follows: In Section 2 the BUS approach is formally introduced and the role of the scaling constant c in BUS is discussed. In Section 3 the post-processing

*Corresponding author
Email addresses: wolfgang.betz@tum.de (Wolfgang Betz), jimbeck@caltech.edu (James L. Beck), iason.papaioannou@tum.de (Iason Papaioannou), straub@tum.de (Daniel Straub)

step for BUS is proposed that can cope with $c \cdot L_{\max} < 1$. In Section 4 the proposed post-processing step is investigated in combination with rejection sampling for different example problems. Section 5 briefly summarizes the obtained findings.

2. Bayesian updating with structural reliability methods

2.1. The idea behind BUS

Straub and Papaioannou show in [4] that a Bayesian updating problem can be interpreted as a structural reliability problem. The principal idea behind BUS (Bayesian Updating with Structural reliability methods) is to augment the space of random variables spanned by $\boldsymbol{\theta}$ with an additional uniformly distributed random variable π with support $[0, 1]$. The updating problem is then expressed as a structural reliability problem in the so-obtained augmented random variable space $[\boldsymbol{\theta}, \pi]$. The "failure" domain Ω of this reliability problem is defined as:

$$\Omega = \{\pi \leq c \cdot L(\boldsymbol{\theta}|\mathbf{d})\} \tag{3}$$

where c is a positive constant chosen such that $c \cdot L(\boldsymbol{\theta}|\mathbf{d}) \leq 1$ is maintained for all $\boldsymbol{\theta}$. The domain Ω is illustrated in Fig. 1. Note that Ω denotes both the failure domain and the corresponding event. The link between the domain Ω and the actual Bayesian updating problem is: Samples from the prior distribution of $\boldsymbol{\theta}$ that are in Ω follow the posterior distribution [4]. In reliability analysis, the limit-state function is defined such that: $g(\boldsymbol{\theta}, \pi) \leq 0$ if $[\boldsymbol{\theta}, \pi] \in \Omega$; and $g(\boldsymbol{\theta}, \pi) > 0$ if $[\boldsymbol{\theta}, \pi]$ is outside of Ω (see Fig. 1). A limit-state function $g(\boldsymbol{\theta}, \pi)$ that describes the "failure" domain Ω defined in Eq. (3) is:

$$g(\boldsymbol{\theta}, \pi) = \pi - c \cdot L(\boldsymbol{\theta}|\mathbf{d}) \tag{4}$$

Optimally, the constant c should be chosen as the reciprocal of the maximum of the likelihood function, denoted $L_{\max}$ [4]. However, $L_{\max}$ is not always known in advance. In such cases, it is difficult to select c appropriately. The implications of choosing c too large or too small are discussed in Section 2.5.

2.2. Estimating the evidence in BUS

An estimate for the evidence c_{E} is obtained as a by-product of BUS. Let p_Ω be the probability that samples $[\boldsymbol{\theta}, \pi]$ from the prior distribution fall into Ω, i.e.:

$$p_\Omega = \Pr[\Omega] = \Pr[g(\boldsymbol{\theta}, \pi) \leq 0] \tag{5}$$

p_Ω is the target quantity of interest in a reliability analysis and called the *probability of failure*. In BUS, p_Ω is directly linked to the evidence c_{E} through the constant c [4]:

$$c_{\mathrm{E}} = \frac{p_\Omega}{c} \tag{6}$$

Note that some reliability methods allow evaluating uncertainty bounds for the estimate of p_Ω. In this case, the statistical uncertainty in the estimated evidence c_{E} can be quantified directly, as the evidence is directly proportional to p_Ω. Uncertainty bounds for the estimated reliability can be evaluated with most simulation methods. For Monte Carlo simulation this is straight-forward to do. Moreover, approximate bounds can be derived for any importance sampling method employing a normal approximation of the estimator (e.g. [9]). For Subset Simulation, some bounds were proposed in [10, 11].

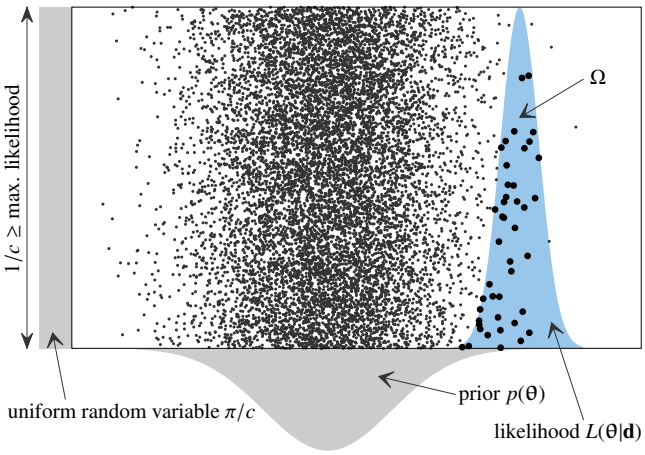

Figure 1: Illustration of the principle of the rejection sampling algorithm – *Algorithm (1)*. The highlighted region is the domain Ω defined in Eq. (3). The limit-state function $g(\boldsymbol{\theta}, \pi)$ introduced in Eq. (4) is smaller or equal than *zero* within Ω (it is *zero* at the boundary of Ω), and larger than *zero* outside of Ω. Samples "below" the likelihood (i.e., the samples contained in Ω) are independent samples from the posterior distribution. In this example, 43 out of 10^4 samples are accepted.

2.3. Structural reliability methods in BUS

BUS employs structural reliability methods to perform Bayesian updating. The most straight-forward application of BUS is rejection sampling – which corresponds to crude Monte Carlo simulation in the context of structural reliability. Rejection sampling within BUS is explained in detail in Section 2.4. As was pointed out in [4], it is possible to use other structural reliability methods instead of the simple rejection sampling algorithm (i.e., instead of a Monte Carlo simulation). BUS is often combined with Subset Simulation (see for example [4, 12, 13, 14, 15, 16]), because this method is efficient for very small failure probabilities and its performance does not depend on the dimension M of the vector of uncertain model parameters $\boldsymbol{\theta}$. Apart from rejection sampling and Subset Simulation, the BUS approach has been combined with the first order reliability method (FORM) and line sampling in [8]. FORM solves the reliability problem only approximately by linearizing the limit-state function at the most probable point of failure [17, 18]. The line sampling method computes a correction factor for the linearized solution by performing a specified number of line searches parallel to an important direction pointing to the failure surface [19, 20, 21].

To generate realizations of the posterior, an additional post-processing step is required: besides computing the probability of failure, samples located in Ω have to be returned. In Monte Carlo simulation and Subset Simulation, samples located in Ω are directly generated during the reliability analysis. Importance sampling based reliability methods require an additional

re-sampling step based on the importance weights associated with the "failed" samples to produce samples of equal weights. In case of FORM, samples of the approximated "failure" domain can be easily generated. Samples from the posterior distribution can be further used for posterior prediction of quantities of interest. A special application is the use of BUS for updating the probability of rare events based on observed system response. In such case, as the target quantity of interest is the posterior probability of failure, no posterior samples have to be generated and the updating problem can be directly solved by structural reliability methods [8, 22].

2.4. The rejection sampling algorithm

The most trivial application of BUS results in the *rejection sampling* algorithm [4, 23] (see *Algorithm (1)* below). This algorithm repeatedly proposes a sample $[\tilde{\theta}, \tilde{\pi}]$ from the prior distribution and accepts the sample if it is located in the "failure" domain; i.e., if $[\tilde{\theta}, \tilde{\pi}] \in \Omega$. The accepted sample $\tilde{\theta}$ is a sample from the posterior distribution. The algorithm is repeated until K posterior samples are generated. The posterior samples resulting from the rejection sampling algorithm are statistically independent.

