WILEY

Bayesian inference with subset simulation in varying dimensions applied to the Karhunen–Loève expansion

Felipe Uribe[1] | **Iason Papaioannou[1]** | **Jonas Latz[2]** | **Wolfgang Betz[1]** | **Elisabeth Ullmann[2]** | **Daniel Straub[1]**

[1]Engineering Risk Analysis Group, Technische Universität München, München, Germany

[2]Chair of Numerical Analysis, Technische Universität München, Garching b.M., Germany

Correspondence
Felipe Uribe, Engineering Risk Analysis Group, Technische Universität München, Arcisstr. 21, 80333 München, Germany.
Email: felipe.uribe@tum.de

Abstract

Uncertainties associated with spatially varying parameters are modeled through random fields discretized into a finite number of random variables. Standard discretization methods, such as the Karhunen–Loève expansion, use series representations for which the truncation order is specified a priori. However, when data is used to update random fields through Bayesian inference, a different truncation order might be necessary to adequately represent the posterior random field. This is an inference problem that not only requires the determination of the often high-dimensional set of coefficients, but also their dimension. In this article, we develop a sequential algorithm to handle such inference settings and propose a penalizing prior distribution for the dimension parameter. The method is a variable-dimensional extension of BUS (Bayesian Updating with Structural reliability methods), combined with subset simulation (SuS). The key idea is to replace the standard Markov Chain Monte Carlo (MCMC) algorithm within SuS by a trans-dimensional MCMC sampler that is able to populate the discrete-continuous parameter space. To address this task, we consider two types of MCMC algorithms that operate in a fixed-dimensional saturated space. The performance of the proposed method with both MCMC variants is assessed numerically for two examples: a 1D cantilever beam with spatially varying flexibility and a 2D groundwater flow problem with uncertain permeability field.

KEYWORDS
Bayesian model choice, inverse problems, Karhunen–Loève expansion, random fields, trans-dimensional MCMC, uncertainty quantification

1 | INTRODUCTION

Numerical approximations to partial differential equations (PDEs) require the specification of input parameters that are typically unknown and/or intrinsically random. Uncertainties in the values of these quantities can be reduced by incorporating observations of the physical system into the numerical model. This represents an inverse problem in which the

objective is to identify the model parameters that are compatible with the available information. The complexity of inverse problems increases in model choice situations, whereby a single model needs to be selected from a predefined collection of plausible models. Each model in the set can have different parameters or can represent different mathematical assumptions. The uncertainty in the model and its parameters can be treated in a unified manner within the Bayesian inference framework.[1-3]

In the Bayesian approach to solve inverse problems the uncertain parameters are represented as random variables. The idea is to update the (*prior*) probability distribution of the parameters by including information about the PDE model and observed data (*likelihood*). Solving the Bayesian inverse problem then amounts to estimating or characterizing the updated (*posterior*) distribution. In the case of model choice, inference involves the estimation of a conditional posterior distribution induced by the model class and determining the plausibility of the respective model class. The prior distribution is then hierarchically structured into a prior for the parameters conditioned on the model and a second prior for the model itself.[2,4]

Closed-form expressions of the model and parameter posterior are oftentimes cumbersome to obtain. As a result, approximate solutions are computed in practice. One common approach in model choice situations is to use methods that are based on the likelihood function of individual models, these include, Akaike and Bayesian information (see, e.g., References 5,6). Another flexible way to approach the model choice problem is via sampling methods. In this case, the characterization of the posterior requires the exploration of a discrete-continuous space. This task can be performed by two main strategies: (i) fixing the model and solving the inverse problem for each case; an approach referred to as *within-model simulation* (see, e.g., Reference 7 and several methods described in the seminal work[8]); and (ii) performing simultaneous inference on both, model and parameters, using Markov chain Monte Carlo (MCMC) algorithms that explore the parameter space by moving between different models. This approach is called *across-model simulation*, and the standard algorithm is the reversible jump MCMC algorithm[9,10] (other MCMC samplers for across-model simulation include[11-14]). Most of the disadvantages of standard MCMC, such as convergence rate deterioration with increasing dimensions and burn-in/thinning periods requirements, are also present in the MCMC samplers used in across-model simulation.[15] In standard inference settings, specialized sequential algorithms that gradually approach the posterior distribution alleviate several of these issues.[16,17] Some of these algorithms have been adapted to perform across-model simulation, for example, the sequential importance sampling with reversible jump MCMC[18] and the population-based reversible jump MCMC.[19]

The problem of inferring both, models and parameters, also applies to cases involving a single mathematical model that has parameters with variable dimension. Common examples include, mixture models with an unknown number of components,[20] polynomial regression where the degree of the polynomial is variable,[17] or general functional representations that use series expansions for which the number of terms is unknown. The latter is of particular relevance in the context of learning spatially varying parameters represented via random fields.[21] Random fields increase the complexity of the inverse problem since the posterior distribution is defined over an infinite-dimensional space. Series representations are typically applied to project the random field to a finite-dimensional space; some approaches include, the Karhunen–Loève expansion,[22,23] the expansion optimal linear estimation,[24] the perturbation method,[25] spectral representation,[26] among others. In particular, the Karhunen–Loève (KL) expansion discretizes the field using the eigenvalues and eigenfunctions of its autocovariance operator to construct a series expansion with random coefficients.[27] It is common practice to truncate the KL expansion after a finite number of terms based on some variance-representation criterion. This heuristic is generally valid in prior situations when no information or observations about the field are available. In the inversion case, the optimal number of terms in the series expansion is unknown and it is controlled by the data.[28,29]

In this article, we propose a sequential methodology that is able to perform inference in parameter spaces of different dimension. The method is an extension of the classical BUS (Bayesian updating with structural reliability methods) framework, which expresses a Bayesian inverse problem as an equivalent rare event simulation task.[30] For an efficient and sequential solution of the inverse problem, BUS is combined with subset simulation (SuS),[31] although other rare event estimation methods can be employed (see, e.g., Reference 32). The main idea is to incorporate the discrete dimension random variable into the parameter space defined by BUS, such that the resulting sequence of intermediate distributions also depend on the dimension. The associated intermediate densities become *trans-dimensional* and the standard MCMC algorithms used within BUS with SuS are no longer valid. Therefore, we investigate a class of trans-dimensional and dimension-independent MCMC algorithms that explore a so-called *saturated* or composite parameter space[14] in an alternating manner. In this space, the dimension is fixed to a maximum upper value, which is selected conservatively based on prior information. Particularly, we discuss a Metropolis-within-Gibbs algorithm and develop a step-wise sampler as a simplified reversible jump MCMC in the saturated space. Since this space is typically high-dimensional, the core of

these algorithms is the preconditioned Crank–Nicolson sampler.[33] The efficiency and accuracy of the method is tested on engineering models involving random field parameters represented with the KL expansion: one example for which reference solution of the dimension posterior is available, and a second example that requires the estimation of the posterior at some reference dimensions using within-model simulation runs, to verify the dimension posterior estimated by our algorithm.

We also address the specification of the model/dimension prior by defining a discrete distribution that penalizes increasing dimensionality. In model choice problems, imposing a penalty to complicated models is necessary (see References 1,5,34 for a discussion). A model with more parameters usually fits the data better than a model with less parameters. However, the actual modeling improvement might be negligible or possible over-fitting of the data can arise.[28,29] The proposed dimension prior is defined to avoid such situations and we build it based on the geometry of the parameter space and prior information about the random fields.

The organization of the article is as follows: in Section 2, we present fundamental concepts of random field modeling and the KL representation; we also formulate the Bayesian inversion problem in the fixed- and variable-dimensional settings. At the end of the section, we propose a prior for the specification of the dimension parameter. The major contribution of this work is introduced in Section 3, where we explain the trans-dimensional BUS algorithm. This methodology is based on trans-dimensional MCMC samplers, which are discussed in Section 4. Next, the proposed method is demonstrated by means of two numerical experiments in Section 5. The article finalizes with a discussion of results in Section 6 and a summary of the work in Section 7.

2 | MATHEMATICAL FORMULATION

2.1 | Random fields and the Karhunen–Loève expansion

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $D \subseteq \mathbb{R}^d$ an index set representing a physical domain. A real-valued random field is a function $H(\boldsymbol{x}, \omega) : D \times \Omega \to \mathbb{R}$, with arguments $\boldsymbol{x} \in D$ a spatial coordinate and $\omega \in \Omega$ an outcome of the sample space.[35,36] Intuitively, a random field can be interpreted either as a single random variable that takes values in a function space, or as a collection of random variables indexed in space.

Random fields are represented in terms of a finite set of random variables using *stochastic discretization* algorithms. Popular dimensionality reduction techniques are based on finite expansions of random variables and deterministic functions. These include the Karhunen–Loève expansion,[22,23] which expresses a random field as a linear combination of orthogonal functions chosen as the eigenfunctions resulting from the spectral decomposition of the covariance operator. Since all positive-definite functions have an unique spectral representation (see Bochner's Theorem [35, section 3]), one can define an orthonormal basis, unique and optimal in the mean-squared sense, consisting of the eigenfunctions of the covariance operator together with a sequence of real and non-negative eigenvalues [37, p. 248]. One can use such basis to represent a second-order random field as

$$H(\boldsymbol{x}, \omega) \approx \widehat{H}(\boldsymbol{x};\ k, \boldsymbol{\theta}(\omega)) := \mu(\boldsymbol{x}) + \sum_{i=1}^{\infty} \mathbb{1}\ (i \leq k)\ \sqrt{\lambda_i} \phi_i(\boldsymbol{x})\ \theta_i(\omega), \tag{1}$$

where $\widehat{H}(\boldsymbol{x};\ k, \boldsymbol{\theta}(\omega))$ denotes the approximated field, $\mathbb{1}\ (\cdot)$ is the indicator function, k is the truncation order of the expansion, $\theta_i(\omega) : \Omega \to \mathbb{R}$ is a set of mutually uncorrelated random variables with mean zero and unit variance, $\lambda_i \in [0, \infty)$ are the eigenvalues of the covariance operator (satisfying $\lambda_i \geq \lambda_{i+1}$, $\lim_{i\to\infty} \lambda_i = 0$, $\sum_{i=1}^{\infty} \lambda_i < \infty$), and $\phi_i(\boldsymbol{x}) : D \to \mathbb{R}$ are the eigenfunctions of the covariance operator. For Gaussian random fields, the variables $\theta_i(\omega)$ are independent standard Gaussian. In the general case, the distribution of $\theta_i(\omega)$ is cumbersome to estimate. The series expansion in (1) follows from Mercer's Theorem (details are provided in Reference 38) and it is referred to as the *Karhunen–Loève* (KL) expansion. We remark that the set of eigenpairs $\{\lambda_i, \phi_i\}$ is computed through the solution of an homogeneous Fredholm integral equation of the second kind,[27] which can be solved using different approaches, such as projection (collocation, Galerkin)[39] or Nyström methods.[40]

The KL expansion is often employed to reduce the dimensionality and parameterize random fields. Consider the square-integrable random vector $\boldsymbol{\theta}(\omega) \in \Theta_k \subseteq \mathbb{R}^k$ resulting from truncating the KL series expansion (1) at the kth term. This truncation yields an approximated field, which is optimal in the mean-squared-error sense as compared to any other spectral projection algorithm.[27] Since the eigenpairs associated to the covariance operator are deterministic quantities,

the parameter $\theta(\omega)$ characterizes the randomness of the field. Hence, the KL construction only depends on the vector of random coefficients θ and the truncation order k.

2.2 | Bayesian inverse problems in fixed dimension

We begin by considering the forward problem $\boldsymbol{y} = G(H(\boldsymbol{x}, \omega))$, where G is a *solution operator* expressing the relationship between the input parameters and the model response. We are interested in cases where the operator G implies the solution of differential equation models with random fields parameters $H(\boldsymbol{x}, \omega)$. These are encountered for instance, in continuum mechanics where material parameters are spatially variable (e.g., Young's modulus and Poisson ratio). In this case, G operates on function spaces defined on D since both, input and output, are random field realizations of two different quantities on the physical domain. The dimension of the forward problem can be reduced using the parameterized random field in the expansion (1), such that the input parameter space is now given by Θ_k. Assuming a fixed truncation order k, we write the approximated field as $\hat{H}(\boldsymbol{x}; \theta)$ and the parameter space as $\Theta \subseteq \mathbb{R}^k$.

In inverse problems, the aim is to infer the parameters $\theta \in \Theta$ given noisy observations of the system response $\widetilde{\boldsymbol{y}} \in \mathcal{Y} := \mathbb{R}^m$, with m denoting the number of observations and $\mathcal{Y}$ the data space. Assuming an additive observation error, the objective is:

$$\text{find } \theta \in \Theta \; : \; \widetilde{\boldsymbol{y}} = \mathcal{G}\left(\hat{H}(\boldsymbol{x}; \theta)\right) + \boldsymbol{\eta}, \tag{2}$$

where $\mathcal{G} = \mathcal{O} \circ G : \Theta \to \mathcal{Y}$ is the *forward response operator*, defined as the composition of the solution operator G with an observation operator $\mathcal{O}$ that maps the forward solution to the data space; and $\boldsymbol{\eta} \in \mathbb{R}^m$ is the observation noise which is assumed to be Gaussian distributed with mean zero and non-singular covariance matrix $\boldsymbol{\Sigma}_{\text{obs}} \in \mathbb{R}^{m \times m}$.

The inverse problem (2) is generally ill-posed. Bayesian statistical methods offer a framework that integrates the observations with prior information, providing a mechanism of regularization. In Bayesian inverse problems, the components of the parameter vector θ are modeled as random variables and are assumed to have an initial *prior density* $\pi_{\text{pr}}(\theta)$. The *likelihood function* $\mathrm{L}(\theta; \widetilde{\boldsymbol{y}}) \propto \pi_{\text{like}}(\widetilde{\boldsymbol{y}} \mid \theta)$ is proportional to the probability density of $\widetilde{\boldsymbol{y}}$ given a parameter state θ and provides a link between the computational model and data. After including observations, the updated belief about θ is represented by the *posterior density* $\pi_{\text{pos}}(\theta \mid \widetilde{\boldsymbol{y}})$. Through Bayes' Theorem this conditional density is

$$\pi_{\text{pos}}\left(\theta \mid \widetilde{\boldsymbol{y}}\right) = \frac{1}{Z_{\widetilde{\boldsymbol{y}}}} \pi_{\text{pr}}(\theta) \mathrm{L}\left(\theta; \widetilde{\boldsymbol{y}}\right) \propto \exp\left(-\frac{1}{2}\left\|\boldsymbol{\Sigma}_{\text{pr}}^{-1/2}\left(\theta - \boldsymbol{\mu}_{\text{pr}}\right)\right\|_2^2 + \ln \mathrm{L}\left(\theta; \widetilde{\boldsymbol{y}}\right)\right), \tag{3}$$

where $Z_{\widetilde{\boldsymbol{y}}} = \int_{\Theta} \pi_{\text{pr}}(\theta) \mathrm{L}\left(\theta; \widetilde{\boldsymbol{y}}\right) d\theta$ is the normalizing constant of $\pi_{\text{pos}}\left(\theta \mid \widetilde{\boldsymbol{y}}\right)$, called the *model evidence*.

Remark 1. In the right-hand side of (3), we assumed that the prior distribution is Gaussian $\theta \sim \mathcal{N}(\boldsymbol{\mu}_{\text{pr}}, \boldsymbol{\Sigma}_{\text{pr}})$. Particularly, under the KL parametrization the prior mean and covariance become $\boldsymbol{\mu}_{\text{pr}} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{\text{pr}} = \mathbf{I}_k$, where $\mathbf{I}_k \in \mathbb{R}^{k \times k}$ is an identity matrix. The information about the mean and covariance structure of the random field enters directly in the definition of the likelihood function via the forward operator.

2.3 | Bayesian inverse problems in varying dimension

Consider a more general inference setting for which the set of observed data $\widetilde{\boldsymbol{y}}$ is associated not only with one, but with a finite collection of plausible models $\mathcal{K} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k, \ldots, \mathcal{M}_{k_{\max}}\}$, where $k \in \mathcal{K}$ is a model indicator index, and $k_{\max} < \infty$ is a prescribed limit on the collection. The resulting discrete-continuous parameter space can be written as $\mathcal{X} = \cup_{k \in \mathcal{K}}(\{k\} \times \Theta_k)$. Observe that there exist different uncertain parameter vectors $\theta_k \in \Theta_k \subseteq \mathbb{R}^k$ for each particular model k, and thus, the goal is to extract information from the data to infer jointly the pairs $(k, \theta_k) \in \mathcal{X}$. For the sake of simplicity in notation, we shall henceforth use the model indicator index k to denote the model $\mathcal{M}_k$.