The quantity p_Ω is the probability that a proposed sample is accepted; p_Ω is also referred to as the *acceptance probability*. An unbiased estimate $\hat{p}_\Omega$ of p_Ω is [24]:

$$p_\Omega \approx \hat{p}_\Omega = \frac{K-1}{n-1} \tag{7}$$

where n is the total number of prior samples that were proposed to generate K posterior samples. Note that the estimator K/n produces a biased estimate for p_Ω [24]. An unbiased estimate of the variance of $\hat{p}_\Omega$ is [25]:

$$\mathrm{Var}\left[\hat{p}_\Omega\right] \approx \frac{(1-\hat{p}_\Omega)\,\hat{p}_\Omega}{n-2} \tag{8}$$

The estimates given in Eqs. (7) and (8) are frequentist estimates. To appropriately quantify the uncertainty about p_Ω based on the outcome of a particular run of rejection sampling, we employ a Bayesian approach. The number n of prior samples needed to generate K posterior samples in the rejection sampling algorithm follows a negative binomial distribution. Thus, having observed a certain n for a given K, the likelihood of p_Ω is:

$$L(p_\Omega|n,K) = \binom{n-1}{K-1}(p_\Omega)^K(1-p_\Omega)^{n-K} \tag{9}$$

where $\binom{n-1}{K-1}$ denotes the binomial coefficient. The beta distribution acts as conjugate prior for the Bayesian updating. If one selects the prior distribution for p_Ω based on the Principle of Maximum Information Entropy [26, 27] with only normalization as constraint, then the resulting prior distribution is the uniform distribution on $[0, 1]$, which is a special case of the beta distribution. In this case, the posterior is also a beta distribution and can be expressed as:

$$p(p_\Omega|K,n) = \frac{{p_\Omega}^K \cdot (1-p_\Omega)^{n-K}}{\mathrm{B}(K+1, n-K+1)} \tag{10}$$

where B denotes the *beta function*. The most probable value of p_Ω a posteriori is K/n. The expectation of p_Ω is:

$$\mathrm{E}\left[p_\Omega|K,n\right] = \frac{K+1}{n+2} \tag{11}$$

The variance of the distribution in Eq. (10) can be derived as:

$$\mathrm{Var}\left[p_\Omega|K,n\right] = \frac{(K+1)\cdot(n-K+1)}{(n+2)^2\cdot(n+3)} \tag{12}$$

For increasing K and n, Eqs. (7) and (11), as well as Eqs. (8) and (12), approach the same values.

Algorithm (1): rejection sampling

As input the algorithm requires:

- K, the total number of samples to draw from the posterior distribution.
- c, selected such that $c^{-1} \geq L_{\max}$.

The algorithm evaluates the evidence c_E and returns K unweighted and statistically independent posterior samples $\theta_{(k)}$ with $k = 1, \ldots, K$.

1. Initialize counters $k = 1$ and $n = 0$.
2. **while** $(k \leq K)$ **do**:
   (a) Propose sample $[\tilde{\theta}, \tilde{\pi}]$:
      i. Draw $\tilde{\theta}$ from the prior distribution $p(\theta)$.
      ii. Draw $\tilde{\pi}$ from the uniform distribution that has support $[0, 1]$.
   (b) **if** $\left(g(\tilde{\theta}, \tilde{\pi}) \leq 0\right)$ **then**:
      i. Increase the counter $k = k + 1$.
      ii. Accept the proposed sample $\tilde{\theta}$ as a posterior sample, i.e.:
      set $\theta_{(k)} = \tilde{\theta}$.
   (c) Increase the counter $n = n + 1$.
3. Estimate p_Ω by means of Eq. (7) or Eq. (10).
4. Evaluate the evidence $c_\mathrm{E} = p_\Omega / c$

On average, the algorithm requires K/p_Ω samples from the prior distribution to generate K samples from the posterior distribution. The principle of the *rejection sampling* algorithm is illustrated in Fig. 1. Note that *Algorithm (1)* is similar to a Monte Carlo simulation for solving the structural reliability problem that has limit-state function $g(\theta, \pi)$ and random variables $[\theta, \pi]$. The difference is that in a Monte Carlo simulation typically the total number of samples n is fixed whereas in *Algorithm (1)* the number K of samples to be generated in the domain Ω is specified.

The main advantage of rejection sampling is that it produces independent samples from the posterior distribution. However, if the posterior distribution does not match the prior distribution well, p_Ω becomes small, which renders the rejection sampling algorithm inefficient. As a consequence, rejection sampling is typically inefficient and usually more advanced reliability methods are employed – see Section 2.3. However, because rejection sampling is straightforward and easy to understand, we use it here to demonstrate the consequences of choosing the constant c such that $c^{-1} < L_{\max}$.

2.5. The constant c in BUS

An appropriate selection of the constant c is crucial in BUS. The constant c in BUS is motivated by rejection sampling. For the equivalent rejection sampling problem targeted in BUS, the value of $c^{-1} \cdot {c_\mathrm{E}}^{-1}$ is a scaling factor for the proposal distribution $p(\Theta)$ that must be chosen such that the scaled proposal distribution is an envelope function for the target distribution [28]; i.e., $c^{-1} \cdot {c_\mathrm{E}}^{-1} \cdot p(\Theta) \geq p(\Theta|\mathbf{d})$ must be maintained for all Θ, which reduces to $c^{-1} \geq L_\mathrm{max}$ if $p(\Theta|\mathbf{d})$ is replaced with Eq. (1). On the one hand, if a value of c^{-1} that is larger than the maximum L_max of the likelihood function is selected, the efficiency of the approach decreases; the acceptance probability p_Ω of BUS decreases linearly with c, as seen from Eq. (6). On the other hand, if a value of c^{-1} that is smaller than the maximum L_max of the likelihood function is selected, the scaled proposal distribution does not envelope the target distribution for all Θ and, thus, BUS does not produce samples that follow the posterior distribution. Instead, in such cases, BUS produces samples that follow a distribution $p^{(c)}_\mathrm{trunc}$ that is proportional to the part of the target distribution that is contained within the envelope of the proposal distribution. The distribution $p^{(c)}_\mathrm{trunc}$ can be derived as:

$$p^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d}) = \frac{L^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d}) \cdot p(\Theta)}{{c_\mathrm{E}}^{(c)}} = \frac{c \cdot L^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d}) \cdot p(\Theta)}{p^{(c)}_\Omega} \tag{13}$$

where ${c_\mathrm{E}}^{(c)} = p^{(c)}_\Omega / c$ and $p^{(c)}_\Omega$ are the associated evidence and acceptance probability that correspond to the selected $c^{-1} < L_\mathrm{max}$. $L^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d})$ is defined as:

$$L^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d}) = \min\left(L(\Theta|\mathbf{d}), c^{-1}\right) \tag{14}$$

Let $w^{(c)}$ be a correction factor such that

$$c_\mathrm{E} = w^{(c)} \cdot {c_\mathrm{E}}^{(c)} \tag{15}$$

One can expand the correction factor $w^{(c)}$:

$$\begin{aligned} w^{(c)} &= \frac{c_\mathrm{E}}{{c_\mathrm{E}}^{(c)}} && (16)\\ &= \frac{\int_\Theta L(\Theta|\mathbf{d})\, p(\Theta)\, \mathrm{d}\Theta}{{c_\mathrm{E}}^{(c)}} && (17)\\ &= \int_\Theta \frac{L(\Theta|\mathbf{d})}{L^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d})} \cdot \frac{L^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d}) \cdot p(\Theta)}{{c_\mathrm{E}}^{(c)}}\, \mathrm{d}\Theta && (18)\\ &= \int_\Theta \frac{L(\Theta|\mathbf{d})}{L^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d})} \cdot p^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d})\, \mathrm{d}\Theta && (19) \end{aligned}$$