Let $\pi_{\text{pr}}(\theta_k \mid k)$ be a prior density imposed on the parameter θ_k given the model k, and $\overline{\pi}_{\text{pr}}(k)$ a discrete prior mass specified over the models k (we use the notation $\overline{\pi}$ to indicate probability mass functions). The joint posterior density over both, model and parameters, is computed from Bayes' Theorem as

$$\pi_{\text{pos}}(k, \theta_k \mid \widetilde{\boldsymbol{y}}) = \frac{1}{Z_{\widetilde{\boldsymbol{y}}}} \overline{\pi}_{\text{pr}}(k) \pi_{\text{pr}}(\theta_k \mid k) \mathrm{L}\left(k, \theta_k; \widetilde{\boldsymbol{y}}\right) \propto \overline{\pi}_{\text{pr}}(k) \exp\left(-\frac{1}{2}\left\|\boldsymbol{\Sigma}_{\text{pr},k}^{-1/2}\left(\theta_k - \boldsymbol{\mu}_{\text{pr},k}\right)\right\|_2^2 + \ln \mathrm{L}\left(k, \theta_k; \widetilde{\boldsymbol{y}}\right)\right), \tag{4}$$

wherein the prior parameters and likelihood function depend on k, and the evidence is given by the law of total probability:

$$\overline{Z}_{\widetilde{\mathbf{y}}} = \sum_{k' \in \mathcal{K}} \overline{\pi}_{\mathrm{pr}}(k') Z_{\widetilde{\mathbf{y}}}(k') = \sum_{k' \in \mathcal{K}} \overline{\pi}_{\mathrm{pr}}(k') \int_{\Theta_{k'}} \pi_{\mathrm{pr}}(\boldsymbol{\theta}_{k'} \mid k') \mathrm{L}\left(k', \boldsymbol{\theta}_{k'}; \widetilde{\mathbf{y}}\right) \mathrm{d}\boldsymbol{\theta}_{k'}, \tag{5}$$

here $Z_{\widetilde{\mathbf{y}}}(k)$ is the evidence of the individual model k. The posterior density of the models is obtained by integrating out the parameters in (4) as

$$\overline{\pi}_{\mathrm{pos}}(k \mid \widetilde{\mathbf{y}}) = \frac{\overline{\pi}_{\mathrm{pr}}(k) \int_{\Theta_k} \pi_{\mathrm{pr}}(\boldsymbol{\theta}_k \mid k) \mathrm{L}\left(k, \boldsymbol{\theta}_k; \widetilde{\mathbf{y}}\right) \mathrm{d}\boldsymbol{\theta}_k}{\sum_{k' \in \mathcal{K}} \overline{\pi}_{\mathrm{pr}}(k') \int_{\Theta_{k'}} \pi_{\mathrm{pr}}(\boldsymbol{\theta}_{k'} \mid k') \mathrm{L}\left(k', \boldsymbol{\theta}_{k'}; \widetilde{\mathbf{y}}\right) \mathrm{d}\boldsymbol{\theta}_{k'}} = \overline{\pi}_{\mathrm{pr}}(k) \frac{Z_{\widetilde{\mathbf{y}}}(k)}{\overline{Z}_{\widetilde{\mathbf{y}}}}. \tag{6}$$

The model posterior in (6) can be used to perform (i) *model choice or selection*, which requires the computation of the maximum a posteriori probability (MAP) estimator, $k_{\mathrm{MAP}} = \arg\max_{k \in \mathcal{K}} \overline{\pi}_{\mathrm{pos}}(k \mid \widetilde{\mathbf{y}})$, or (ii) *model mixing or averaging* which requires the consideration of the whole collection of parameters weighted by $\overline{\pi}_{\mathrm{pos}}(k \mid \widetilde{\mathbf{y}})$. Model choice is used as indicator of model complexity, that is, the model that provides the best alignment with the observed data should be preferred over unnecessarily complicated ones. The model mixing solution consists of the model posterior predictive distribution. In this case, all the collection of models is used for future decisions and the underestimation of uncertainty resulting from choosing only a single model is avoided. Since this process leads to a higher computational cost, only models that are sufficiently likely compared to the MAP estimator may be considered in the analysis. Occam's window and Bayes' factors are used to perform such a model reduction [2, p. 368].

The formulations in (4) and (6) are also applicable to Bayesian non-parametric settings where in fact there exists only a single mathematical model, but one with variable-dimension parameter.[41] We are interested in the latter since this corresponds to Bayesian inverse problems involving random fields represented by a series expansion whose number of terms is variable. In the KL expansion, the set of models is defined by the model indicator indices $\mathcal{K} = \{1, 2, \ldots, k, \ldots k_{\max}\}$, where each element defines a truncation order. This truncation specifies the dimension of the standard Gaussian random coefficients of the KL expansion, thus, each particular model/dimension k involves a vector of uncertain parameters $\boldsymbol{\theta}_k \in \Theta_k$. The aim is to perform simultaneous inference on the discrete random variable k (dimension), and the associated random vector $\boldsymbol{\theta}_k$ (coefficients) of the KL expansion. In the following, we often use the terms model and dimension interchangeably.

2.4 | Selection of the prior distribution for the KL truncation

It has been shown numerically in Reference 29 that the model evidence associated to the dimensions of the KL discretization does not reveal the classical Bayesian penalization appearing for instance in regression models, where constantly increasing the polynomial order, eventually reduces the model evidence values (revealing potential over-fitting). In the KL expansion, the model evidence keeps increasing as one adds more terms (see also Reference 28). Another observation in Reference 29 is that the information gained by continuously increasing KL terms becomes negligible once an 'optimal' truncation is achieved. We conjecture that this behavior is related to the representation of the posterior covariance since its approximation improves as $k \to \infty$, however, a rigorous proof of this result is yet to be derived.

These observations motivate the definition of a prior for the dimension parameter k that penalizes increasing dimensionality. We employ a truncated geometric distribution (more details concerning this choice are given in the appendix):

$$\overline{\pi}_{\mathrm{pr}}(k) = \frac{(1-p)^{k-1}p}{1 - (1-p)^{k_{\max}}}, \qquad k = 1, \ldots, k_{\max}, \tag{7}$$

where $k_{\max}$ is the upper truncation level, and the success probability $p \in (0, 1)$ marks the decay rate of the probability mass. This parameter allows us to control the shape of the distribution. In practice, the parameter space is bounded and typically some prior knowledge about these bounds is available. We select the parameter p by regulating the behavior of the distribution tails, such that $\mathbb{P}\left[k \leq k_{\min}\right] = \overline{F}_{\mathrm{pr}}(k_{\min}; p) = \alpha$, where $k_{\min}$ is a prescribed threshold which is associated with probability α, and $\overline{F}_{\mathrm{pr}}(k_{\min}; p)$ is the cumulative distribution function associated to (7) (as function of p). This defines a nonlinear equation that is solved for p, given values of $k_{\min}$, $k_{\max}$, and α.

Based on our experiments, $k_{\min}$ is chosen as the number of terms in the KL expansion that retains 50% of the variability in the prior random field, and we assign to the event $\{k \leq k_{\min}\}$ a probability of $\alpha = 0.10$. Furthermore, the truncation value $k_{\max}$ is selected as the number of terms in the expansion that retains 99% of the variability. Note that in this strategy, approximately 90% of the probability mass is concentrated on truncation orders higher than those yielding the 50% variability and smaller that those producing the 99% variability. We found that this heuristic produces a decay p that does not excessively penalize high-order KL terms.

We remark that other prior models for the dimension parameter have been derived in the context of random fields, for example, an exponential prior with fixed decay rate for the truncation order in the KL expansion,[33] and a penalized complexity prior for different values of a re-parameterized Matérn kernel.[42]

3 | BAYESIAN INFERENCE WITH SUBSET SIMULATION IN SPACES OF VARYING DIMENSION

The characterization of the posterior distribution using standard MCMC algorithms can be inefficient because not only a large number of iterations are required to compute accurate statistics, but also tuning and post-processing steps need to be implemented to reduce correlation (e.g., burn-in and lag periods). The task is even more complex when the posterior distribution is high- and trans-dimensional. Therefore, a common approach is to embed standard MCMC samplers into algorithms that start from the prior and sequentially approach the posterior distribution.[43] The idea is to explore the posterior on-the-fly by constructing a set of intermediate measures that converge to the full posterior. An approach that belongs to this class of algorithms is BUS (Bayesian Updating with Structural reliability methods), which reformulates the Bayesian inverse problem as a classical reliability analysis (rare event estimation) problem. This construction allows one to employ efficient reliability estimation algorithms to sample from the posterior. In this section, we discuss the combination of BUS with subset simulation and its extension to variable-dimensional inference problems.

3.1 | The BUS formulation

In practice, the posterior density is only known up to its scaling constant, that is, $\pi_{\mathrm{pos}}\left(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{y}}\right) \propto \pi_{\mathrm{pr}}(\boldsymbol{\theta}) \mathrm{L}\left(\boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right) = \pi(\boldsymbol{\theta})$. The posterior can still be characterized by drawing samples from the unnormalized target density $\pi(\boldsymbol{\theta})$. Particularly, the *rejection sampling algorithm* generates samples from $\pi(\boldsymbol{\theta})$ using a proposal density $q(\boldsymbol{\theta})$. The proposal is selected such that it dominates the target function. This means that $q(\boldsymbol{\theta})$ must have equal or heavier tails than those of $\pi(\boldsymbol{\theta})$. Therefore, the proposal satisfies the relation

$$\sup_{\theta \in \Theta}\left(\frac{\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right) \leq \overline{c} < \infty \quad \text{for some } \textit{covering} \text{ constant } \overline{c} \in (1, \infty) \tag{8}$$

and $\mathrm{supp}\left(\pi(\boldsymbol{\theta})\right) \subseteq \mathrm{supp}\left(q(\boldsymbol{\theta})\right)$. Thereafter, samples drawn from $q(\boldsymbol{\theta})$ are rejected strategically to make the accepted samples distributed according to $\pi(\boldsymbol{\theta})$. A simple choice for the proposal density is the prior distribution $\pi_{\mathrm{pr}}(\boldsymbol{\theta})$. In this case, the acceptance probability α in rejection sampling[16] becomes

$$\alpha = \frac{\pi(\boldsymbol{\theta})}{\overline{c}\, q(\boldsymbol{\theta})} = \frac{\pi_{\mathrm{pr}}\left(\boldsymbol{\theta}\right) \mathrm{L}\left(\boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right)}{\overline{c}\, \pi_{\mathrm{pr}}(\boldsymbol{\theta})} = c\, \mathrm{L}\left(\boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right), \tag{9}$$

where $c = 1/\overline{c} \in (0, \infty)$ and the covering constant is selected such that $\overline{c} \geq L_{\max} = \max(\mathrm{L}\left(\boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right))$. Rejection sampling then amounts to (i) drawing a standard uniform random number $\upsilon \sim \mathrm{Unif}[0, 1]$, (ii) sampling a candidate from the prior $\boldsymbol{\theta} \sim \pi_{\mathrm{pr}}(\boldsymbol{\theta})$, and (iii) accepting the candidate if $\upsilon \leq \alpha = c\, \mathrm{L}\left(\boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right)$. This particular acceptance-rejection mechanism allows us to generate the space

$$\mathcal{H} = \{(\boldsymbol{\theta}, \upsilon) \in \boldsymbol{\Theta} : h(\boldsymbol{\theta}, \upsilon) \leq 0\}, \quad \text{where} \quad h(\boldsymbol{\theta}, \upsilon) = \upsilon - c\, \mathrm{L}\left(\boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right) \tag{10}$$

and $\boldsymbol{\Theta} = [\Theta, \Upsilon]$ is an augmented parameter space ($\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$ and $\upsilon \in \Upsilon \subseteq [0, 1]$).

In the context of reliability analysis and rare event simulation, the space $\mathcal{H}$ defines a failure domain with *limit-state function* (LSF) $h(\theta, v)$. If one is able to generate samples $\{(v, \theta)\} \in \Theta$ that fall into $\mathcal{H}$, the parameter samples $\{\theta\} \in \Theta$ will be distributed according to the posterior $\pi_{\text{pos}}(\theta \mid \widetilde{y})$. Note that the $\{v\}$ samples are auxiliary. This connection is the foundation of the BUS approach, since one can employ existing methods from rare event simulation to perform Bayesian inference. Indeed, the previous rejection sampling algorithm corresponds to applying standard Monte Carlo simulation for the solution of a rare event estimation problem defined by the LSF $h(\theta, v)$ over the space Θ.

The main objective in reliability analysis is to estimate the probability of failure. When employing the BUS framework, this value is associated to the probability that the samples belong to the domain $\mathcal{H}$, that is, $p_{\mathcal{H}} = \mathbb{P}[\mathcal{H}] = \mathbb{P}[h(\theta, v) \le 0]$. This probability, which is obtained as a by-product of BUS, is used to estimate the model evidence as[30]

$$Z_{\widetilde{y}} = c^{-1} p_{\mathcal{H}} = \overline{c}\, p_{\mathcal{H}}. \tag{11}$$

Note that the application of BUS requires the knowledge of the constant $c = 1/\overline{c}$. From (9), it is seen that the covering constant is optimally chosen as the maximum of the likelihood function $\overline{c} = L_{\max}$. If $\overline{c} < L_{\max}$, the resulting samples will be distributed according to a truncated posterior distribution.[44] Conversely, if $\overline{c} > L_{\max}$, the efficiency of BUS decreases because the value of $p_{\mathcal{H}}$ will be small, requiring more samples for its estimation. Since in many cases $L_{\max}$ is not known in advance and its computation poses an additional cost, we employ the strategy introduced in Reference 44 where the constant c is adaptively computed at each step of the simulation. We discuss this method in the variable-dimensional context in subsection 3.3.

3.2 | BUS with subset simulation in fixed dimensions

In order to efficiently compute samples from the posterior distribution, BUS is often combined with subset simulation (SuS).[31] The combination of BUS with SuS (denoted BUS-SuS), performs Bayesian inversion sequentially. This is because SuS transforms the task of estimating the rare event $\{h(\theta, v) \le 0\}$ into a sequence of problems involving more frequent events.

In BUS-SuS, the parameter space Θ is divided into a decreasing sequence of nested subsets or *intermediate levels*, starting from the whole space and narrowing down to the target posterior space, that is, $\Theta = \mathcal{H}_0 \supset \mathcal{H}_1 \supset \cdots \supset \mathcal{H}_{N_{\text{lv}}} = \mathcal{H}$, such that $\mathcal{H} = \bigcap_{j=0}^{N_{\text{lv}}} \mathcal{H}_j$, where N_{lv} is the number of intermediate subsets. Based on the general product rule of probability, the probability that the prior samples fall into the posterior space, $p_{\mathcal{H}}$ is given by

$$p_{\mathcal{H}} = \mathbb{P}\left[\cap_{j=0}^{N_{\text{lv}}} \mathcal{H}_j\right] = \mathbb{P}\left[\mathcal{H}_{N_{\text{lv}}} \mid \cap_{j=0}^{N_{\text{lv}}-1} \mathcal{H}_j\right] \mathbb{P}\left[\cap_{j=0}^{N_{\text{lv}}-1} \mathcal{H}_j\right] = \prod_{j=1}^{N_{\text{lv}}} \mathbb{P}\left[\mathcal{H}_j \mid \mathcal{H}_{j-1}\right], \tag{12}$$

where $\mathbb{P}\left[\mathcal{H}_j \mid \mathcal{H}_{j-1}\right]$ represents the conditional probability at level $(j-1)$. Each intermediate level is defined as the set $\mathcal{H}_j = \{(\theta, v) \in \Theta : h(\theta, v) \le \tau_j\}$, where $\infty = \tau_0 > \tau_1 > \cdots > \tau_j > \cdots > \tau_{N_{\text{lv}}} = 0$, is a decreasing sequence of threshold levels. In practice, it is not possible to make an optimal a priori selection of the sequence $\{\tau_j\}_{j=0}^{N_{\text{lv}}}$. Therefore, they are adaptively selected as the p_0-percentile of the LSF values of the samples simulated at intermediate level $\mathcal{H}_{j-1}$.[31] This implies fixing the conditional probabilities to a common value $p_0 = \mathbb{P}\left[\mathcal{H}_j | \mathcal{H}_{j-1}\right]$ (with usually chosen as $p_0 \in [0.1, 0.3]$).