The quantity defined in Eq. (19) can be estimated from the samples $\{\Theta^{(c)}_{(k)}\}_{k=1,\dots,K}$ generated with BUS and $c^{-1} < L_\mathrm{max}$, which follow distribution $p^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d})$:

$$\begin{aligned} w^{(c)} &\approx \frac{1}{K} \sum_{k=1}^{K} \frac{L(\Theta^{(c)}_{(k)}|\mathbf{d})}{L^{(c)}_\mathrm{trunc}(\Theta^{(c)}_{(k)}|\mathbf{d})} && (20)\\ &= \frac{1}{K} \sum_{k=1}^{K} \max\left(1, c \cdot L(\Theta^{(c)}_{(k)}|\mathbf{d})\right) && (21) \end{aligned}$$

Using Eq. (15) and the estimate for $w^{(c)}$ given in Eq. (20), one can correct the evidence computed with $c^{-1} < L_\mathrm{max}$. Let p_Ω be the probability of Eq. (5) for a choice of $c = 1/L_\mathrm{max}$. From Eq. (16), it follows that

$$p_\Omega = p^{(c)}_\Omega \cdot w^{(c)} \cdot \frac{c^{-1}}{L_\mathrm{max}} \tag{22}$$

where ${c_\mathrm{E}}^{(c)} = p^{(c)}_\Omega / c$ and $c_\mathrm{E} = p_\Omega \cdot L_\mathrm{max}$ is used (Eq. (6) does not hold if $c^{-1} < L_\mathrm{max}$). Note that $c^{-1} \leq L_\mathrm{max}$ implies that $p_\Omega \leq p^{(c)}_\Omega$, as the relative size of the failure domain increases with decreasing c^{-1} according to Eq. (3) ($c^{-1} \to 0$ implies $p^{(c)}_\Omega \to 1$). For $c^{-1} > L_\mathrm{max}$, $p_\Omega > p^{(c)}_\Omega$ implies $p_\Omega = p^{(c)}_\Omega \cdot c^{-1}/L_\mathrm{max}$. Moreover, $L(\Theta^{(c)}_{(k)}|\mathbf{d}) \geq L^{(c)}_\mathrm{trunc}(\Theta^{(c)}_{(k)}|\mathbf{d})$ clearly holds independent of c, due to Eq. (14). Thus, the values that $w^{(c)}$ can take are bounded: $1 \leq w^{(c)} \leq c \cdot L_\mathrm{max}$. For $c^{-1} \geq L_\mathrm{max}$ the correction factor is $w^{(c)} = 1$. For $c^{-1} \leq L_\mathrm{max}$ the correction factor must be smaller than $c \cdot L_\mathrm{max}$ in order to maintain $p_\Omega \leq p^{(c)}_\Omega$ (see Eq. (22)). Additionally, it is clear that ${c_\mathrm{E}}^{(c)}$ underestimates the actual evidence c_E.

The relative error in the evidence associated with selecting c^{-1} too small (i.e., $c^{-1} < L_\mathrm{max}$) is:

$$\begin{aligned} \varepsilon^{(c)}_{c_\mathrm{E}} &= 1 - \frac{1}{w^{(c)}} = 1 - \frac{{c_\mathrm{E}}^{(c)}}{c_\mathrm{E}} = 1 - \frac{p^{(c)}_\Omega}{p_\Omega} \frac{c^{-1}}{L_\mathrm{max}} && (23)\\ &= 1 - \int_\Theta \frac{L^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d})}{L(\Theta|\mathbf{d})} \cdot \frac{L(\Theta|\mathbf{d}) \cdot p(\Theta)}{c_\mathrm{E}}\, \mathrm{d}\Theta && (24)\\ &= 1 - \mathrm{E}_{\Theta|\mathbf{d}}\left[\frac{L^{(c)}_\mathrm{trunc}(\Theta|\mathbf{d})}{L(\Theta|\mathbf{d})}\right] && (25) \end{aligned}$$

For $c^{-1} \to 0$, the error $\varepsilon^{(\infty)}_{c_\mathrm{E}}$ approaches *one* and the generated samples follow the prior distribution. For $c^{-1} = L_\mathrm{max}$, it is $\varepsilon^{(1/L_\mathrm{max})}_{c_\mathrm{E}} = 0$ and the samples follow the posterior distribution.

In the following, post-processing steps for BUS are proposed, that allow c^{-1} in BUS to be chosen smaller than the maximum of the likelihood function. The proposed post-processing steps are applicable to all reliability methods used within the BUS framework. However, the performance of the proposed approach still depends strongly on the selected c. For the particular combination of BUS with Subset Simulation, the authors have recently proposed a different strategy [12], denoted aBUS, to handle the constant c in BUS. In aBUS, the constant c does not have to be specified initially and is learned adaptively during the simulation. If Subset Simulation is used as reliability method in BUS, we recommend to use the aBUS approach, as preliminary comparative numerical studies of performance indicated aBUS to be more efficient than the post-processing steps proposed in the following.

3. Proposed post-processing steps for BUS

3.1. Evidence

As discussed in the previous section, BUS does not return samples from the posterior distribution if the constant c^{-1} is

selected smaller than $L_{\max}$. However, employing Eqs. (16) and (20), the evidence of the investigated problem can be estimated even if c is not selected properly. Note that Eq. (20) does not require $L_{\max}$ to be known.

3.2. Posterior samples

Since the aim is to obtain samples from the posterior distribution, the distribution of the generated samples needs to be corrected, as the samples produced by BUS with $c^{-1} < L_{\max}$ follow the distribution $p^{(c)}_{\text{trunc}}(\Theta|\mathbf{d})$ in Eq. (13). One viable strategy is to consider the samples from distribution $p^{(c)}_{\text{trunc}}(\Theta|\mathbf{d})$ as weighted samples from the posterior distribution. In this case, one can directly work with the weighted posterior samples or perform a re-sampling step based on the sample weights to get unweighted samples that asymptotically follow the posterior distribution. Alternatively, a Metropolis-Hastings [29, 30] step can be added to *Algorithm (1)* to ensure that the generated samples follow the posterior distribution asymptotically. We refer to this algorithm as *M-H rejection sampling – Algorithm (2)*; it is based on an algorithm presented in [31].

In this contribution, we focus on generating unweighted posterior samples. More specifically, K unweighted posterior samples are generated from K weighted posterior samples. Using this approach, information is lost: samples with large weights tend to be repeated and samples with small weights tend to be discarded. If applicable, the following two strategies can be more efficient instead: (i) Instead of generating unweighted posterior samples, the weighted posterior samples obtained with BUS and $c^{-1} < L_{\max}$ are directly used. (ii) Based on K weighted posterior samples, considerably more than K unweighted posterior samples are generated by means of a re-sampling strategy.