At the first subset $\mathcal{H}_0$, samples are generated from the prior distribution using standard Monte Carlo. Thereafter, BUS-SuS employs a modified MCMC algorithm to draw samples from each intermediate conditional density $\pi(\theta, v \mid \mathcal{H}_j)$. The Markov chains are initialized from $N_s = N p_0$ samples conditional on $\mathcal{H}_{j-1}$ for which $h(\theta, v) \le \tau_j$. The process is repeated until the target posterior domain is reached (see, e.g., Reference 45 for further details). At the last subset, the probability $p_{\mathcal{H}}$ in (12) is estimated as $\widehat{p}_{\mathcal{H}} = p_0^{N_{\text{lv}}-1} \widehat{p}_{N_{\text{lv}}}$, where $\widehat{p}_{N_{\text{lv}}}$ represents the last conditional probability, which is estimated by Monte Carlo as the ratio of the number of samples that lie in $\mathcal{H}$ and the number of samples per level N. The $(\widehat{p}_{N_{\text{lv}}} N)$ samples that lie in $\mathcal{H}$ are used as seeds to generate the final batch of N draws conditional on $\mathcal{H}$. The resulting samples are uniformly weighted but correlated samples of the posterior distribution, and the probability estimate $\widehat{p}_{\mathcal{H}}$ is used to compute the model evidence via (11).

Remark 2. It is common practice to solve reliability problems in the standard Gaussian space. Due to the BUS formulation, this also translates to the Bayesian inversion. Hence, a new standard Gaussian parameter vector $\vartheta = [\theta, \theta_v]^T \in \mathbb{R}^{k+1}$ is created by combining the KL coefficients θ and the transformed auxiliary uniform variable $\theta_v = \Phi^{-1}(v)$, where $\Phi(\cdot)$

denotes the standard Gaussian cumulative distribution. Furthermore, in order to guarantee a smooth transition between the intermediate levels, as well as for numerical stability, the LSF in the right-hand side of (10) is expressed in terms of the log-likelihood. Applying the natural logarithm to each term of $h(\boldsymbol{\theta}, v)$ in (10) yields[46]

$$h_{\ln}(\boldsymbol{\vartheta}) = \ln(\Phi(\theta_v)) - \ln(c\, \mathrm{L}\left(\boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right)) = \ln(\Phi(\theta_v)) + \ln(\overline{c}) - \ln \mathrm{L}\left(\boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right). \qquad (13)$$

3.3 | BUS with subset simulation in varying dimensions

We now extend the concepts of subsection 3.1 and 3.2 to the variable-dimensional case. The basic idea is to re-augment the parameter space by including the discrete dimension variable. This requires minor modifications of the target LSF, and the application of trans-dimensional MCMC algorithms to sample the intermediate conditional densities. We denote this trans-dimensional BUS-SuS methodology as *tBUS-SuS*.

Consider the general Bayesian inverse problem (4). The joint posterior distribution is characterized by a target function in a discrete-continuous space, $\pi(k, \boldsymbol{\theta}) = \overline{\pi}_{\mathrm{pr}}(k)\pi_{\mathrm{pr}}(\boldsymbol{\theta}_k \mid k)\mathrm{L}\left(k, \boldsymbol{\theta}_k; \widetilde{\boldsymbol{y}}\right) \propto \overline{\pi}_{\mathrm{pos}}\left(k, \boldsymbol{\theta}_k \mid \widetilde{\boldsymbol{y}}\right)$. We choose the proposal distribution to be equal to the joint prior $q(k, \boldsymbol{\theta}_k) = \overline{\pi}_{\mathrm{pr}}(k)\pi_{\mathrm{pr}}(\boldsymbol{\theta}_k \mid k)$. The acceptance probability in rejection sampling becomes

$$\alpha = \frac{\pi(k, \boldsymbol{\theta})}{\overline{r}\, q(k, \boldsymbol{\theta})} = \frac{\overline{\pi}_{\mathrm{pr}}(k)\pi_{\mathrm{pr}}(\boldsymbol{\theta}_k \mid k)\mathrm{L}\left(k, \boldsymbol{\theta}_k; \widetilde{\boldsymbol{y}}\right)}{\overline{r}\, \overline{\pi}_{\mathrm{pr}}(k)\, \pi_{\mathrm{pr}}(\boldsymbol{\theta}_k \mid k)} = r\, \mathrm{L}\left(k, \boldsymbol{\theta}_k; \widetilde{\boldsymbol{y}}\right), \qquad (14)$$

where $r = 1/\overline{r} \in (0, \infty)$. By analogy with the fixed-dimensional setting, the covering constant $\overline{r}$ can be optimally chosen as $L_{\mathrm{max,all}} = \max(\mathrm{L}\left(k, \boldsymbol{\theta}_k; \widetilde{\boldsymbol{y}}\right))$, that is, the maximum of the likelihood function across different dimensions. Thereafter, samples drawn from the priors $k \sim \overline{\pi}_{\mathrm{pr}}(\cdot)$ and $\boldsymbol{\theta}_k \sim \pi_{\mathrm{pr}}(\cdot \mid k)$ are accepted if $v \le \alpha = r\, \mathrm{L}\left(k, \boldsymbol{\theta}_k; \widetilde{\boldsymbol{y}}\right)$, otherwise they are rejected. In this case, the $\mathcal{H}$-space and the LSF in (10) are re-defined as

$$\mathcal{H} = \left\{ (k, \boldsymbol{\theta}_k, v) \in \overline{\boldsymbol{\Theta}} : h(k, \boldsymbol{\theta}_k, v) \le 0 \right\}, \quad \text{where} \quad h(k, \boldsymbol{\theta}_k, v) = v - r\, \mathrm{L}\left(k, \boldsymbol{\theta}_k; \widetilde{\boldsymbol{y}}\right) \qquad (15)$$

and $\overline{\boldsymbol{\Theta}} = [K, \Theta_k, \Upsilon]$ is the re-augmented discrete-continuous parameter space ($k \in K \subseteq \mathbb{Z}_{>0}$, $\boldsymbol{\theta}_k \in \Theta_k \subseteq \mathbb{R}^k$ and $v \in \Upsilon \subseteq [0, 1]$). As it will be seen in Section 4, we employ trans-dimensional MCMC algorithms that work in a saturated space for which $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{k_{\max}}$, and thus we re-define the parameter space without dependence on k, that is, $\overline{\boldsymbol{\Theta}} = [K, \Theta, \Upsilon]$.

The BUS-SuS algorithm can be extended analogously to solve the Bayesian inverse problem associated to (15). Each intermediate domain is now defined as the set $\mathcal{H}_j = \{(k, \boldsymbol{\theta}, v) \in \overline{\boldsymbol{\Theta}} : h(k, \boldsymbol{\theta}, v) \le \tau_j\}$, with the threshold level sequence $\{\tau_j\}_{j=0}^{N_{\mathrm{lv}}}$ adaptively selected as in the fixed-dimensional case. Under the LSF in the right-hand side of (15), the standard MCMC algorithms used within BUS-SuS are no longer suitable for sampling the intermediate densities $\pi(k, \boldsymbol{\theta}, v \mid \mathcal{H}_j)$. Trans-dimensional MCMC methods are required to perform such task. In the tBUS-SuS case, we modify this class of samplers to simulate the intermediate densities conditional on events defined by the sequence of levels $\{\tau_j\}$; this will be shown in Section 4. Moreover, instead of the LSF in (13), we employ its variable-dimensional extension

$$h_{\ln}(k, \boldsymbol{\vartheta}) = \ln(\Phi(\theta_v)) + \overline{\mathscr{r}} - \ln \mathrm{L}\left(k, \boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right), \qquad (16)$$

where $\boldsymbol{\vartheta} = [\boldsymbol{\theta}, \theta_v] \in \mathbb{R}^{k_{\max}+1}$, and $\overline{\mathscr{r}} = \ln(\overline{r})$ is optimally the maximum of the log-likelihood function across the different dimensions. We summarize the tBUS-SuS method in Algorithm 1.

Since finding the constant $\overline{\mathscr{r}} = \ln(\overline{r})$ poses an additional computational cost, it is convenient to introduce a tBUS-SuS algorithm for which the covering constant $\overline{r}$ is not required as an input. We employ the adaptive BUS-SuS methodology proposed in[44] for the trans-dimensional setting. In this case, the covering constant is updated at each level, leading to a set of values $\{\overline{r}_j\}_{j=0}^{N_{\mathrm{lv}}}$. In order to guarantee the nestedness of the intermediate domains, the threshold levels τ_j are corrected after updating the value $\overline{r}_j$. First note that from (16), a jth intermediate domain is defined as the set

$$\mathcal{H}_j = \{(k, \boldsymbol{\vartheta}) : \ln(\Phi(\theta_v)) + \overline{\mathscr{r}}_j - \ln \mathrm{L}\left(k, \boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right) \le \tau_j\}, \qquad (17)$$

where $\overline{\mathscr{r}}_j = \ln(\overline{r}_j)$ is the maximum of the log-likelihood function observed at the jth sampling level. The idea of Reference 44 is that the event $\mathcal{H}_j$ associated with the log-scaling constant $\overline{\mathscr{r}}_j$ and the threshold τ_j, can be equivalently expressed

---

Algorithm 1. tBUS-SuS

1: **Input**: number of samples per level N, conditional probability p_0, covering constant $\overline{r}$, maximum dimension $k_{\max}$, log-likelihood function $\ln L\left(\cdot, \cdot; \widetilde{y}\right)$, dimension prior $\overline{\pi}_{\text{pr}}(k)$

2: Draw N samples from the dimension prior, $k_0 \sim \overline{\pi}_{\text{pr}}(\cdot)$

3: Draw N samples from the standard Gaussian distribution, $\boldsymbol{\vartheta}_0 = [\theta_0, \theta_{v,0}] \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_{\max}+1})$

4: Compute the initial log-likelihood function values, $L_{\text{eval}} \leftarrow \ln L\left(k_0, \theta_0; \widetilde{y}\right)$

5: Set $j \leftarrow 0$ and $\tau_0 \leftarrow \infty$

6: **while** $\tau_j > 0$ **do**

7:     Increase intermediate level counter, $j \leftarrow j + 1$

8:     Compute LSF values, $h_{\text{eval}} \leftarrow \ln(\Phi(\theta_{v,j-1})) + \overline{r} - L_{\text{eval}}$

9:     Sort h_{eval} in ascending order and create a vector `idx` to store the indices of this sorting

10:     Create $k_{\text{sort}}, \boldsymbol{\vartheta}_{\text{sort}}$ as the dimension and parameter samples $k_{j-1}, \boldsymbol{\vartheta}_{j-1}$ sorted according to `idx`

11:     Set the intermediate threshold level τ_j as the p_0-percentile of the values in h_{eval}

12:     Compute the number of samples in the jth intermediate level, $N_{\mathcal{H}_j} \leftarrow \sum_{i=1}^{N}(h_{\text{eval}}^{(i)} \leq \max(0, \tau_j))$

13:     **if** $\tau_j > 0$ **then**

14:         $p_{\mathcal{H}_{j-1}} \leftarrow p_0$

15:     **else**

16:         $\tau_j \leftarrow 0$ and $p_{\mathcal{H}_{j-1}} \leftarrow N_{\mathcal{H}_j}/N$

17:     **end if**

18:     Select seeds for the MCMC step, $(k_{\text{seed}}, \boldsymbol{\vartheta}_{\text{seed}}) \leftarrow \{k_{\text{sort}}^{(i)}, \boldsymbol{\vartheta}_{\text{sort}}^{(i)}\}_{i=1}^{N_{\mathcal{H}_j}}$

19:     Generate next level values $\{k_j^{(i)}, \boldsymbol{\vartheta}_j^{(i)}, L_{\text{eval}}^{(i)}\}_{i=1}^{N}$ from the seeds $(k_{\text{seed}}, \boldsymbol{\vartheta}_{\text{seed}})$ and intermediate level τ_j using a trans-dimensional MCMC algorithm. Here, each seed is used to construct a chain with $N_c = \texttt{floor}(N/N_s)$ states, where $N_s = N_{\mathcal{H}_j}$ is the number of seeds

20: **end while**

21: Set the posterior samples, $k_{\text{pos}} \leftarrow k_j$ and $\boldsymbol{\vartheta}_{\text{pos}} \leftarrow \boldsymbol{\vartheta}_j$

22: **for** $k \leftarrow 1$ *to* $k_{\max}$ **do**

23:     Find the number of posterior samples that lie in dimension k, $N_k \leftarrow \sum_{i=1}^{N}(k_{\text{pos}}^{(i)} = k)$

24:     Estimate the model posterior, $\widehat{\overline{\pi}}_{\text{pos}}^{(k)} \leftarrow N_k/N$

25: **end for**

26: **Output**: sequence of posterior dimension samples $\{k_{\text{pos}}^{(i)}\}_{i=1}^{N}$ and parameter samples $\{\boldsymbol{\vartheta}_{\text{pos}}^{(i)}\}_{i=1}^{N}$, and model posterior $\widehat{\overline{\pi}}_{\text{pos}}$.

---

by a scaling $\overline{r}_j'$ and a modified threshold τ_j' selected as $\tau_j' = \tau_j - \overline{r}_j + \overline{r}_j'$. This allows one to sequentially update the covering constant $\overline{r}_j$ to a new value $\overline{r}_j'$, without compromising the distribution of the samples. Adjusting the threshold value from τ_j to τ_j' (after updating the scaling from $\overline{r}_j$ for $\overline{r}_j'$) defines the same intermediate domain since $\mathcal{H}_j = \left\{(k, \boldsymbol{\vartheta}) : \ln(\Phi(\theta_v)) \leq \tau_j' - \overline{r}_j' + \ln L\left(k, \theta; \widetilde{y}\right)\right\}$ is equivalent to (17). In principle, τ_j' corrects the level τ_j using the residual of maximum log-likelihood values observed at different levels. At the last simulation level, when $\overline{r}_j' = \overline{r}_j$, the covering reaches a value that is close or equal to the actual maximum log-likelihood, that is, $\overline{r}_{N_{\text{lv}}} \leq \ln(L_{\max,\text{all}})$. In the limit $N \to \infty$, the value $\overline{r}_{N_{\text{lv}}}$ converges to $\ln(L_{\max,\text{all}})$. Despite the fact that $\overline{r}_{N_{\text{lv}}}$ is likely to be smaller than $\ln(L_{\max,\text{all}})$, the samples generated by the algorithm follow the posterior distribution; this is shown in Reference 44. The adaptive tBUS-SuS method is described in Algorithm 2.

4 | MCMC ALGORITHMS IN SPACES OF VARYING DIMENSION

In variable-dimensional problems, MCMC methods must explore a discrete-continuous parameter space. In this section, we present an overview of such algorithms and discuss two special MCMC samplers that are used in combination with the proposed tBUS-SuS algorithm. We note that the methods discussed here are applicable to model updating problems whenever the variables of the different models have a nested structure (see, e.g., Reference 13,14).

Algorithm 2. Adaptive tBUS-SuS

1: **Input**: number of samples per level N, conditional probability p_0, maximum dimension $k_{\max}$, log-likelihood function $\ln L\left(\cdot, \cdot; \widetilde{\boldsymbol{y}}\right)$, dimension prior $\overline{\pi}_{\mathrm{pr}}(k)$

2: Repeat Lines 2-4 of Algorithm 1

3: Compute the initial maximum log-likelihood, $\overline{\mathscr{r}}_0 \leftarrow \max(L_{\mathrm{eval}})$

4: Set $j \leftarrow 0$ and $\tau_0 \leftarrow \infty$

5: **while** $\tau_j > 0$ **do**

6:     Increase intermediate level counter, $j \leftarrow j + 1$

7:     Compute LSF values, $h_{\mathrm{eval}} \leftarrow \ln(\Phi(\theta_{v,j-1})) + \overline{\mathscr{r}}_j - L_{\mathrm{eval}}$

8:     Repeat Lines 9-19 of Algorithm 1

9:     Compute a new value of the maximum log-likelihood, $\overline{\mathscr{r}}_j' \leftarrow \max\left(\overline{\mathscr{r}}_j, \{L_{\mathrm{eval}}^{(i)}\}_{i=1}^N\right)$

10:     Compute the modified intermediate threshold level, $\tau_j \leftarrow \tau_j - \overline{\mathscr{r}}_j + \overline{\mathscr{r}}_j'$, and update $\overline{\mathscr{r}}_j \leftarrow \overline{\mathscr{r}}_j'$

11: **end while**

12: Repeat Lines 21-25 and Output of Algorithm 1

4.1 | General remarks

In across-model simulation, the standard algorithm to sample from the joint posterior in (4) is the *reversible jump* MCMC (RJMCMC) method.[9] The idea is to generate a Markov chain that is able to jump between models with parameter spaces of different dimension. If the current and proposed states have the same dimension, the proposal move explores different locations within the same parameter space. In this case, the so-called detailed balance condition is guaranteed by a standard MCMC sampler.[9] If the current and proposed dimensions are different, the detailed balance holds by defining a proposal move that satisfies a dimension matching condition. This is achieved by constructing a one-to-one deterministic transformation (jumping function) ensuring that the image and the domain of the transformation have the same dimension. The acceptance probability in RJMCMC resembles the one of the classical Metropolis–Hastings algorithm, where the proposal distribution is decomposed into a discrete density for the dimension, a continuous density for the parameters, and the Jacobian of the jumping transformation is also taken into account (see Reference 41 for further details). The RJMCMC can suffer from poor sampling performance associated to the definition of the jumping function and the proposal distribution. The potential inefficiency of the method has motivated several tuning step procedures (see, e.g., References 47,48).

Another class of algorithms are the *saturated space* approaches[11,13,14,47] (also referred to as product or composite space). The main characteristic is that the parameter space is not particularized to a given dimension k, instead the parameters lie in a space whose dimension contains all dimensions of interest, that is, $k_{\max}$. The joint posterior in the saturated space is Reference 14

$$\pi_{\mathrm{pos}}(k, \boldsymbol{\theta} \mid \widetilde{\boldsymbol{y}}) = \frac{1}{\overline{Z}_{\widetilde{\boldsymbol{y}}}} \overline{\pi}_{\mathrm{pr}}(k) \pi_{\mathrm{pr}}(\boldsymbol{\theta}_k \mid k) \pi_{\mathrm{pr}}(\boldsymbol{\theta}_{\sim k} \mid k, \boldsymbol{\theta}_k) L\left(k, \boldsymbol{\theta}_k; \widetilde{\boldsymbol{y}}\right), \tag{18}$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_k, \boldsymbol{\theta}_{\sim k}]$ and the additional component is the so-called linking density or *pseudo-prior* $\pi_{\mathrm{pr}}(\boldsymbol{\theta}_{\sim k} \mid k, \boldsymbol{\theta}_k)$, where the $\boldsymbol{\theta}_{\sim k}$ denotes the parameters that are not used by the model k. This formulation allows one to employ standard MCMC procedures to variable-dimensional problems. We now motivate these techniques from the viewpoint of nested and variable selection models.

4.2 | Nested and variable selection models

For problems involving *nested models* the parameters of the model k have the same interpretation as the first k parameters of the model $k + 1$. These are also related to *variable selection models*, where the dimension change amounts to switch on and off components in the parameter vector, that is, exclusion of a component is equivalent to setting a parameter to zero.[14] Under the KL expansion (1), the likelihood in (4) is independent of θ_i when $i > k$, and the prior of parameter vector $\boldsymbol{\theta}$ in the saturated space becomes independent of k. Therefore, the (saturated) joint posterior in (18) can be

written as

$$\pi_{\text{pos}}(k, \boldsymbol{\theta} \mid \widetilde{\boldsymbol{y}}) \propto \overline{\pi}_{\text{pr}}(k)\pi_{\text{pr}}(\boldsymbol{\theta})\text{L}\left(k, \boldsymbol{\theta}; \widetilde{\boldsymbol{y}}\right), \tag{19}$$

with $\pi_{\text{pr}}(\boldsymbol{\theta})$ denoting the prior of the parameters in the saturated space $\Theta \subseteq \mathbb{R}^{k_{\max}}$. In the context of tBUS-SuS, the posterior distribution (19) can be re-written by conditioning on the region $\mathcal{H}$ in (15) and marginalizing over the auxiliary uniform random variable υ,

$$\pi_{\text{pos}}(k, \boldsymbol{\theta} \mid \widetilde{\boldsymbol{y}}) \propto \overline{\pi}_{\text{pr}}(k)\pi_{\text{pr}}(\boldsymbol{\theta}) \int_0^1 \mathbb{1}_{\mathcal{H}}(k, \boldsymbol{\theta}, \upsilon) \, \mathrm{d}\upsilon, \tag{20}$$

where $\mathbb{1}_{\mathcal{H}}$ is the indicator function, which is equal to one, if $(k, \boldsymbol{\theta}, \upsilon) \in \mathcal{H}$, and zero otherwise. Due to the sequential structure of tBUS-SuS, we require MCMC algorithms that sample conditional densities on each intermediate domain $\mathcal{H}_j$, that is, $\pi(k, \boldsymbol{\theta}, \upsilon \mid \mathcal{H}_j) = \overline{\pi}_{\text{pr}}(k)\pi_{\text{pr}}(\boldsymbol{\theta})\mathbb{1}_{\mathcal{H}_j}(k, \boldsymbol{\theta}, \upsilon)$. Particularly, we work in a saturated standard Gaussian space in which the KL coefficients and the auxiliary variable can be grouped to define the parameter vector $\boldsymbol{\vartheta} = [\boldsymbol{\theta}, \theta_\upsilon]$ (cf. Remark 2). As a result, the intermediate densities are defined as $\pi(k, \boldsymbol{\vartheta} \mid \mathcal{H}_j)$, the discrete dimension space is $K \subseteq \mathbb{Z}_{[1,k_{\max}]}$, the saturated parameter space is $\Theta \subseteq \mathbb{R}^{k_{\max}+1}$, and the full discrete-continuous space is re-defined as $\overline{\Theta} = [K, \Theta]$.

The saturated space is oftentimes high-dimensional in random field applications. In order to avoid convergence deterioration with increasing k, dimension-independent MCMC algorithms are applied. These samplers are based on numerical discretizations to stochastic differential equations (SDE) that preserve the reference prior or posterior measures. A main requirement for an MCMC algorithm to be dimension-independent is that of being well-defined in function spaces. For instance, the *preconditioned Crank–Nicolson* (pCN) algorithm is derived in Reference 33 by discretizing a prior-preconditioned overdamped Langevin dynamic SDE using a Crank–Nicolson scheme. Given the high-dimensional nature of random fields and the structure of the KL expansion, we focus on saturated space approaches for which the pCN algorithm can be utilized, namely: the step-wise and Metropolis-within-Gibbs algorithms.

4.2.1 | Step-wise sampler

We construct a step-wise algorithm based on the pCN proposal that only requires one acceptance probability step for both, dimension and parameters. The foundations and converge properties of this sampler follow from References 13,14. Consider a proposal density $q(k, \boldsymbol{\vartheta}) = q_1(k) \, q_2(\boldsymbol{\vartheta})$ across the full state space $\overline{\Theta}$. This density takes into account the proposal for the dimension $q_1(k)$ and the proposal for the parameters in the saturated space $q_2(\boldsymbol{\vartheta})$. Under these assumptions the acceptance probability of the standard Metropolis–Hastings algorithm becomes[14]

$$\alpha(k, \boldsymbol{\vartheta}; k^\star, \boldsymbol{\vartheta}^\star) = \min \left\{ 1, \frac{q_1(k^\star) \, q_2(\boldsymbol{\vartheta}^\star) \, \pi(k^\star, \boldsymbol{\vartheta}^\star \mid \widetilde{\boldsymbol{y}})}{q_1(k) \, q_2(\boldsymbol{\vartheta}) \, \pi(k, \boldsymbol{\vartheta} \mid \widetilde{\boldsymbol{y}})} \right\}. \tag{21}$$

We employ the pCN proposal for the parameter vector $\boldsymbol{\vartheta}$ in the saturated space. In this case, the proposal $q_2(\boldsymbol{\vartheta})$ cancels out with the saturated parameter prior in the target posterior (see, e.g., Reference 33). Moreover, since k is a discrete variable, the proposal distribution for the dimension $q_1(k)$ can be represented as a proposal matrix $\mathbf{Q} \in \mathbb{R}^{k_{\max} \times k_{\max}}$. This is a right-stochastic matrix containing the probabilities of the moves. Such probabilities can be assigned using a discrete probability law controlled by a *spread parameter* $\rho \in [1, k_{\max}]$, defining the width or jump lengths of the proposal. The resulting acceptance probability simplifies to

$$\alpha(k, \boldsymbol{\vartheta}; k^\star, \boldsymbol{\vartheta}^\star) = \min \left\{ 1, \frac{\text{L}\left(k^\star, \boldsymbol{\vartheta}^\star; \widetilde{\boldsymbol{y}}\right)}{\text{L}\left(k, \boldsymbol{\vartheta}; \widetilde{\boldsymbol{y}}\right)} \frac{\overline{\pi}_{\text{pr}}(k^\star) \, \mathbf{Q}(k^\star, k)}{\overline{\pi}_{\text{pr}}(k) \, \mathbf{Q}(k, k^\star)} \right\}, \tag{22}$$

which in the context of tBUS-SuS is equivalent to

$$\alpha(k, \boldsymbol{\vartheta}; k^\star, \boldsymbol{\vartheta}^\star) = \min \left\{ 1, \mathbb{1}_{\mathcal{H}_j}\left(k^\star, \boldsymbol{\vartheta}^\star\right) \frac{\overline{\pi}_{\text{pr}}(k^\star) \, \mathbf{Q}(k^\star, k)}{\overline{\pi}_{\text{pr}}(k) \, \mathbf{Q}(k, k^\star)} \right\} = \mathbb{1}_{\mathcal{H}_j}\left(k^\star, \boldsymbol{\vartheta}^\star\right) \underbrace{\min \left\{ 1, \frac{\overline{\pi}_{\text{pr}}(k^\star) \, \mathbf{Q}(k^\star, k)}{\overline{\pi}_{\text{pr}}(k) \, \mathbf{Q}(k, k^\star)} \right\}}_{(*)}. \tag{23}$$

Algorithm 3. State update in the step-wise sampler for tBUS-SuS in the standard Gaussian space

---

1: **Input**: Let $(k, \boldsymbol{\vartheta})$ be the current state of the Markov chain and β the pCN proposal scaling

2: /* Step 1: sample the dimension */

3: Draw a candidate dimension $k^\star \sim \mathbf{Q}(k, :)$

4: /* Step 2: sample the coefficients using pCN */

5: Draw candidate parameters, $\boldsymbol{\vartheta}^\star \leftarrow \sqrt{1 - \beta^2}\, \boldsymbol{\vartheta} + \beta\boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_{\max}+1})$

6: Compute the acceptance probability $\alpha_{k,\boldsymbol{\vartheta}}$ as per Equation (23)

7: Sample, $U_{k,\boldsymbol{\vartheta}} \sim \text{Unif}(0, 1)$

8: **if** $U_{k,\boldsymbol{\vartheta}} < \alpha_{k,\boldsymbol{\vartheta}}$ **then**

9: $\quad k_{\text{next}} \leftarrow k^\star$ and $\boldsymbol{\vartheta}_{\text{next}} \leftarrow \boldsymbol{\vartheta}^\star$

10: **else**

11: $\quad k_{\text{next}} \leftarrow k$ and $\boldsymbol{\vartheta}_{\text{next}} \leftarrow \boldsymbol{\vartheta}$

12: **end if**

13: **Output**: $(k_{\text{next}}, \boldsymbol{\vartheta}_{\text{next}})$

This Metropolis–Hastings implementation on the saturated space proceeds in a *step-wise* manner as follows: in the first step, a candidate dimension $k^\star$ is proposed according to the proposal matrix $\mathbf{Q}$. In the second step, a candidate parameter $\boldsymbol{\vartheta}^\star$ is proposed using the pCN.[33] Afterwards, the candidate pair $(k^\star, \boldsymbol{\vartheta}^\star)$ is rejected or accepted jointly according to the probability (23). Algorithm 3 describes this procedure in detail. Note that we can alternatively implement the right term of (23), such that (i) a model $k^\star$ is proposed and accepted with probability (∗) in (23), (ii) $\theta^\star$ is drawn from a pCN proposal, and (iii) the pair $(k^\star, \theta^\star)$ is accepted, if it lies in the domain $\mathcal{H}_j$ using the indicator function.

4.2.2 | Metropolis-within-Gibbs sampler

The Metropolis-within-Gibbs (MwG) algorithm[49] updates the parameters $\boldsymbol{\vartheta}$ and the dimension k in an alternating manner. In the saturated space, MwG explores the joint posterior using a Gibbs sampling version of the algorithm in Reference 11, after including Metropolis–Hastings steps (details are provided in Reference 14). The algorithm can also be derived by writing the posterior (20) as the product of the dimension posterior and the dimension-specific parameter posterior[9]

$$\pi_{\text{pos}}(k, \boldsymbol{\vartheta} \mid \widetilde{\boldsymbol{y}}) = \overline{\pi}(k \mid \widetilde{\boldsymbol{y}})\pi(\boldsymbol{\vartheta} \mid k, \widetilde{\boldsymbol{y}}). \tag{24}$$

The densities $\pi(\boldsymbol{\vartheta} \mid k, \widetilde{\boldsymbol{y}})$ may differ abruptly for small changes in the variable k and thus, the chain might always remain in some state. However, under the KL formulation (1), the coefficients and the dimension are independent *a priori*. This property alleviates potential poor mixing properties in MwG.[50]

The idea of MwG is to sample each conditional density in (24). Recall that for tBUS-SuS, these densities need to be defined with respect to the intermediate levels $\mathcal{H}_j$. In the first step, we fix the parameter $\boldsymbol{\vartheta}$ and employ the conditional distribution $\overline{\pi}(k \mid \cdot)$ to propose a candidate dimension $k^\star$ using a Metropolis–Hastings sampler. In the second step, we fix the variable k (accepted in the first step), and sample the conditional distribution $\pi(\boldsymbol{\vartheta} \mid \cdot)$ to obtain a candidate parameter $\boldsymbol{\vartheta}^\star$ using the pCN proposal. The state update in MwG for tBUS-SuS is formally described in Algorithm 4. Observe that this approach requires two LSF (likelihood) evaluations for the generation of one chain state.

Remark 3. We apply an adaptive version of the pCN algorithm used within the trans-dimensional Algorithms 3 and 4. The idea is to control the pCN scaling β to keep the acceptance rate around a near-optimal value through the simulation. The optimality is defined in terms of the smallest error in the approximation of the model posterior. The adaptation procedure follows from Reference 45.

5 | NUMERICAL EXAMPLES

We test the proposed method on two examples. The first problem allows to verify the approximations performed by tBUS-SuS, since a reference model posterior can be computed analytically. In the second example, a closed-form

---

Algorithm 4. State update in the MwG sampler for tBUS-SuS in the standard Gaussian space

---

1: **Input**: Let (k, ϑ) be the current state of the Markov chain and β the pCN proposal scaling

2: /* Step 1: for fixed ϑ, sample the conditional distribution $\overline{\pi}(k\,|\,\widetilde{y})$ */

3: Sample the dimension, $k^\star \sim \mathbf{Q}(k, :)$

4: Compute the acceptance probability

$$\alpha_k \leftarrow \mathbb{1}_{\mathcal{H}_j}\left(k^\star, \vartheta\right) \min\left\{1, \frac{\overline{\pi}_{\mathrm{pr}}(k^\star)\,\mathbf{Q}(k^\star, k)}{\overline{\pi}_{\mathrm{pr}}(k)\,\mathbf{Q}(k, k^\star)}\right\}$$

5: Sample, $U_k \sim \mathrm{Unif}(0, 1)$

6: **if** $U_k < \alpha_k$ **then**

7: $\quad k_{\mathrm{next}} \leftarrow k^\star$

8: **else**

9: $\quad k_{\mathrm{next}} \leftarrow k$

10: **end if**

11: /* Step 2: for fixed k_{next}, sample the conditional distribution $\pi(\vartheta\,|\,k_{\mathrm{next}}, \widetilde{y})$ */

12: Sample the parameters using pCN, $\vartheta^\star \leftarrow \sqrt{1 - \beta^2}\,\vartheta + \beta\xi$, where $\xi \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_{\max}+1})$

13: Compute the acceptance probability $\alpha_\vartheta \leftarrow \mathbb{1}_{\mathcal{H}_j}\left(k_{\mathrm{next}}, \vartheta^\star\right)$

14: **if** $\alpha_\vartheta = 1$ **then**

15: $\quad \vartheta_{\mathrm{next}} \leftarrow \vartheta^\star$

16: **else**

17: $\quad \vartheta_{\mathrm{next}} \leftarrow \vartheta$

18: **end if**

19: **Output**: $(k_{\mathrm{next}}, \vartheta_{\mathrm{next}})$

---

expression is not available. Thus, we compute several posterior dimension snapshots using a within-model BUS-SuS approach to verify the solution estimated by tBUS-SuS. In all cases, the intermediate conditional probabilities are fixed at $p_0 = 0.1$. The methods presented in the article are implemented in MATLAB® R2018b, while the results are generated in Matplotlib v3.0.1.