3.3. M-H rejection sampling algorithm

The initial part of the proposed algorithm (Steps (1) and (2) in *Algorithm (2)*), which produces samples from distribution $p^{(c)}_{\text{trunc}}(\Theta|\mathbf{d})$, is equivalent to rejection sampling (*Algorithm (1)*). These Steps (1) and (2) can in principle be replaced by any reliability method. In case the applied reliability method produces weighted samples from $p^{(c)}_{\text{trunc}}(\Theta|\mathbf{d})$, the weights of Eq. (20) need to be multiplied by the weights resulting from the reliability analysis. In Step (3), the evidence is computed based on the correction factor $w^{(c)}$ derived in Section 2.5. If (i) either c_{E} or p_Ω are the quantities of interest, or (ii) weighted samples from the posterior distribution are sufficient, the procedure can be terminated at this point (i.e., after Step (3)). Assuming that the applied reliability method resulted in unweighted samples from $p^{(c)}_{\text{trunc}}(\Theta|\mathbf{d})$, the importance weight of the jth generated sample is $L(\Theta^{(c)}_{(j)}|\mathbf{d})/L^{(c)}_{\text{trunc}}(\Theta^{(c)}_{(j)}|\mathbf{d})$. Since the aim is to generate posterior samples with equal weights, the following two additional steps are performed (alternatively, a re-sampling step could be used): In Step (4), an initial seed is selected to initiate the Metropolis-Hastings algorithm. This seed is selected through a re-sampling step such that it asymptotically follows the posterior distribution. Finally, in Step (5), the distribution of the samples is corrected by means of MCMC, employing the Metropolis-Hastings algorithm proposed in [31]. As the set of samples generated in *steps (1) and (2)* are used as candidate samples of the MCMC algorithm, the employed proposal distribution is $p^{(c)}_{\text{trunc}}(\Theta|\mathbf{d})$. Therefore, the Metropolis-Hastings accept/reject-ratio of the kth candidate sample $\Theta^{(c)}_{(k)}$ is:

$$r_k = \min\left(1, \frac{p\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right)}{p\left(\Theta_{(k-1)}|\mathbf{d}\right)} \cdot \frac{p^{(c)}_{\text{trunc}}\left(\Theta_{(k-1)}|\mathbf{d}\right)}{p^{(c)}_{\text{trunc}}\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right)}\right) \tag{26}$$

$$= \min\left(1, \frac{L\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right)}{L\left(\Theta_{(k-1)}|\mathbf{d}\right)} \cdot \frac{L^{(c)}_{\text{trunc}}\left(\Theta_{(k-1)}|\mathbf{d}\right)}{L^{(c)}_{\text{trunc}}\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right)}\right) \tag{27}$$

where $k \in \{2, \ldots, K\}$. The candidate sample is accepted with probability r_k and rejected with probability $1 - r_k$. It is a standard result that this Metropolis-Hastings algorithm has stationary distribution $p(\Theta|\mathbf{d})$ and, thus, the generated samples asymptotically follow the posterior distribution. The ratio r_k can be simplified to:

$$r_k = \min\left(1, \frac{\max\left(L\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right), c^{-1}\right)}{L\left(\Theta_{(k-1)}|\mathbf{d}\right)}\right) \tag{28}$$

This can be readily shown by distinguishing four different cases:

Case (1): $L\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right) \leq c^{-1}$ and $L\left(\Theta_{(k-1)}|\mathbf{d}\right) \leq c^{-1}$
$L(\Theta|\mathbf{d}) = L^{(c)}_{\text{trunc}}(\Theta|\mathbf{d})$ holds and, consequently, both Eq. (27) and Eq. (28) become *one*.

Case (2): $L\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right) > c^{-1}$ and $L\left(\Theta_{(k-1)}|\mathbf{d}\right) \leq c^{-1}$
As in the previous case, Eqs. (27) and (28) become *one*.

Case (3): $L\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right) \leq c^{-1}$ and $L\left(\Theta_{(k-1)}|\mathbf{d}\right) > c^{-1}$
Eqs. (27) and (28) result in $c^{-1}/L\left(\Theta_{(k-1)}|\mathbf{d}\right)$.

Case (4): $L\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right) > c^{-1}$ and $L\left(\Theta_{(k-1)}|\mathbf{d}\right) > c^{-1}$
$L^{(c)}_{\text{trunc}}(\Theta|\mathbf{d}) = c^{-1}$ ans so both Eq. (27) and Eq. (28) translate to $\min\left(1, L\left(\Theta^{(c)}_{(k)}|\mathbf{d}\right)/L\left(\Theta_{(k-1)}|\mathbf{d}\right)\right)$.

Consequently, Eqs. (27) and (28) are indeed equivalent.

The burn-in period of the Markov chain in Step (5) is very short for reasonably large K, as the initial seed asymptotically follows the stationary distribution of the chain. Note that the likelihood function needs to be evaluated only in Step (2) of the algorithm. If the evaluation of the model behind the likelihood is computationally demanding, the main computational burden of the algorithm is in Step (2). All other steps have a small impact on the run-time in this case.

Algorithm (2): M-H rejection sampling

As input the algorithm requires:

- K, the total number of samples to draw from the posterior distribution.
- c, where c^{-1} may be selected smaller than $L_{\max}$.

The algorithm evaluates the evidence c_{E} and returns K unweighted but dependent posterior samples $\theta_{(k)}$ with $k = 1, \ldots, K$.

1. Initialize counters $k = 1$ and $n = 0$.
2. **while** $(k \leq K)$ **do**:
   (a) Propose sample $[\tilde{\theta}, \tilde{\pi}]$:
      i. Draw $\tilde{\theta}$ from the prior distribution $p(\theta)$.
      ii. Draw $\tilde{\pi}$ from the uniform distribution that has support $[0, 1]$.
   (b) **if** $\left(g(\tilde{\theta}, \tilde{\pi}) \leq 0\right)$ **then**:
      i. Accept the proposed sample $\tilde{\theta}$ as a posterior sample of the truncated posterior distribution, i.e.:
      set $\theta_{(k)}^{(c)} = \tilde{\theta}$.
      ii. Increase the counter $k = k + 1$.
   (c) Increase the counter $n = n + 1$.
3. Evaluate the evidence c_{E}
   (a) Estimate $p_{\Omega}^{(c)}$ by means of Eq. (7) or Eq. (10).
   (b) Evaluate the quantity $w^{(c)}$, defined in Eq. (20).
   (c) An estimate $\hat{c}_{\mathrm{E},K}$ of c_{E} is obtained through Eq. (15), where ${c_{\mathrm{E}}}^{(c)} = p_{\Omega}^{(c)}/c$.
4. Pick sample $\theta_{(1)}$:
   (a) Resampling step: draw index i randomly from the list $\{1, \ldots, K\}$ where the jth element of the list is associated with probability $\frac{L(\theta_{(j)}^{(c)}|\mathbf{d})}{L_{\mathrm{trunc}}^{(c)}(\theta_{(j)}^{(c)}|\mathbf{d})} / \left(w^{(c)} \cdot K\right)$.
   (b) **swap** $\theta_{(i)}^{(c)}$ with $\theta_{(1)}^{(c)}$
   (c) Set $\theta_{(1)} = \theta_{(1)}^{(c)}$
   (d) Set $k = 2$
5. **while** $(k \leq K)$ **do**:
   (a) Draw a sample u randomly from a uniform distribution with support $[0, 1]$.
   (b) Evaluate the accept/reject-ratio r_k defined in Eq. (28); i.e.:
   $r_k = \min\left(1, \frac{\max\left(L\left(\theta_{(k)}^{(c)}\middle|\mathbf{d}\right), c^{-1}\right)}{L\left(\theta_{(k-1)}\middle|\mathbf{d}\right)}\right)$
   (c) **if** $(u \leq r_k)$ **then** accept the candidate sample:
   Set $\theta_{(k)} = \theta_{(k)}^{(c)}$
   else; i.e., if $(u > r_k)$ then reject the candidate sample:
   Set $\theta_{(k)} = \theta_{(k-1)}$.
   (d) Increase the counter $k = k + 1$.