5.1 | 1D cantilever beam

The first example involves an ordinary differential equation (ODE) that describes the equilibrium of a cantilever beam. In this case, the solution of the Bayesian inverse problem can be derived analytically.[29] The physical domain is the interval $D = [0, L]$, where $L = 5$ m is the length of the beam. The beam is subjected to a deterministic point load $P = 20$ kN at its free right end. The vertical displacements are constrained at the left edge of the beam. Let $F(x) = (IE(x))^{-1}$ denote the flexibility of the beam (with $x \in D$), where E is the elastic modulus, and I the moment of inertia. The deflection response $w(x)$, for a given flexibility and load configuration, is governed by the Euler–Bernoulli ODE:

$$M(x) = -F^{-1}(x)\frac{\mathrm{d}^2 w(x)}{\mathrm{d}x^2} \quad \Rightarrow \quad w(x) = -P\int_0^x \int_0^s (L - t)F(t)\,\mathrm{d}t\mathrm{d}s, \tag{25}$$

here we use the fact that the bending moment of a cantilever beam is $M(x) = (L - x)P$.

The flexibility is modeled by a Gaussian random field prior $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \boldsymbol{\Sigma}_{\mathrm{pr}})$, with constant mean $\mu_{\mathrm{pr}} = 10^{-4}$ (kN $^{-1}$m $^{-2}$) and covariance matrix $\boldsymbol{\Sigma}_{\mathrm{pr}}$ defined through a Matérn kernel with smoothing parameter $\nu = 0.5$, which yields the following exponential autocovariance function, $C(x, x') = \sigma_{\mathrm{pr}}^2 \exp\left(-|x - x'|/\ell\right)$, for $x, x' \in D$. We set the prior standard deviation to $\sigma_{\mathrm{pr}} = 0.35\mu_{\mathrm{pr}} = 3.5 \times 10^{-5}$ and perform a parameter study on the correlation length ℓ. Note that for the exponential covariance kernel, the KL eigenvalue problem has analytical solution.[27]

The *true* flexibility field is a realization from the prior random field (with correlation length $\ell_{\mathrm{true}} = 2$ m). The realization is computed from the full random field using a finely discretized mean and covariance (501 equally spaced points). Partial observations of the deflection field are generated by simulating the ODE (25) using this underlying realization.

The data is collected at $m = 10$ equally-spaced points of the domain D (Figure 1). This generates a measurement vector $\widetilde{y} \in \mathbb{R}^m$ with additive and spatially correlated error described by a Gaussian PDF, $\eta \sim \mathcal{N}(0, \Sigma_{obs})$, where the covariance of the error is constructed from an exponential kernel with standard deviation $\sigma_{obs} = 10^{-3}$ and correlation length $\ell_{obs} = 1$ m.

In this example, closed-form expressions of the model evidence for each dimension k are available (see, e.g., Reference 29), this allows us to derive the model posterior analytically. We consider different correlation lengths to evaluate its influence on the model posterior estimation. Each correlation length defines different dimension priors as follows:

- for $\ell = 0.1$, $k_{min} = 17$ and $k_{max} = 1\,014$; this yields $p = 6.17 \times 10^{-3}$.
- for $\ell = 0.5$, $k_{min} = 4$ and $k_{max} = 204$; this yields $p = 2.59 \times 10^{-2}$.
- for $\ell = 0.9$, $k_{min} = 2$ and $k_{max} = 114$; this yields $p = 5.12 \times 10^{-2}$;

recall that the success probability p, marking the decay of the geometric prior distribution, is computed by solving a nonlinear equation given the truncation parameters k_{min}, k_{max} (see subsection 2.4). The priors together with the analytical model posterior and evidence are shown in Figure 2. We employ these closed-form solutions as reference to test the performance of the proposed tBUS-SuS approach.

Before solving the inverse problem, we start with several parameter studies to assess the performance of tBUS-SuS. These are carried out by monitoring three posterior quantities of interest (QoIs), namely, the dimension parameter k, the flexibility random field at the middle of the beam F_{mid}, and the deflection random field at the tip of the beam w_{tip}.

Remark 4. We evaluated the performance of tBUS-SuS for different proposal scalings, this includes the parameter ρ of the jump proposal $\mathbf{Q}$ and the parameter β of the pCN proposal (the results are omitted for the sake of brevity). From these
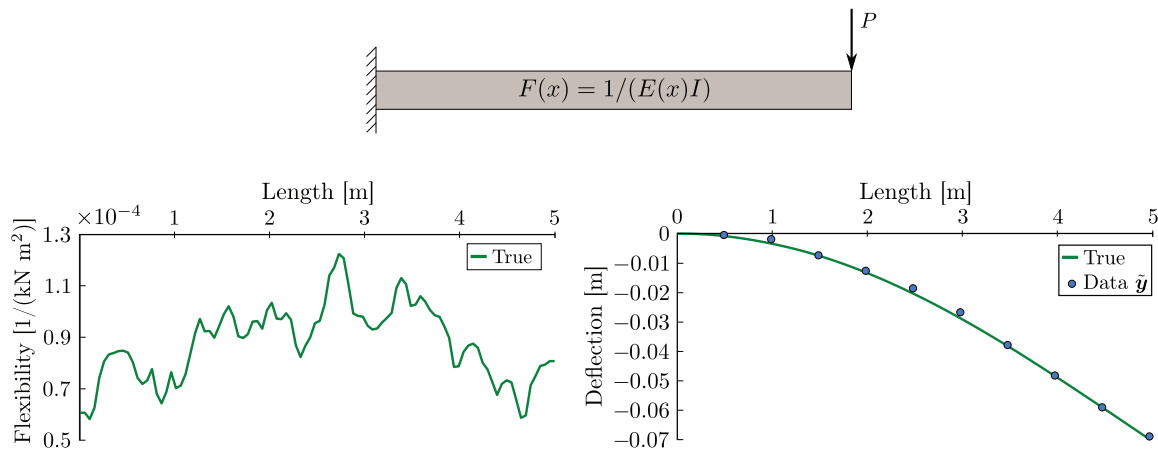
FIGURE 1 Cantilever beam problem: model description, true values and set of deflection measurements
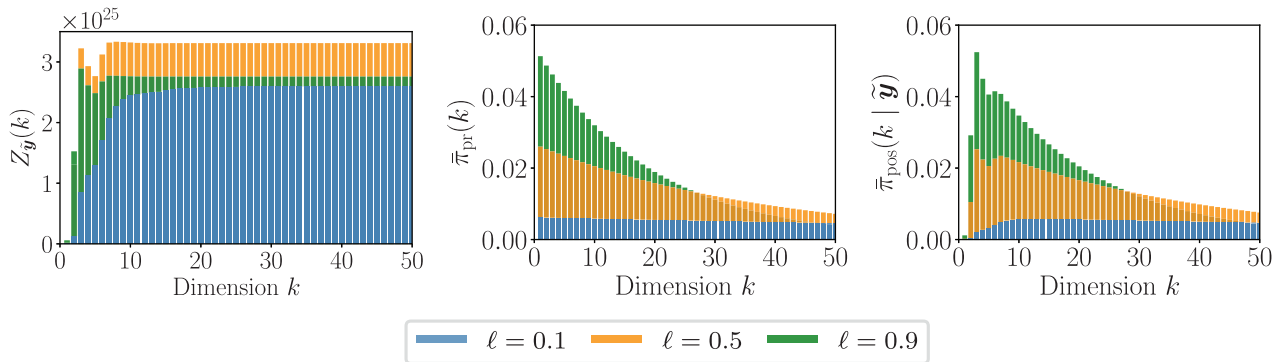
FIGURE 2 Closed-form solutions for different correlation lengths in the prior flexibility random field. Left: model evidence. Center: model/dimension prior. Right: model/dimension posterior

	Adaptive: $\text{eff}_{\text{tBUS}}(k)$		Standard: $\text{eff}_{\text{tBUS}}(k)$	
ℓ	**MwG**	**Step-wise**	**MwG**	**Step-wise**
0.1	8.01×10^{-2}	4.24×10^{-2}	4.00×10^{-2}	1.79×10^{-2}
0.5	5.49×10^{-2}	2.38×10^{-2}	6.13×10^{-3}	1.85×10^{-2}
0.9	3.59×10^{-2}	2.41×10^{-2}	4.03×10^{-3}	2.08×10^{-3}

TABLE 1 Efficiency metric (B2) of the dimension parameter k for adaptive and standard tBUS-SuS

studies we found that: (i) sampling from the prior instead of using the matrix $\mathbf{Q}$ is beneficial when the change from prior to posterior update is small (cf. Figure 2), and (ii) adapting the pCN scaling β such that a target acceptance rate value of $\alpha \in [0.2, 0.4]$ is maintained through the simulation is a good choice.

5.1.1 | Computational efficiency

We investigate the computational gain of using the across-model tBUS-SuS compared with individual within-model runs of BUS-SuS. The objective is twofold, (i) identify the number of effective independent samples in the resulting set of posterior samples, and (ii) define a metric that is equivalent in within-model and across-model simulation approaches. The efficiency metrics are defined in the appendix; they are expressed as the ratio between the effective number of independent samples and the number of models calls. In the following studies, the results are computed for an average of $N_{\text{sim}} = 100$ independent runs of the algorithms.

For each correlation length $\ell \in \{0.1, 0.5, 0.9\}$, we use the reference variances of the QoIs (computed from the closed-form solution) for the estimation of the effective number of samples in equation (B1):

- $\sigma_k^2 \in \{24\,161.22,\ 1\,250.80,\ 328.56\}$ for the dimension,
- $\sigma_{F_{\text{mid}}}^2 \in \{1.143 \times 10^{-9},\ 7.484 \times 10^{-10},\ 5.364 \times 10^{-10}\}$ for the flexibility, and
- $\sigma_{w_{\text{tip}}}^2 \in \{8.446 \times 10^{-7},\ 9.042 \times 10^{-7},\ 9.159 \times 10^{-7}\}$ for the deflection.

Standard tBUS-SuS vs. Adaptive tBUS-SuS: we compare the efficiency in the approximation of the model posterior between standard tBUS-SuS (Algorithm 1) and adaptive tBUS-SuS (Algorithm 2). We employ $N = 10^4$ samples per level. Recall that the adaptive version is a more general method since the estimation of the constant $\overline{r}$ in the standard version requires extra computational cost.

The efficiency metrics are defined in (B2) as the number of effective samples normalized by the average number of forward model calls. Small efficiency values either indicate that the samples are potentially highly correlated, or that the forward model is evaluated many times. As a result, large efficiencies hint a better sampling method. Table 1 shows overall that the efficiency of the adaptive method is larger or comparable to the efficiency obtained by the standard one.

Moreover, the extra model evaluations required to find the constant $\overline{r} = \ln(\overline{r})$ in standard tBUS-SuS are neglected in the efficiency computation. This makes the efficiency gap between both approaches even larger. In the study, $\overline{r}$ is chosen *a priori* as the maximum of 10^5 independent log-likelihood evaluations; this value is increased conservatively by 25% such that $\overline{r} \geq L_{\text{max,all}}$. This of course leads to a reduced efficiency in standard tBUS-SuS due to the increase of model evaluations. From this study, we employ the adaptive tBUS-SuS algorithm in the remainder of the article.

Within-model BUS-SuS runs vs. tBUS-SuS: we compare the efficiencies in the estimation of the random field QoIs (namely, the mean values of F_{mid} and w_{tip}). These are computed using within-model runs of adaptive BUS-SuS and the proposed adaptive tBUS-SuS. The number of samples per level used in BUS-SuS and tBUS-SuS are $N = 5 \times 10^3$ and $N = 10^4$, respectively. Figure 3 shows the efficiencies (B2) estimated with individual adaptive BUS-SuS runs (1st column) and adaptive tBUS-SuS using the two trans-dimensional MCMC algorithms (2nd and 3rd columns). We point out that in fixed-dimensional BUS-SuS, the number of model calls increases with the dimension since more intermediate levels are required to reach the posterior. Conversely, the cost in across-model tBUS-SuS is a single value for all the dimensions, and thus it needs to be distributed approximately according to the model posterior (cf. first row of Figure 3).

Note in the first row of Figure 3 that the total number of calls in within-model BUS-SuS is larger than tBUS-SuS, even when using a larger number of samples per level in tBUS-SuS; also tBUS-SuS with MwG almost doubles the cost compared to tBUS-SuS with the step-wise sampler (due to the extra likelihood evaluation required in MwG). In the second and third
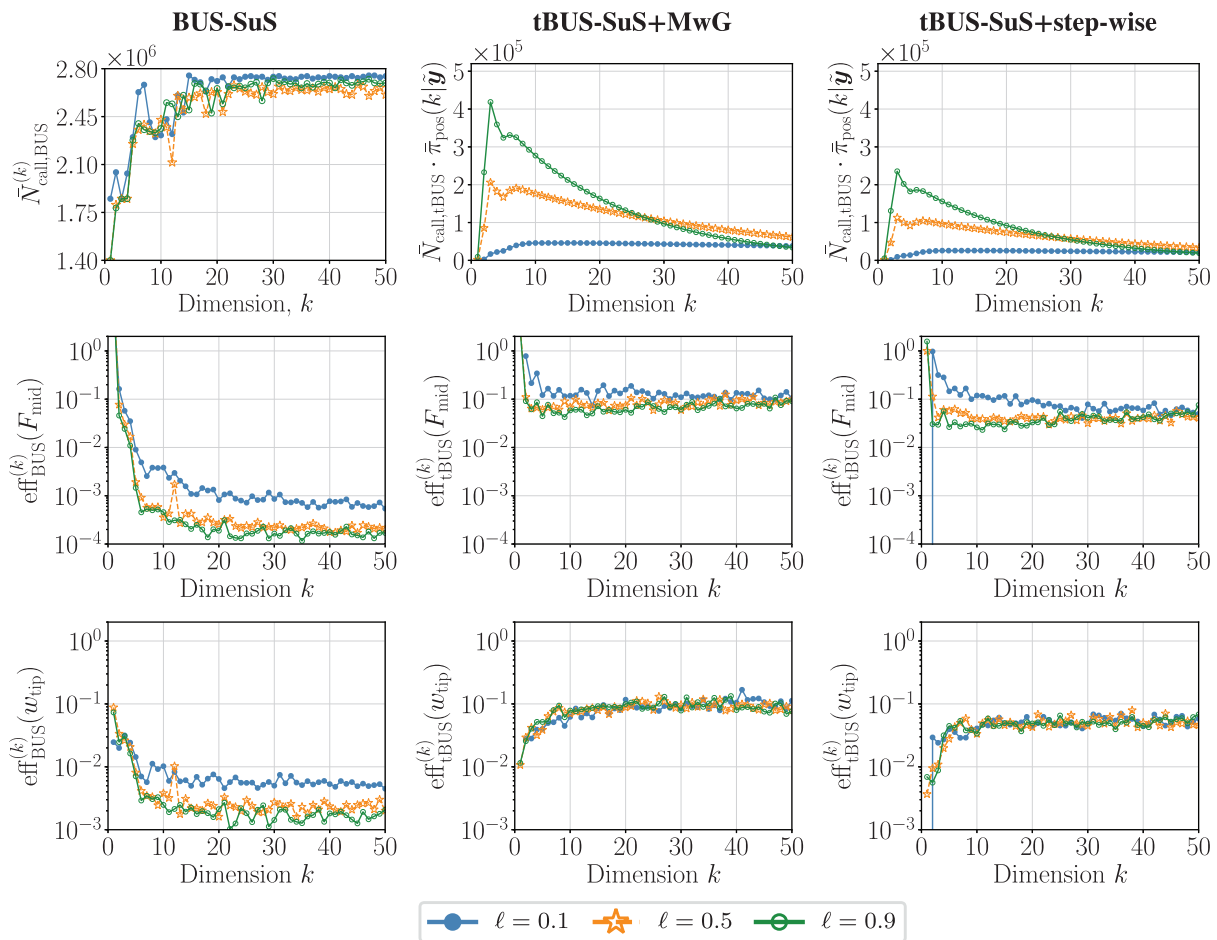
FIGURE 3 Comparison between within-model BUS-SuS and across-model tBUS-SuS: number of model calls and efficiency metrics (B2) for different correlation lengths and random field QoIs $\mu_{F_{\text{mid}}}$ and $\mu_{w_{\text{tip}}}$. Adaptive BUS-SuS (1st col), adaptive tBUS with MwG sampler (2nd col), and adaptive tBUS with step-wise sampler (3rd col)

rows of Figure 3, the efficiencies of the random field QoIs are shown per dimension, since different KL truncation orders yield different random field approximations. The number of effective samples obtained with MwG is larger than using the step-wise method. Nevertheless, the efficiencies in both approaches are similar. This is because the computational cost normalizes the number of effective number of samples in the efficiency metric (B2), and MwG has almost twice the cost of the step-wise sampler. Finally, the efficiencies also reveal the advantage of using across-model simulation algorithms for the solution of Bayesian model choice problems, as compared to single model runs.