The principle of *Algorithm (2)* is illustrated in Fig. 2. Contrary to *rejection sampling*, this algorithm does not produce K independent samples. Instead, the K generated posterior samples are dependent, because of the acceptance/rejection step in the Metropolis-Hastings algorithm (*steps (5b) and (5c)* in *Algorithm (2)*). The smaller the corresponding acceptance rate, the larger is the induced dependency. A particular advantage of *Algorithm (2)* compared to *Algorithm (1)* is that *Algorithm (2)* can be applied without knowing $L_{\max}$. However, the selected c still has a considerable influence on the efficiency of the algorithm. General guidelines on how to select c are given in [4]. A potential strategy to select the value of c in *Algorithm (2)* is to perform an initial Monte Carlo simulation from the prior distribution with a small number of samples, and to select c^{-1} equal to the largest observed likelihood value. The initially generated Monte Carlo samples can be re-used in *Algorithm (2)*.

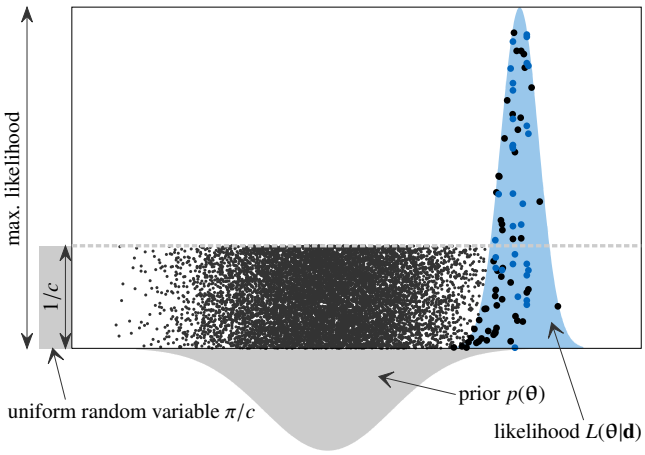

Figure 2: Illustration of the principle of the *M-H rejection sampling* algorithm – *Algorithm (2)*. Note that only the *black* samples are independent; the *blue* samples introduce a dependency (at least one *black* sample has the same θ as a *blue* sample) that decreases the efficiency of the sampling algorithm. For clarity, the ordinate of a sample θ in Ω is selected randomly between *zero* and $L(\theta|\mathbf{d})$.

Note that the performance of the proposed post-processing step does not depend on the dimension of θ. The actual location of a generated sample $\theta_{(k)}$ is not used, only the associated likelihood is considered.

4. Numerical investigations

In this section, the performance of *Algorithm (2)* is investigated by means of three example problems.

4.1. Definitions

For the discussion of the example problems, the following quantities are introduced:

- b acts as a normalized version of c^{-1}: Let $b \in (0, 1]$ be defined as $b = cL_{\max}$; i.e., $b = 1 \Leftrightarrow c^{-1} = L_{\max}$ and $b = 0 \Leftrightarrow c^{-1} = 0$. The performance of the investigated algorithm is assessed for different values of b.

- $b_{10^3,\max}$ represents the largest observed likelihood multiplied with c in a set of 10^3 independent posterior samples. Note that $b_{10^3,\max}$ is a stochastic quantity.

- $c_{\mathrm{E,ref}}$ denotes the actual value of the evidence of the example problem. The quantity $p_{\Omega,\mathrm{ref}}$ is defined as $p_{\Omega,\mathrm{ref}} = c_{\mathrm{E,ref}}/L_{\max}$.

- $\hat{c}_{\mathrm{E},K}$ is the evidence estimated by the investigated algorithm based on K posterior samples. Moreover, $\hat{c}_{\mathrm{E},K}^{(c)}$ denotes the evidence estimated with *Algorithm (1)* and scaling constant c.

– a_K and s_K denote the estimated mean and standard deviation of the first component of the parameter vector in a set of K posterior samples. Note that a_K and s_K are random variables for finite K. If the investigated algorithm produces posterior samples, we have $\mathrm{E}[a_K] = \mathrm{E}[\theta_1|\mathbf{d}]$ and $\mathrm{E}[s_K] = \sigma[\theta_1|\mathbf{d}]$. If the generated posterior samples are independent, then $\sigma[a_K] = \frac{1}{\sqrt{K}} \cdot \sigma[\theta_1|\mathbf{d}]$. For dependent samples, $\sigma[a_K]$ can be expressed as

$$\sigma[a_K] = \sqrt{\frac{1+\gamma}{K}} \cdot \sigma[\theta_1|\mathbf{d}] \tag{29}$$

where $\gamma \geq 0$ quantifies the dependency of the generated samples.

- N_{eff} is the number of effectively independent samples in the generated set of K posterior samples (of the first component θ_1). This quantity specifies how many truly independent posterior samples of θ_1 would give the same variance in the sample mean as $\mathrm{Var}[a_K]$ obtained by BUS.

$$N_{\mathrm{eff}} = \left(\frac{\mathrm{E}[s_K]}{\sigma[a_K]}\right)^2 = \frac{K}{1+\gamma} \tag{30}$$

Note that N_{eff} can be interpreted as a measure for the dependency of the generated posterior samples; for fixed K, the smaller N_{eff} the stronger the dependency.

– n_K is the total number of prior samples needed to generate K posterior samples in *Algorithm (2)*.

– $\hat{\theta}_1$ represents a posterior sample obtained with the investigated algorithm.

4.2. Investigated example problems

– *Example problem 1a:* A one dimensional problem with a standard Gaussian prior. The uncertain parameter is denoted by θ. The likelihood for θ is a Gaussian distribution that has mean $\mu_l = 3$ and standard deviation $\sigma_l = 0.3$. This problem has an analytical solution: the posterior distribution is Gaussian with mean and standard deviation of $\mu_l/\left(\sigma_l^2+1\right) = 2.75$ and $1/\sqrt{1+\sigma_l^{-2}} = 0.287$, respectively. The maximum of the likelihood is $L_{\max} = 1/\left(\sigma_l\sqrt{2\pi}\right) = 1.33$. The evidence of the example problem is $c_{\mathrm{E,ref}} = \varphi\left(\mu_l/\sqrt{1+\sigma_l^2}\right)/\sqrt{1+\sigma_l^2} = 6.16 \cdot 10^{-3}$, where $\varphi(\cdot)$ is the PDF of the standard Gaussian distribution. Consequently, $p_{\Omega,\mathrm{ref}}$ of the rejection sampling algorithm is $4.63 \cdot 10^{-3}$, if $c = 1/L_{\max}$.

– *Example problem 1b:* The formulation of this problem is equivalent to *Example problem 1a*, with the only difference being that the likelihood function for θ has mean $\mu_l = 5$ and standard deviation $\sigma_l = 0.2$. The posterior mean and standard deviation are 4.81 and 0.196, respectively. The evidence for this problem is $c_{\mathrm{E,ref}} = 2.36 \cdot 10^{-6}$. Consequently, $p_{\Omega,\mathrm{ref}}$ of the rejection sampling algorithm is $1.18 \cdot 10^{-6}$, if $c = 1/L_{\max}$ and $L_{\max} = 1.99$.