Number of samples per level: the influence of the number of sampler per level has been assessed for the adaptive BUS-SuS method in Reference 44. We perform a similar study for adaptive tBUS-SuS, but focusing on the estimation of the dimension posterior statistics. For simplicity, we select MwG as the MCMC sampler, and fix the correlation length to $\ell = 0.5$.

Since analytical solutions are available, we compute a normalized root-mean-square error (RMSE) of the approximation of the model posterior, as N increases. This metric is estimated pointwise for each dimension as, $\text{RMSE}^{\star}(k) = \sqrt{\mathbb{E}[(\hat{\bar{\pi}}_{\text{pos}}(k \mid \tilde{\mathbf{y}}) - \bar{\pi}_{\text{pos}}(k \mid \tilde{\mathbf{y}}))^2]}/\bar{\pi}_{\text{pos}}(k \mid \tilde{\mathbf{y}})$, where $\hat{\bar{\pi}}_{\text{pos}}(k)$ denotes the approximated posterior. To obtain a global metric for each sample size, the pointwise errors are averaged through all dimensions. The results are shown in Figure 4. We also plot the absolute bias and standard deviation components of the RMSE. Note that the standard deviation tends to dominate the RMSE. As expected, the approximation of the dimension posterior improves for increasing N.

We remark that the accuracy in the approximation of the quantities of interest in tBUS-SuS depends mostly on the correlation of the resulting MCMC chains. This correlation tends to increase with the number of levels in tBUS-SuS (a property inherited from the SuS method). Since this value is always problem specific, a recommendation is to increase N when the number of intermediate levels is large.

5.1.2 | Approximation of the posterior for the dimension and the random fields

We utilize the adaptive version of tBUS-SuS for the estimation of the model and random field posteriors. We employ $N = 10^4$ samples per level. The approximated dimension posteriors are shown in Figure 5 using the MwG and step-wise samplers. In this case, we plot the mean and standard deviation bounds of the approximation, computed as an average of $N_{sim} = 100$ independent runs. The shape of the reference model posterior is well-captured for all investigated correlation length cases. The variability of the approximation using the MwG sampler is smaller than the one computed by the step-wise algorithm. Note that the differences in both samplers are larger for smaller correlation lengths. The tBUS-SuS simulations require in average $N_{lv} = 5$ intermediate levels to reach the posterior for all the investigated correlation lengths.

We also estimate the posterior flexibility and deflection random fields for different correlation lengths. Closed-form expressions of the posterior random fields are used as Reference 29. Figure 6 shows the model choice and model mixing solutions in terms of the posterior mean and posterior 95% credible intervals (CI); this CI is defined as the region between the 0.025 and 0.975 quantiles of the posterior. For the deflection response field, we compute the difference between the prior mean and the 95% posterior CIs (called *differential* deflection), in order to differentiate the approximations. The model choice estimate is given by the truncation order that yields the maximum model posterior (Figure 5), in this case $k_{best} \in \{10, 3, 3\}$ for the correlation lengths $\ell \in \{0.1, 0.5, 0.9\}$, respectively. The model mixing estimate takes into account the whole dimension spectrum (up to k_{max}). The reference CIs agree closely with the model mixing estimates since we
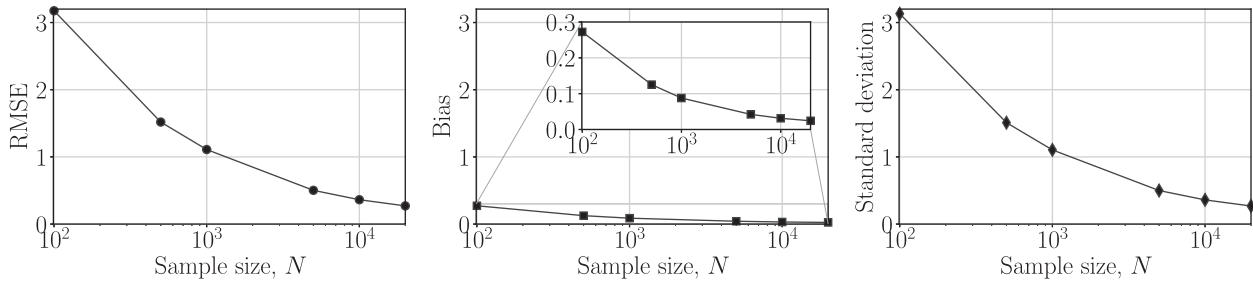
FIGURE 4 Influence of the number of samples per level on tBUS-SuS. Left: RMSE. Center: absolute bias. Right: standard deviation. These quantities are normalized by the exact solution and then averaged through dimensions to obtain a point estimate for each sample size
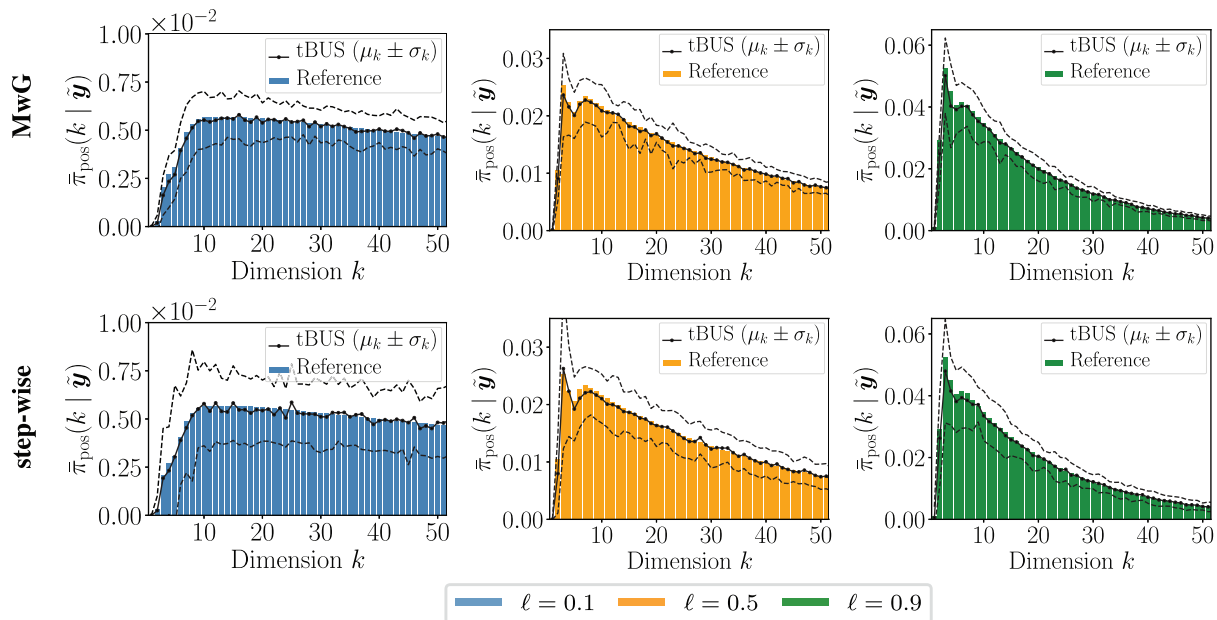
FIGURE 5 Estimation of the model posterior using adaptive tBUS-SuS for different correlation lengths in the prior flexibility random field: MwG (1st row); step-wise (2nd row)
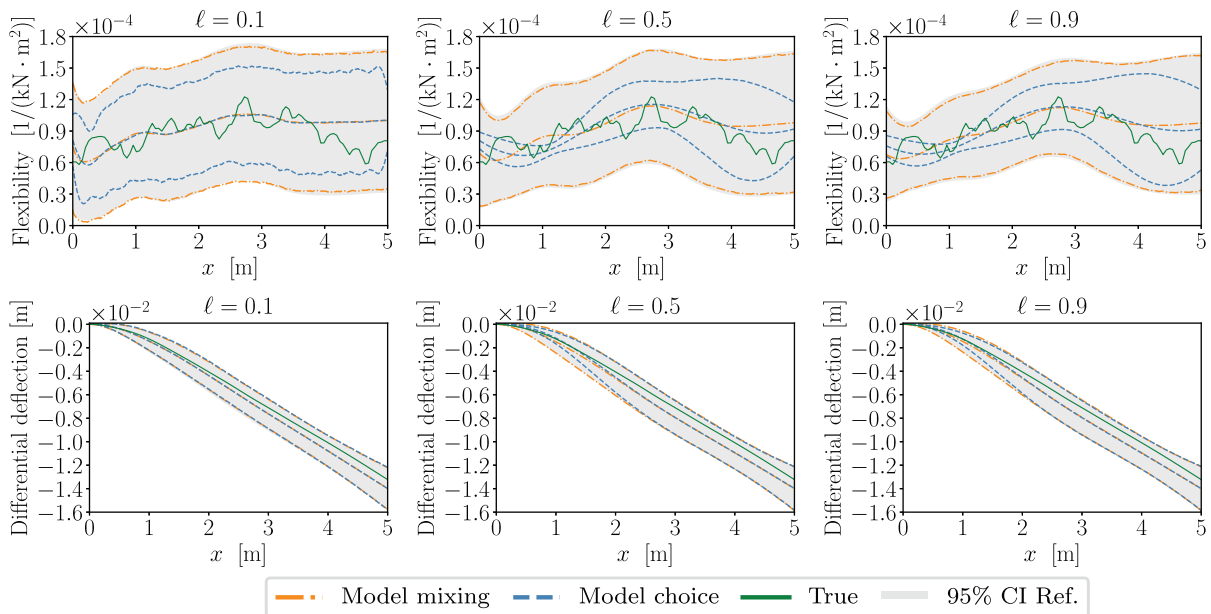
FIGURE 6 Posterior flexibility and deflection random fields for different correlation lengths in the prior flexibility field: estimated mean and 95% CI of the best model (model choice) and the averaging of models (model mixing) using adaptive tBUS-SuS with MwG. The reference 95% CI is highlighted in gray

use all the KL expansions associated to the model posterior for the random field representation. In this case, for the larger correlation length, the model choice solution fails to capture the assumed true flexibility in different intervals of the domain.

Remark 5. To expand further on the results of Figure 6, the same study of this subsection was performed with a smaller prior coefficient of variation. In this case, we use the correlation length $\ell = 0.5$ and the prior standard deviation is chosen as $\sigma_{pr} = 0.1\mu_{pr}$. The main effects of reducing the input coefficient of variation are as follows: (i). The influence of the lower truncation models is decreased; in this case the best model is $k_{best} = 7$; (ii). The 95% CI of the flexibility field becomes tighter, with the model mixing and model choice solutions producing similar results; (iii). The performance of adaptive tBUS-SuS is unaffected and the efficiency metric (B2) remains in the same order of magnitude as in Table 1.

We conclude this subsection by illustrating the evolution of the samples in tBUS-SuS and its relation to the model posterior. The results are shown for the prior correlation length $\ell = 0.5$. Figure 7 shows the prior, second intermediate level, and posterior samples obtained from a single simulation of adaptive tBUS-SuS with MwG. The process of sequentially approximating the posterior is shown by the distribution of the samples, starting from the prior and narrowing down to the target posterior. For $k = 1$, we plot the samples that contribute to the model posterior at the first dimension, that is, the one-dimensional KL coefficient against the auxiliary standard uniform random variable. The tBUS-SuS simulation required $N_{lv} = 5$ levels to reach the posterior region (highlighted in gray), $\tau = [14.74, 5.04, 2.67, 0.31, 0]$. Note that the value of the model posterior at $k = 1$ is almost zero (cf. Figure 5), and hence the amount of samples is considerably reduced as the algorithm evolves from the prior to the posterior measure. Moreover, the maximum log-likelihood at dimension $k_{max} = 204$ is $\bar{c}_{204} = 68.17$, and at dimension $k = 1$ it is $\bar{c}_1 = 56.53$. Due to the nested structure of the KL expansion, the constant $\bar{r} = \ln(\bar{r})$ in the LSF (16) is equal to the maximum log-likelihood at the largest dimension. In this case, the scaling $\bar{r}$ is significantly larger than the value of the covering constant at dimension 1. Thus, we observe that the posterior samples at $k = 1$ are located in a small region of the two-dimensional parameter space. This occurs mainly at lower dimensions, since there exist significant differences between lower- and higher-dimensional likelihood values. We remark that there is an associated reduction of the efficiency, but this does not prevent the algorithm from computing accurate posterior samples. Moreover, with increasing k the values of $\bar{c}_k$ are closer to $\bar{r}$ and the efficiency loss becomes negligible. For $k = 2$, Figure 7 plots the components of the two-dimensional KL coefficients; we also show the contours of the log-likelihood function with fixed dimension $k = 2$. In this case, the reduction in the number of samples when updating from prior to posterior is smaller than at dimension $k = 1$. Note that the value of the model posterior at $k = 2$ is larger than zero, and the difference in the probability mass between prior and posterior at $k = 2$ is less substantial (cf. Figure 5).
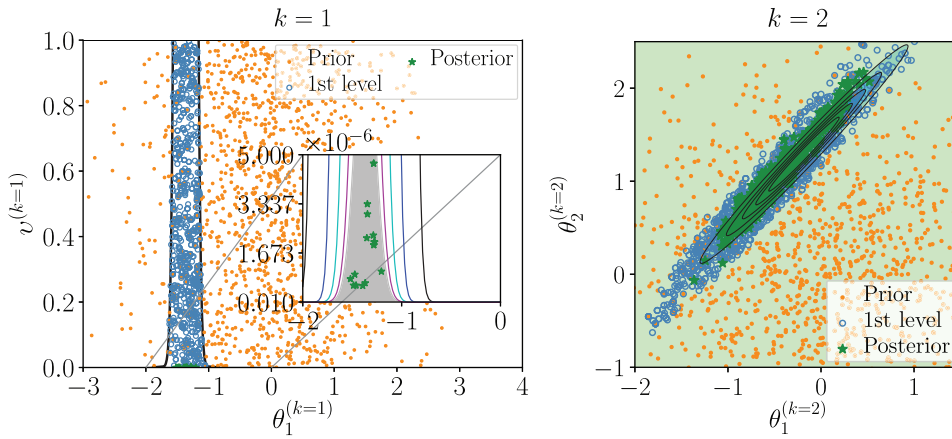
FIGURE 7 tBUS-SuS samples of the KL coefficients at dimensions $k = 1$ (left) and $k = 2$ (right). For $k = 1$, the posterior region is highlighted in gray. For $k = 2$, the contours of the log-likelihood function are also plotted

5.2 | 2D groundwater flow

We consider inference of the hydraulic conductivity field of an aquifer using observations of the hydraulic head measured at specific boreholes (see, e.g., Reference 51). We define an aquifer on the square domain $D = [0,1] \times [0,1]$ km^2 with boundary ∂D. Spatial coordinates are denoted by $\boldsymbol{x} = [x_1, x_2] \in D$. The steady-state Fick's second law of diffusion is used to describe the spatial variation of the hydraulic head inside the aquifer. Hence, for a given hydraulic conductivity of the soil $\kappa(\boldsymbol{x}, \omega)$ and sink or source terms $J(\boldsymbol{x})$, the hydraulic head $u(\boldsymbol{x})$ follows the elliptic PDE

$$-\nabla \cdot [\kappa(\boldsymbol{x}, \omega) \nabla u(\boldsymbol{x})] = J(\boldsymbol{x}), \tag{26}$$

with Dirichlet boundary condition, $u(\boldsymbol{x}) = 0$ for $\boldsymbol{x} \in \partial D$. The source terms are defined as the superposition of nine weighted Gaussian plumes with standard width $\sigma_J = 1 \times 10^{-3}$ km. The plumes have equal and unitary strengths, and are centered at locations $\boldsymbol{\mu}_J = [0.25\,i, 0.25\,j]$ with $i, j = 1, \dots, 3$, that is

$$J(\boldsymbol{x}) = \sum_{i=1}^{9} \mathcal{N}(\boldsymbol{x};\ \boldsymbol{\mu}_J^{(i)}, \sigma_J^2 \mathbf{I}_2). \tag{27}$$

We employ the finite element method to solve the PDE (26) using piecewise linear triangular elements on a regular mesh of 80×80 square blocks (each divided into two triangles). Hence, the total number of elements is $2 \times 80^2 = 1.28 \times 10^4$. This selection is based on parameter studies showing that this mesh is sufficiently fine to account for the random field variability associated to the smallest correlation length used in our experiments. The prior hydraulic conductivity field is modeled as a log-normal random field, $\kappa(\boldsymbol{x}) := \exp(\kappa'(\boldsymbol{x}))$. The underlying Gaussian field $\kappa'(\boldsymbol{x})$ has mean zero and standard deviation $\sigma_{\kappa'} = 1$ km/day. The covariance operator of the Gaussian field is constructed from a Matérn kernel,[35] with smoothing parameter $\nu = 1.5$. The solution of the eigenvalue problem is computed with the Nyström method using 100 Gauss–Legendre points in each direction.[40] Bear in mind that the KL eigenvalue problem is solved only once for $k_{\max}$ terms. This means that this task is not re-computed for each k since we work on a saturated space where the KL coefficients have a nested structure.

The *true* conductivity $\kappa(\boldsymbol{x})$ is a realization of a random field with characteristics similar to the prior. In this case, we set the Matérn kernel parameters as $\nu_{\text{true}} = 2.0$ and $\ell_{\text{true}} = 0.1$. The truncation of the KL expansion used to generate this realization is 312, which captures 99% of the prior variance. The hydraulic head observations $\widetilde{\boldsymbol{y}}$ are obtained at $m = 12$ sensor locations. They are computed from a PDE evaluation of the true conductivity field using a finer finite element mesh. The measurement error is modeled as additive and mutually independent from the random field. It is defined by a joint Gaussian PDF with mean zero and noise covariance matrix $\boldsymbol{\Sigma}_{\text{obs}} = \sigma_{\text{obs}}^2 \mathbf{I}_m$. The variance of the measurement noise is prescribed such that the observations have a signal-to-noise ratio $\mathbb{V}\left[\widetilde{\boldsymbol{y}}\right] / \sigma_{\text{obs}}^2 = 120$. The true hydraulic conductivity and hydraulic head fields, together with the source terms and the synthetic data are shown in Figure 8.

In this example, we evaluate the posterior for different correlation lengths $\ell \in \{0.1, 0.2, 0.3\}$. Each correlation length defines a dimension prior as follows:
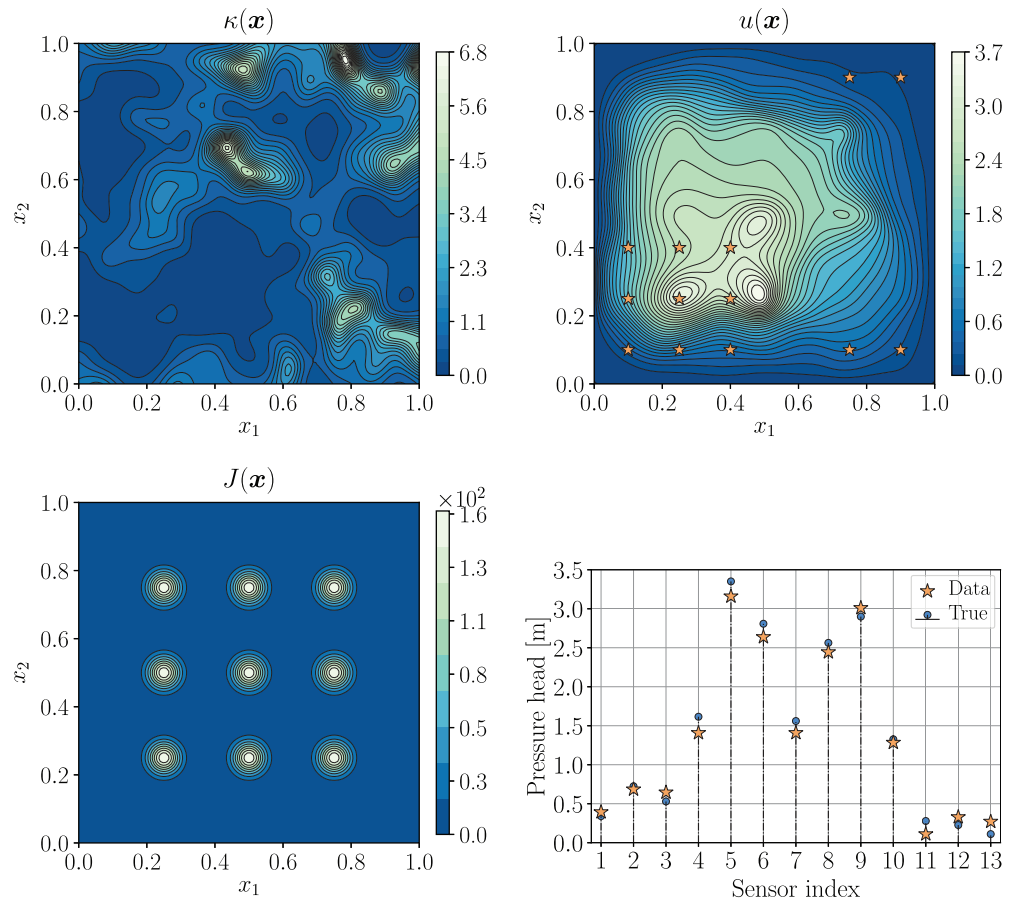
- for $\ell = 0.1$, we obtain $k_{\min} = 16$ and $k_{\max} = 520$; this yields $p = 6.31 \times 10^{-3}$.
- for $\ell = 0.2$, we obtain $k_{\min} = 5$ and $k_{\max} = 138$; this yields $p = 1.94 \times 10^{-2}$.
- for $\ell = 0.3$, we obtain $k_{\min} = 3$ and $k_{\max} = 65$; this yields $p = 2.94 \times 10^{-2}$.

We employ the same proposal scaling settings investigated in the previous example. Moreover, we use a proposal $\mathbf{Q}$ to sample k, in addition to sample from the prior. The jump proposal matrix is constructed from a discrete triangular distribution with jump length $\rho = 0.25\, k_{\max}$, which appears to be a good choice for both MCMC algorithms (cf. Remark 4).

5.2.1 | Approximation of the posterior for the dimension and the random fields

Adaptive tBUS-SuS is used to estimate the posterior of the dimension and the random field. The results are shown for an average of $N_{\text{sim}} = 60$ independent simulation runs using $N = 1.5 \times 10^4$ samples per level. To compare the tBUS-SuS approximations, we compute the reference solution from model evidences estimated by within-model runs of adaptive BUS-SuS using $N = 5 \times 10^3$ samples per level and averaged over $N_{\text{sim}} = 90$ simulations.

For this example, it is not feasible to compute the full reference solution for the posterior of the dimension by means of within-model simulation algorithms. Thus, we estimate the reference at 6 dimension snapshots for each correlation length: $k_{\text{snap}} \in \{40, 45, 50, 55, 60, 70\}$ for $\ell = 0.1$, $k_{\text{snap}} \in \{20, 30, 35, 40, 50, 60\}$ for $\ell = 0.2$, and $k_{\text{snap}} \in \{20, 25, 30, 32, 35, 40\}$ for $\ell = 0.3$. Since the reference solutions are given in terms of the model evidences $Z_{\widetilde{\mathbf{y}}}(k_{\text{snap}})$, we transform them to model posteriors using (6). This requires the knowledge of the evidence of all model classes, which is not available in this case. Instead, we apply the normalization

$$\overline{\pi}_{\text{pos}}(k_{\text{snap}} \mid \widetilde{\mathbf{y}}) = \frac{\overline{\pi}_{\text{pr}}(k_{\text{snap}}) Z_{\widetilde{\mathbf{y}}}(k_{\text{snap}})}{\sum_{k \in k_{\text{snap}}} \overline{\pi}_{\text{pr}}(k) Z_{\widetilde{\mathbf{y}}}(k)} \sum_{k \in k_{\text{snap}}} \hat{\overline{\pi}}_{\text{pos}}(k \mid \widetilde{\mathbf{y}}), \tag{28}$$

such that the sum of the reference $\overline{\pi}_{\text{pos}}(k_{\text{snap}} \mid \widetilde{y})$ match the sum of the tBUS-SuS estimated model posteriors $\hat{\overline{\pi}}_{\text{pos}}(k_{\text{snap}} \mid \widetilde{y})$ at the given snapshots. By using this approach, the reference solution is limited to the tBUS-SuS solution. Nevertheless, it allows us to validate the correct shape of the dimension posterior.

The dimension posteriors estimated by adaptive tBUS-SuS with MwG are shown in Figure 9. We plot the mean and standard deviation bounds of the approximations. The solutions are computed when the dimension is sampled from the prior (1st row) and when it is sampled from the proposal **Q** (2nd row). Both alternatives yield comparable results. In general, we observe an increase in the variability around the MAP estimate. Note also that the dimension posteriors have several modes which can be related to the non-uniform distribution of the measurement locations, together with the symmetry of the KL eigenfunctions. For instance, when using $\ell = 0.3$ there is a jump in the values of the probability mass after the 15th dimension, every 5 dimensions until the MAP estimate. Furthermore, as an indicative of the algorithm performance, tBUS-SuS requires on average $N_{\text{lv}} = 12$ intermediate levels and the proposal scaling β changes from 0.75 in the first level, to 0.09 in the last level (for the correlation length $\ell = 0.1$ and sampling k from the prior).

The approximated dimension posteriors using adaptive tBUS-SuS with the step-wise sampler are shown in Figure 10. In this example, the step-wise sampler is more sensitive to the selection of the dimension proposal than MwG. We observed that the dimension prior is not a good proposal choice to sample the dimensions (the results are omitted). The main issue is that the resulting posterior samples are highly correlated since the values of the scaling β at the last level of the simulation are in the order of 10^{-4}. Therefore, instead of showing a comparison between the dimension proposal schemes, we employ the proposal matrix **Q** with two different settings: using N samples per level, and using $2N$ samples per level. In both MCMC algorithms, increasing the number of samples per level considerably improves the variability of the estimates. Particularly, using $2N$ samples per level in the step-wise sampler yields comparable results to those of MwG since we are evaluating the PDE model approximately the same number of times. The resulting dimension posteriors are able to capture the trend of the reference solutions. In both cases, the estimation is very close to the reference mean value.

Finally, we estimate the posterior hydraulic conductivity and hydraulic head random fields for the investigated correlation lengths. Figure 11 shows the model choice and model mixing solutions in terms of the posterior mean and standard deviation of the hydraulic conductivity (only for $\ell \in \{0.1, 0.3\}$). The model choice estimate is given by the truncation order that yields the maximum model posterior, in this case $k_{\text{best}} \in \{43, 43, 33\}$ for each investigated correlation length (Figure 10, 2nd row). The model mixing solution takes into account the whole dimension spectrum. Note that most of
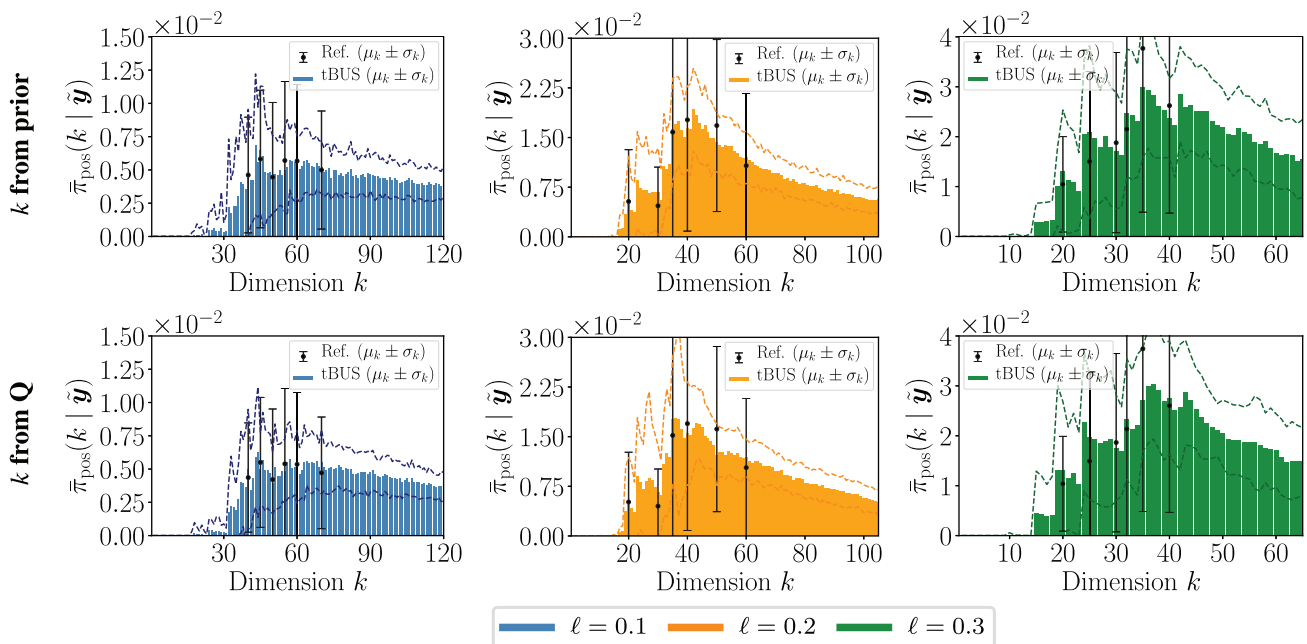

FIGURE 9 Diffusion example: model posterior using adaptive tBUS-SuS with **MwG sampler**, sampling k from the prior (1st row) and from proposal **Q** (2nd row)
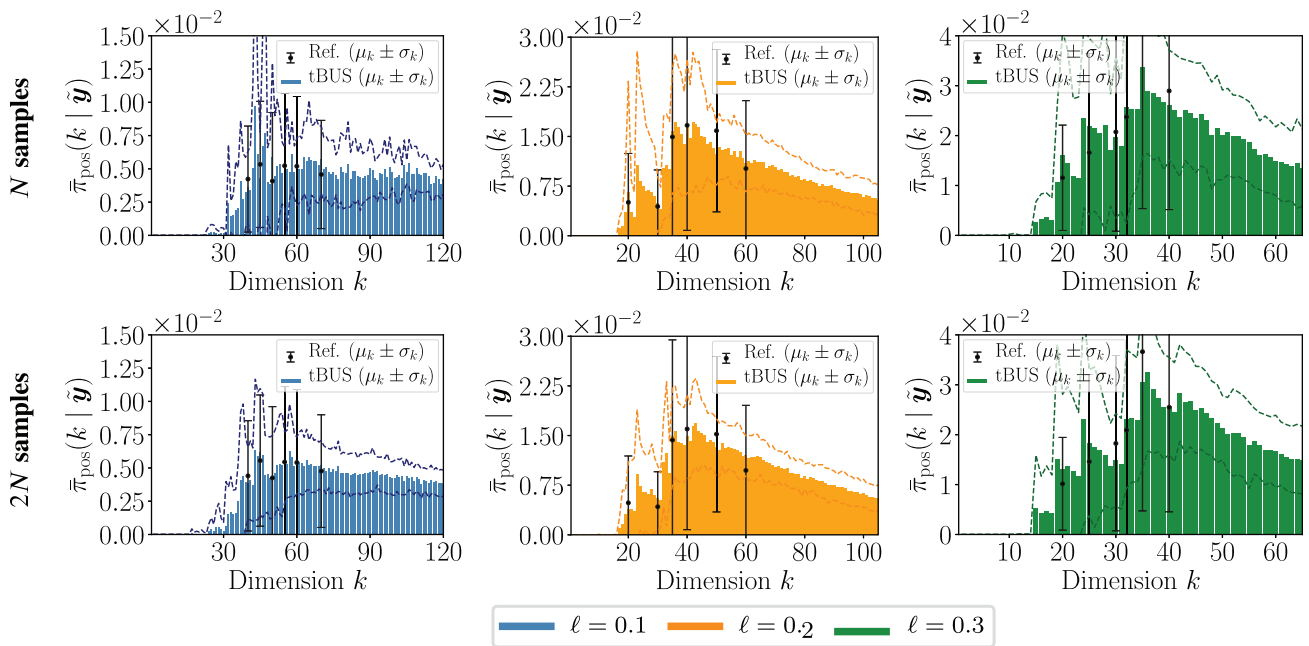
FIGURE 10 Diffusion example: model posterior using adaptive tBUS-SuS with **step-wise sampler** and sampling k from proposal **Q**. Using N samples per level (1st row), using $2N$ samples per level (2nd row)
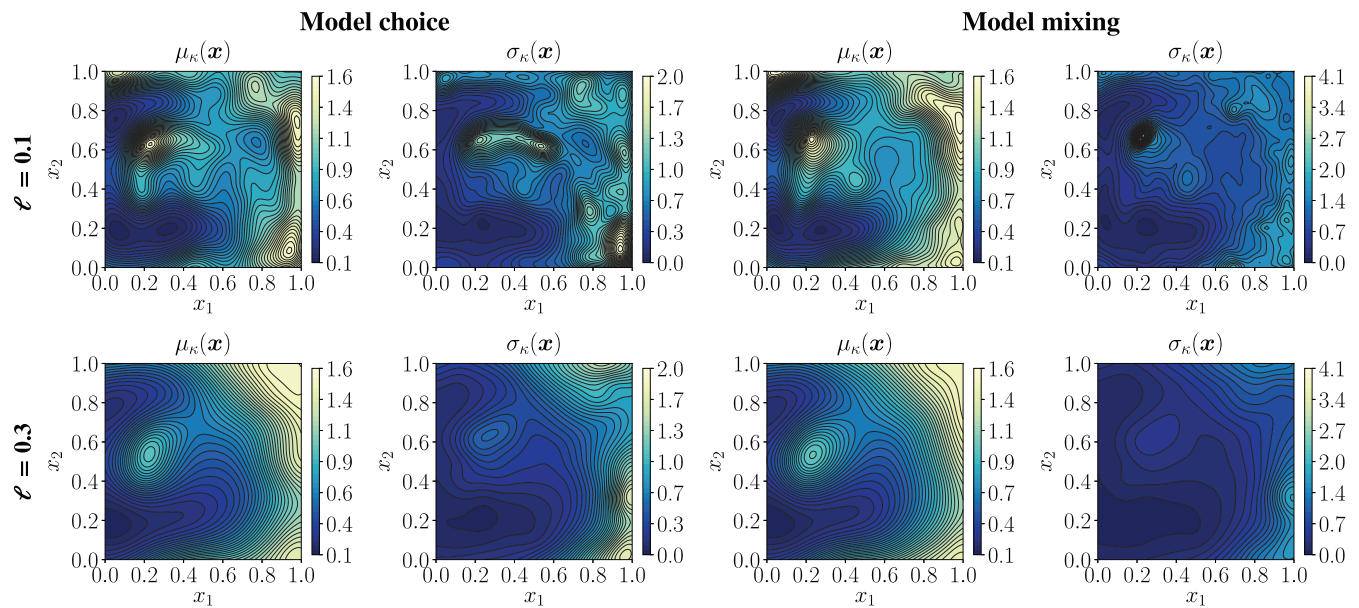