– *Example problem 2:* A 12-dimensional problem with prior $\prod_{i=1}^{12}\varphi(\theta_i)$, where $\varphi(\cdot)$ denotes the standard Gaussian PDF and θ_i is the ith component of the 12-dimensional parameter vector $\boldsymbol{\theta}$. The likelihood function of the problem is $\prod_{i=1}^{12}\varphi\left(\frac{\theta_i-\mu_l}{\sigma_l}\right)/\sigma_l$, with $\sigma_l = 0.6$ where the value μ_l is chosen such that the evidence $c_{\mathrm{E,ref}}$ becomes 10^{-6}; i.e., $\mu_l = 0.462$. The posterior mean and standard deviation of each component of $\boldsymbol{\theta}$ are 0.34 and 0.51, respectively. The theoretical maximum that the likelihood function can take is $L_{\max} = \left(0.6 \cdot \sqrt{2\pi}\right)^{-12} = 7.47 \cdot 10^{-3}$. Thus, $p_{\Omega,\mathrm{ref}}$ of the rejection sampling algorithm is $1.34 \cdot 10^{-4}$, if $c = 1/L_{\max}$.

– *Example problem 3:* A two-story frame structure represented as a two-degree-of-freedom shear building model is investigated. This example problem was originally discussed in [32]. BUS is applied in [4, 13] to solve this problem. The two stiffness coefficients k_1 (first story) and k_2 (second story) of the model are considered uncertain. The uncertainty in k_1 and k_2 is expressed as $k_1 = \theta_1 \cdot k_{\mathrm{n}}$ and $k_2 = \theta_2 \cdot k_{\mathrm{n}}$, where θ_1 and θ_2 are uncertain parameters and $k_{\mathrm{n}} = 29.7 \cdot 10^6\mathrm{N/m}$. The prior distributions of θ_1 and θ_2 are modeled as independent log-Normal distributions with modes 1.3 and 0.8 and standard deviation 1.0. The lumped story masses m_1 (first story) and m_2 (second story) are considered deterministic and have masses $m_1 = 16.5 \cdot 10^3\mathrm{kg}$ and $m_2 = 16.1 \cdot 10^3\mathrm{kg}$. The influence of damping is neglected. Bayesian updating is performed based on the measured first two eigen-frequencies of the system: $\tilde{f}_1 = 3.13\mathrm{Hz}$ and $\tilde{f}_2 = 9.83\mathrm{Hz}$. The likelihood of the problem is expressed as $L(\boldsymbol{\theta}) = \exp\left(-0.5 \cdot J(\boldsymbol{\theta})/\sigma_\varepsilon^2\right)$, where $\sigma_\varepsilon = 1/16$ and $J(\boldsymbol{\theta}) = \sum_{j=1}^{2}\lambda_j^2\left(\frac{f_j^2(\boldsymbol{\theta})}{\tilde{f}_j^2}-1\right)^2$ with $\lambda_1 = \lambda_2 = 1$ and $f_j(\boldsymbol{\theta})$ as the jth eigen-frequency predicted by the model. The posterior distribution of this problem is bimodal [4, 32]. The reference evidence is: $c_{\mathrm{E,ref}} = p_{\Omega,\mathrm{ref}} = 1.52 \cdot 10^{-3}$ (since $L_{\max} = 1$). Moreover, $\mathrm{E}[k_1|\mathbf{d}] = 1.12$ and $\sigma[k_1|\mathbf{d}] = 0.66$.

The reference solutions of the presented example problems are summarized in Table 1. In addition to the quantities $c_{\mathrm{E,ref}}$, $L_{\max}$, $p_{\Omega,\mathrm{ref}}$, $\mathrm{E}[\theta_1|\mathbf{d}]$ and $\sigma[\theta_1|\mathbf{d}]$, some statistics of quantity $b_{10^3,\max}$ are listed in the last four rows. The statistics for $b_{10^3,\max}$ can be computed explicitly for *Example problems 1a*, *1b* and *2*, and are evaluated by means of a large number of repeated runs of the rejection sampling algorithm with $K = 10^3$ for *Example problems 3*. It is obvious that *Example problem 2* differs from the other problems with respect to the statistics of $b_{10^3,\max}$: For *Example problem 2*, $\mathrm{E}[b_{10^3,\max}] = 0.46$ and $\Pr\left[b_{10^3,\max} > 0.8\right] = 1 \cdot 10^{-4}$, whereas $\mathrm{E}[b_{10^3,\max}] \approx 1$ and $\Pr\left[b_{10^3,\max} > 0.8\right] = 1$ for all other example problems. Consequently, it is extremely unlikely that a $b_{10^3,\max}$ close to *one* will be observed in *Example problem 2* in a set of 10^3 posterior samples.

4.2.1. Assessment of the performance for $b < 1$

We assess the performance of the introduced example problems for *Algorithm (2)* and different values of $b \in (0, 1]$. The

Table 1: Reference solution of the investigated example problems.

	Example problem 1a	**Example problem 1b**	**Example problem 2**	**Example problem 3**
$c_{\mathrm{E,ref}}$	$6.16 \cdot 10^{-3}$	$2.36 \cdot 10^{-6}$	$1.00 \cdot 10^{-6}$	$1.52 \cdot 10^{-3}$
$L_{\max}$	1.33	1.99	$7.47 \cdot 10^{-3}$	1.00
$p_{\Omega,\mathrm{ref}}$	$4.63 \cdot 10^{-3}$	$1.18 \cdot 10^{-6}$	$1.34 \cdot 10^{-3}$	$1.52 \cdot 10^{-3}$
$\mathrm{E}[\theta_1]$	2.75	4.81	0.34	1.12
$\sigma[\theta_1]$	0.287	0.196	0.51	0.66
$\mathrm{E}[b_{10^3,\max}]$	1	1	0.46	0.999
$\Pr\left[b_{10^3,\max} > 0.8\right]$	1	1	$1 \cdot 10^{-4}$	1
$\Pr\left[b_{10^3,\max} < 0.99\right]$	10^{-37}	10^{-32}	1	$1 \cdot 10^{-4}$
$\Pr\left[b_{10^3,\max} < 0.999\right]$	$5 \cdot 10^{-12}$	$1 \cdot 10^{-10}$	1	0.38

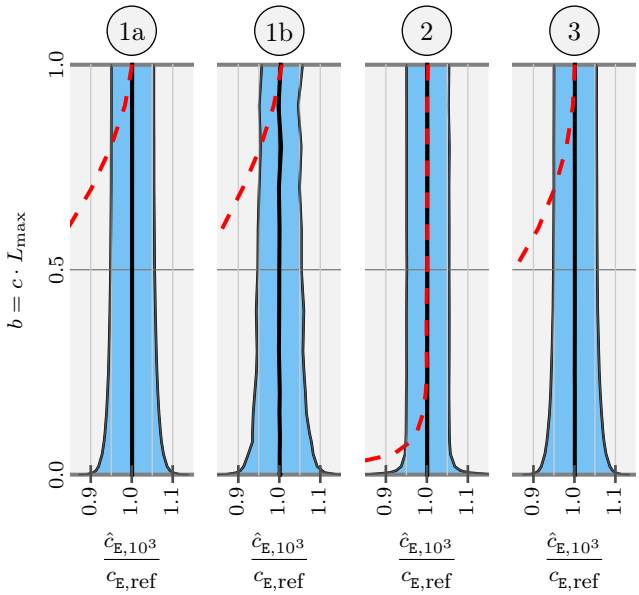

Figure 3: Statistics of the estimated evidence $\hat{c}_{\mathrm{E},10^3}$ divided by the reference value of the evidence $c_{\mathrm{E,ref}}$ for different values of $b = L_{\max} \cdot c$. The **thick black line** represents the average obtained with *Algorithm (2)* – as the average is 1.0 with good approximation, the estimate $\hat{c}_{\mathrm{E},10^3}$ can be considered unbiased. The highlighted area shows the 90% confidence interval of the estimated $\hat{c}_{\mathrm{E},10^3}/c_{\mathrm{E,ref}}$ computed with *Algorithm (2)*. The dashed red line shows the average obtained with *Algorithm (1)*; the bias in the estimated evidence clearly increases for decreasing b. The underlying data was generated by solving the updating problem repeatedly, generating $K = 10^3$ posterior samples in each run.

number of posterior samples generated per updating run is $K = 10^3$. The smallest value of b investigated in the studies is 10^{-4}. The results are presented in Figs. (3)–(5). Fig. 3 shows the statistics of the estimated evidence, Fig. 4 shows the statistics of the generated posterior samples of θ_1, and Fig. 5 shows the statistics of both $\mathrm{E}[n_{10^3}]$ and $\sigma[a_{10^3}]$. The data used to plot Figs. (3)–(5) was generated by solving the updating problem repeatedly, generating $K = 10^3$ posterior samples in each run. The number of times the problem was solved is $2 \cdot 10^5$, 400, $9 \cdot 10^3$ and $8 \cdot 10^4$ for *Example problems 1a*, *1b*, *2* and *3*, respectively.

The statistics of the estimated evidence $\hat{c}_{\mathrm{E},10^3}$ as a function of b are presented in Fig. 3. Independent of the value of b, no bias in the estimate of the evidence can be observed. Additional to the mean, the 90% confidence interval of $\hat{c}_{\mathrm{E},10^3}/c_{\mathrm{E,ref}}$ is shown. For $b \in [0.5, 1]$, the influence of b on the interval can be considered negligible, independent of the example problem. The mean of $\hat{c}^{(c)}_{\mathrm{E},10^3}/c_{\mathrm{E,ref}}$ computed with *Algorithm (1)* and $c^{-1} < L_{\max}$ is also shown in Fig. 3. The results clearly show that the estimate of the evidence $\hat{c}^{(c)}_{\mathrm{E},10^3}$ obtained with *Algorithm (1)* and $c^{-1} < L_{\max}$ underestimates the true evidence $c_{\mathrm{E,ref}}$ of the example problem.

Fig. 4 shows that the estimates of both the mean and the 90% confidence interval of the posterior samples $\hat{\theta}_1$ obtained with the *M-H rejection sampling* algorithm (*Algorithm (2)*) are unbiased, independent of the choice of b. The same cannot be said about *Algorithm (1)* and $c^{-1} < L_{\max}$. In this case, the bias in the mean and the deviation from the 90% confidence interval of the samples θ_1 generated with *Algorithm (2)* increases with decreasing b. This effect is more dominant in *Example problems 1a* and *1b* than in *Example problems 2* and *3*.

The samples produced by the standard rejection sampling are independent, whereas the samples generated with *Algorithm (2)* and $b < 1$ are dependent. The dependency of the samples increases with decreasing b. This effect is illustrated in the right part of Fig. 5. The effective number of independent posterior samples N_{eff} decreases with b. Thus, the efficiency of *Algorithm (2)* with respect to the generated posterior samples decreases with decreasing b. However, the computational cost to generate 10^3 (dependent) posterior samples decreases also with decreasing b – see left part of Fig. 5.

For $b \in [0.5, 1]$, the influence of b on the 90% confidence interval of $\hat{c}_{\mathrm{E},10^3}/c_{\mathrm{E,ref}}$ (shown in Fig. 3) was found to be negligible, independent of the example problem. However, this does not imply that it is irrelevant whether the analysis is performed with $b = 1$ or, for example, with $b = 0.5$. The reason is that although the discussed confidence interval remains stable, the computational cost decreases with b. The computational cost is expressed in terms of the average number of prior samples $\mathrm{E}[n_{10^3}]$ required to perform the analysis (see left part of Fig. 5). For $b \in [0.5, 1]$ the value of $\mathrm{E}[n_{10^3}]$ is – with good approximation – proportional to b. Thus, working with a reduced b is computationally efficient.

The plots shown in Fig. 5 demonstrate that the relative size of $\mathrm{E}[n_{10^3}]$ decreases faster than the relative size of N_{eff}. Thus,

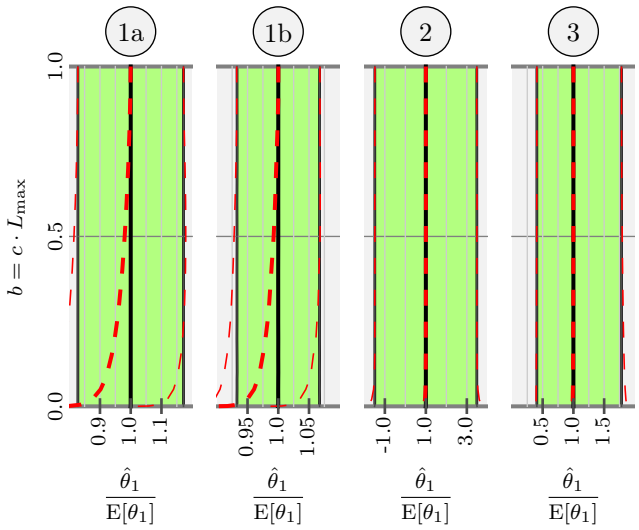

Figure 4: Statistics of the estimated posterior samples $\hat{\theta}_1$ divided by the reference mean of θ_1 for different values of $b = L_{\max} \cdot c$. The **thick black line** represents the average obtained with *Algorithm (2)* – as the average is 1.0 with good approximation, the sample mean can be considered unbiased. The highlighted area shows the 90% confidence interval obtained with *Algorithm (2)*. The dashed red lines show the average and confidence intervals obtained with *Algorithm (1)*. The underlying data was generated by solving the updating problem repeatedly, generating $K = 10^3$ posterior samples in each run.

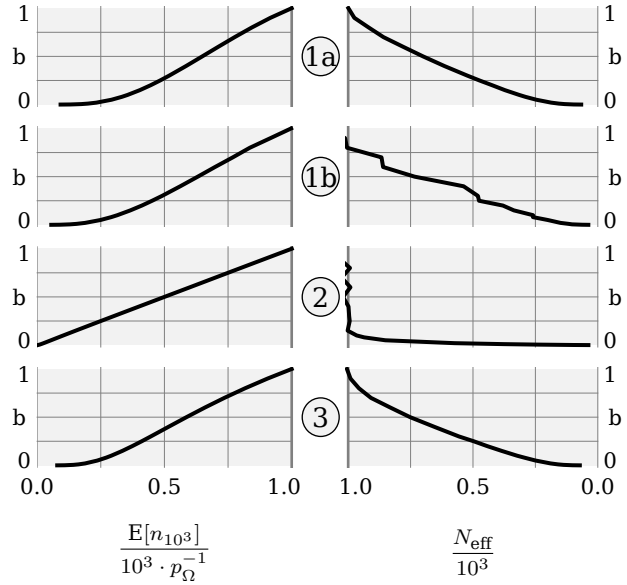

Figure 5: The average number of model calls (left side of plot, x-axis) and the standard deviation of the estimate a_{10^3} (right side of plot, x-axis) are shown for $b \in [0, 1]$ (y-axis of sub-plots) and *Algorithm (2)*. The quantity b is plotted on the y-axis of the sub-plots and ranges from *zero* to *one*. Horizontally, the figure is split in two parts: (**left-part**) On the left-hand side of the figure the average number $\mathrm{E}[n_{10^3}]$ of prior samples required to solve the problem is divided by the average number of prior samples needed in standard rejection sampling; i.e., by $10^3/p_\Omega$. For $b \in [0, 1]$, this quantity is within $[0, 1]$; it measures how many prior samples were needed to solve the problem compared to standard rejection sampling. (**right-part**) On the right-hand side of the figure the effective number of independent posterior samples (N_{eff}) divided by the total number (10^3) of posterior samples is shown.

Algorithm (2) is more efficient for smaller b than for larger b. In general, the results presented in this section demonstrate that neither *Algorithm (1)* nor *Algorithm (2)* is a particularly favorable strategy. Instead, if a total number n of model calls that one is willing to invest can be specified, it is better to generate n samples from the prior distribution and regard the prior samples as weighted samples from the posterior distribution (i.e., $b \to 0$). In this case, the sample weights are proportional to the likelihood function, and the evidence can be estimated as the sample mean of the likelihood function obtained with the samples from the prior distribution (see e.g., [33]). However, the advantage of *Algorithm (2)* (and the BUS approach in general) is that the rejection sampling step (*step (2)*) can be replaced by any structural reliability method. A reliability method that is particularly well suited for BUS is Subset Simulation, it can handle problems with many uncertain parameters and can efficiently estimate very small probabilities that may arise within BUS (see e.g. [4]). Moreover, an estimate for the evidence c_{E} is obtained as a by-product of BUS.

5. Concluding remarks

Bayesian Updating with Structural reliability methods (BUS) reinterprets the Bayesian updating problem as a structural reliability problem. The evidence of the inference problem is obtained as a by-product of BUS. In this paper, we propose a post-processing step for BUS that does not require the scaling constant c to be chosen such that $c \cdot L_{\max} \leq 1$. By means of the proposed post-processing step, an unbiased estimate for the evidence as well as samples from the posterior distribution can be obtained even if the maximum of the likelihood function $L_{\max}$ is not known. The performance of the proposed post-processing step was demonstrated by means of different example problems and rejection sampling. In practice, the proposed post-processing step for BUS can be combined with any structural reliability method.

Acknowledgments

The authors acknowledge the support of the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative. The first author also would like to thank the California Institute of Technology (Caltech) for his research stay at Caltech, which provided the basis for this work.

References

[1] J. L. Beck, "Bayesian system identification based on probability logic," *Structural Control and Health Monitoring*, vol. 17, no. 7, pp. 825–847, 2010.

[2] D. J. MacKay, "Bayesian interpolation," *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.

[3] L. Wasserman, "Bayesian model selection and model averaging," *Journal of mathematical psychology*, vol. 44, no. 1, pp. 92–107, 2000.

[4] D. Straub and I. Papaioannou, "Bayesian updating with structural reliability methods," *Journal of Engineering Mechanics*, 2015.

[5] O. Ditlevsen and H. O. Madsen, *Structural reliability methods*. Wiley New York, 1996.
[6] R. E. Melchers, *Structural reliability analysis and prediction*. John Wiley & Son Ltd, 1999.
[7] D. Straub, "Engineering risk assessment," in *Risk – A Multidisciplinary Introduction*, pp. 333–362, Springer, 2014.
[8] D. Straub, I. Papaioannou, and W. Betz, "Bayesian analysis of rare events," *Journal of Computational Physics*, vol. 314, pp. 538–556, 2016.
[9] M. Lemaire, *Structural reliability*. John Wiley & Sons, 2013.
[10] S.-K. Au and J. L. Beck, "Estimation of small failure probabilities in high dimensions by subset simulation," *Probabilistic Engineering Mechanics*, vol. 16, no. 4, pp. 263–277, 2001.
[11] K. M. Zuev, J. L. Beck, S.-K. Au, and L. S. Katafygiotis, "Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions," *Computers & structures*, vol. 92, pp. 283–296, 2012.
[12] W. Betz, I. Papaioannou, J. L. Beck, and D. Straub, "Bayesian inference with subset simulation: strategies and improvements," *Computer Methods in Applied Mechanics and Engineering*, vol. 331, pp. 72–93, 2018.
[13] F. DiazDelaO, A. Garbuno-Inigo, S. Au, and I. Yoshida, "Bayesian updating and model class selection with Subset Simulation," *Computer Methods in Applied Mechanics and Engineering*, 2017.
[14] W. Betz, I. Papaioannou, and D. Straub, "Adaptive variant of the BUS approach to Bayesian updating," in *Proceedings of 9th European Conference on Structural Dynamics*, 2014.
[15] I. Papaioannou, W. Betz, and D. Straub, "Bayesian model updating of a tunnel in soft soil with settlement measurements," *Geotechnical Safety and Risk IV*, p. 351, 2013.
[16] W. Betz, C. M. Mok, I. Papaioannou, and D. Straub, "Bayesian model calibration using structural reliability methods: application to the hydrological abc model," in *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management* (M. Beer, S.-K. Au, and J. W. Hall, eds.), pp. 2734–2743, ASCE, 2014.
[17] H. Lind, "An exact and invariant first-order reliability format.," *Journal of Engineering mechanics*, vol. 100, 1974.
[18] R. Rackwitz and B. Flessler, "Structural reliability under combined random load sequences," *Computers & Structures*, vol. 9, no. 5, pp. 489–494, 1978.
[19] M. Hohenbichler and R. Rackwitz, "Improvement of second-order reliability estimates by importance sampling," *Journal of Engineering Mechanics*, vol. 114, no. 12, pp. 2195–2199, 1988.
[20] P. Koutsourelakis, H. Pradlwarter, and G. Schuëller, "Reliability of structures in high dimensions, Part I: algorithms and applications," *Probabilistic Engineering Mechanics*, vol. 19, no. 4, pp. 409–417, 2004.
[21] R. Rackwitz, "Reliability analysis – a review and some perspectives," *Structural safety*, vol. 23, no. 4, pp. 365–395, 2001.
[22] D. Straub, "Reliability updating with equality information," *Probabilistic Engineering Mechanics*, vol. 26, no. 2, pp. 254–258, 2011.
[23] A. F. Smith and A. E. Gelfand, "Bayesian statistics without tears: a sampling–resampling perspective," *The American Statistician*, vol. 46, no. 2, pp. 84–88, 1992.
[24] J. Haldane, "On a method of estimating frequencies," *Biometrika*, vol. 33, no. 3, pp. 222–225, 1945.
[25] D. Finney, "On a method of estimating frequencies," *Biometrika*, vol. 36, no. 1/2, pp. 233–234, 1949.
[26] E. T. Jaynes, *Papers on probability, statistics and statistical physics*. Dordrecht, Holland: D. Reidel Publishing, 1983.
[27] E. T. Jaynes, *Probability theory: the logic of science*. Cambridge university press, 2003.
[28] B. D. Flury, "Acceptance–rejection sampling made easy," *SIAM Review*, vol. 32, no. 3, pp. 474–476, 1990.
[29] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys*, vol. 21, p. 1087, 1953.
[30] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
[31] L. Tierney, "Markov chains for exploring posterior distributions," *the Annals of Statistics*, pp. 1701–1728, 1994.
[32] J. L. Beck and S.-K. Au, "Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation," *Journal of Engineering Mechanics*, vol. 128, no. 4, pp. 380–391, 2002.
[33] N. Chopin, "A sequential particle filter method for static models," *Biometrika*, vol. 89, no. 3, pp. 539–552, 2002.