FIGURE 11 Diffusion example: posterior mean (1st and 3rd cols) and standard deviation (2nd and 4th cols) of the **hydraulic conductivity** random field using adaptive tBUS-SuS with the step-wise sampler for different prior correlation lengths (rows) and employing model choice or model mixing

the measurements are concentrated in the lower left corner of the aquifer; also at this location the true hydraulic conductivity is small. Despite the values of the posterior mean are smaller than those of the assumed truth, we observe that the statistics are revealing the locations of lower and higher conductivity values. Moreover, the random field modeled with the smaller correlation length is able to represent the small fluctuations better than those with larger correlation lengths. In contrast to the conductivity field, the differences in the model choice and mixing solutions for the hydraulic head random field are small (Figure 12). This quantity is computed by integrating the PDE model which can be seen as an averaging operation that reduces the effect of the spatial variability, similar to Example 1.
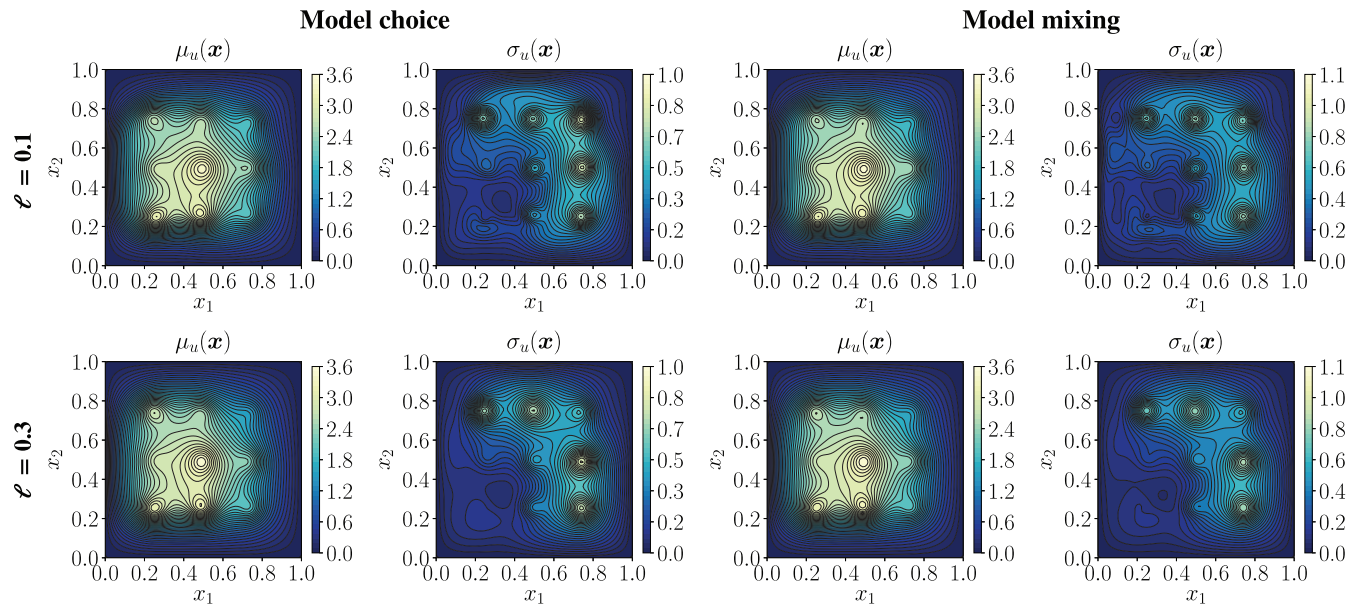
FIGURE 12 Diffusion example: posterior mean (1st and 3rd cols) and standard deviation (2nd and 4th cols) of the **hydrostatic pressure** random field using adaptive tBUS-SuS with the step-wise sampler for different prior correlation lengths (rows) and employing model choice or model mixing

6 | DISCUSSION OF RESULTS

The proposed method is an extension of the BUS formulation for the solution of Bayesian inverse problems where the dimension of the parameter space is variable. Such type of inferences are common in random field updating tasks, since the optimal number of terms in the random field series expansion is unknown a priori, and hence, it can be modeled probabilistically.

The main findings of this contribution in terms of the random field modeling are: (i) If one employs a uniform prior for the dimension, the values of the model evidence define the model posterior themselves. In this case, visits to models with high dimension have the same probability of occurrence, for example, in the beam example for the correlation length $\ell = 0.5$, a model with 10 terms in the KL expansion is evaluated as many times as the model with 50 terms, despite the fact that the quality of both models is essentially the same (Figure 2). Thus, the inclusion of a prior that penalizes models with increased number of parameters is beneficial in the context of the KL expansion. We use the proposed dimension prior to achieve a trade-off between the dimensions with large model evidence and the computational cost, in order to reduce the evaluation of unnecessary high-dimensional models. (ii) Our examples show that the model choice solution is not guaranteed to produce the best results. It can be seen as a 'point' estimate of the dimension posterior, which in some cases can fail to represent the whole variability of the underlying random field. In these cases, using the model mixing solution (or a subset of it) is recommended. However, when the prior coefficient of variation is small, the model mixing and model choice solutions produce comparable results (cf. Remark 5). (iii) One could argue that performing model choice or mixing is not fundamentally necessary in random field applications, and instead one can fix the number of KL terms to a large value (e.g., $k_{\max}$) and perform a standard within-model Bayesian inversion. One of the advantages of our method is the reduction of the computational cost. Note that the value of $k_{\max}$ can be in principle very large, whereas a smaller value (or a set of smaller values) could be sufficient for an appropriate posterior field representation. This is of course relevant in modeling scenarios where the PDE model is expensive or that require gradient computations. (iv) Non-Gaussian random fields problems can be addressed by our approach as long as they can be represented by the KL expansion (or any other spectral expansion in which the parameters have a nested structure). This includes, for instance, the so-called translation fields[52] which apply nonlinear transformations of Gaussian ones.

The main findings in terms of the computational aspects of the method are: (i) In order to populate the discrete-continuous parameter space, the algorithm requires large number of samples per level N to achieve an accurate solution (cf. Figure 4). This is especially relevant if high-order moments of response quantities are to be estimated. As shown in Example 2, the variability of the simulations results reduces with larger N. However, this cost is small when

compared to standard within-model simulation approaches, standalone trans-dimensional MCMC algorithms, or other sequential algorithms used to solve Bayesian model choice problems.[18,19] (ii) The proposed step-wise sampler is a simplification of the RJMCMC sampler in the saturated or composite space that is based on the pCN algorithm. The combination of the step-wise sampler with tBUS-SuS produces a method that is more 'automatic' in its definition, since the tuning of a jumping function is not required. In more general model choice situations (other than nested models), the inclusion of a jumping function is necessary and beneficial, not only to satisfy the detailed balance condition, but also to improve the efficiency of the algorithm. This is relevant in inference cases where the parameters associated with the models have a completely different physical meaning; for example, if the models correspond to different PDEs. These settings are beyond the scope of this article and can be investigated in a future study. (iii) The choice of the dimension proposal within the trans-dimensional MCMC algorithms plays an important role in the estimation of the posterior for the KL dimension. In general, we observe in both examples that the step-wise sampler is more sensitive to this tuning step. Particularly, Example 1 shows that if the discrepancy between model prior and posterior is minimal, drawing samples from the prior gives the best results. However, using a proposal matrix to control the dimension moves is beneficial when the change in the prior-to-posterior update is significant, as seen in Example 2. From the examples we see that using a jump length between 10% and 50% of the maximum truncation order yields good results. (iv) The variability in the dimension posterior estimate is larger when employing the step-wise sampler. Recall that MwG evaluates the joint posterior in two levels: first, one fixes the parameters and samples the dimension, and second, one fixes the dimension and samples the parameters. This mechanism requires twice the number of PDE model evaluations as compared to the step-wise algorithm, which only uses a single joint acceptance step. If one doubles the number of samples per level in the step-wise algorithm such that one obtains a similar cost to MwG, both algorithms yield comparable results, as shown in Example 2.

We remark that one possible improvement of the method is to use the coverings $\bar{c}_k|_{k=1}^{k_{max}}$ for each particular dimension, instead of choosing the constant $\bar{r} = \max(\bar{c}_k|_{k=1}^{k_{max}})$. As discussed at the end of subsection 5.1.2, the constant $\bar{r}$ can be considerably larger than the values of the maximum log-likelihood at the lower dimensions, which induces an efficiency loss in those dimensions. The algorithm can be designed such that one keeps track of the dimensions that are being evaluated and define the LSF (16) as a function of the individual $\bar{c}_k$. This will also require the definition of multiple intermediate levels since every dimension variable defines a particular posterior domain.

7 | CONCLUSIONS

We introduce a sequential methodology for the solution of Bayesian model choice problems, whereby the models are dimensions of the parameter space. The method (called tBUS-SuS) is a variable-dimensional extension of the BUS formulation when used in combination with subset simulation. The approach is applicable to nested models and variable selection models arising in random fields represented by series expansions whose number of terms is unknown. The main ideas behind tBUS-SuS are the re-definition of the parameter space to account for the discrete dimension random variable, and the application of trans-dimensional MCMC algorithms to sample the resulting sequence of intermediate discrete-continuous distributions.

For choosing the prior distribution of the model, we define a dimension prior that penalizes increasing model complexity. This prior is tailored to random field models parameterized with the Karhunen–Loève expansion. Different heuristics are defined in order to adapt the proposed prior to any particular problem. Two MCMC algorithms that work in a saturated space where the dimension is fixed are examined. Particularly, we discuss a Metropolis-within-Gibbs sampler and derive a step-wise sampler based on a simplification of the reversible jump MCMC algorithm in the saturated space. The results in the settings of our experiments show that under the same computational cost, the step-wise sampler has a comparable performance to the Metropolis-within-Gibbs algorithm. After combining the MCMC algorithms with tBUS-SuS, both samplers yield similar performance in terms of efficiency and accuracy in the estimation of the dimension posterior. Nevertheless, the step-wise algorithm is recommended in cases where the PDE model evaluation is expensive, since only one acceptance/rejection step is required for dimension and parameters.

Our applications involve spatially varying parameters modeled as random fields and discretized with the Karhunen–Loève expansion. The results compared to reference solutions show that the tBUS-SuS algorithm with the two MCMC variants perform well in high- and trans-dimensional nested parameter spaces, that is, it is able to estimate not only the posterior random fields but also the posterior distribution of the dimension in the discretization. Finally, the experiments indicate that in random field problems it is recommended to employ the model mixing solution instead of the single model choice solution; the latter can fail to represent the spatial variability of the underlying random field. This

behavior occurs in particular when the prior coefficient of variation is large. Further analysis of this numerical observation is suggested in future studies.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Felipe Uribe https://orcid.org/0000-0002-1010-8184

REFERENCES
1. Jeffreys H. *Theory of Probability*. 3rd ed. Oxford, UK: Oxford University Press; 2003.
2. Robert CP. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. 2nd ed. New York, NY: Springer; 2007.
3. Beck JL, Katafygiotis LS. Updating models and their uncertainties. I: Bayesian statistical framework. *J Eng Mech*. 1998;124(4):455-461.
4. Tarantola A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); 2005.
5. Beck JL, Yuen KV. Model selection using response measurements: Bayesian probabilistic approach. *J Eng Mech*. 2004;130(2):192-203.
6. Li X, Tsai FT-C. Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod. *Water Resour Res*. 2009;45(9):1-14.
7. Prudencio EE, Bauman PT, Faghihi D, Ravi-Chandar K, Oden JT. A computational framework for dynamic data-driven material damage control, based on Bayesian inference and model selection. *Int J Numer Methods Eng*. 2015;102:379-403.
8. Gelman A, Meng XL. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Stat Sci*. 1998;13(2):163-185.
9. Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*. 1995;82(4):711-732.
10. Green PJ. Trans-dimensional Markov chain Monte Carlo. In: Green PJ, Hjort NL, Richardson S, eds. *Highly Structured Stochastic Systems*. Oxford, UK: Oxford University Press; 2003:179-198.
11. Carlin BP, Chib S. Bayesian model choice via Markov chain Monte Carlo methods. *J Royal Stat Soc Ser B*. 1995;57(3):473-484.
12. Stephens M. Bayesian analysis of mixture models with an unknown number of components–an alternative to reversible jump methods. *Ann Stat*. 2000;28(1):40-74.
13. Besag J. *Markov Chain Monte Carlo for Statistical Inference*. Washington, DC: Center for Statistics and the Social Sciences, University of Washington; 2001.
14. Godsill SJ. On the relationship between MCMC methods for model uncertainty. *J Comput Graph Stat*. 2001;10(2):230-248.
15. Fan Y, Sisson SA. Reversible jump Markov chain Monte Carlo. In: Brooks S, Gelman A, Jones GL, Meng XL, eds. *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: Chapman & Hall/CRC Press; 2011:67-91.
16. Owen AB. Monte Carlo theory, methods and examples; 2013. statweb.stanford.edu/\%simowen/mc/.
17. Rubinstein RY, Kroese DP. *Simulation and the Monte Carlo Method*. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2017.
18. Andrieu C, De Freitas NA, Doucet A. Sequential MCMC for Bayesian model selection. Paper presented at: Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics. (SPW-HOS '99); 1999:130-134.
19. Jasra A, Stephens DA, Holmes CC. Population-based reversible jump Markov chain Monte Carlo. *Biometrika*. 2007;94(4):787-807.
20. Richardson S, Green PJ. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J R Stat Soc Ser B Stat Methodol*. 1997;59(4):731-792.
21. Koutsourelakis PS. A novel Bayesian strategy for the identification of spatially varying material properties and model validation: an application to static elastography. *Int J Numer Methods Eng*. 2012;91:249-268.
22. Loève M. Fonctions aléatoires du second ordre. *La Revue Scientifique*. 1946;84:195-206.
23. Karhunen K. Über lineare methoden in der wahrscheinlichkeitsrechnung. *Ann Acad Sci Fennicæ Ser A.1 Math Phys*. 1947;37:1-79.
24. Li CC, Der Kiureghian A. Optimal discretization of random fields. *J Eng Mech*. 1993;119(6):1136-1154.
25. Kamiński M. *The Stochastic Perturbation Method for Computational Mechanics*. Hoboken, NJ: Wiley; 2013.
26. Shinozuka M, Deodatis G. Simulation of multi-dimensional Gaussian stochastic fields by spectral representation. *Appl Mech Rev*. 1996;49(1):29-53.
27. Ghanem RG, Spanos PD. *Stochastic Finite Elements: A Spectral Approach*. Revised ed. Mineola, New York: Dover Publications; 2012.
28. Elsheikh AH, Wheeler MF, Hoteit I. Hybrid nested sampling algorithm for Bayesian model selection applied to inverse subsurface flow problems. *J Comput Phys*. 2014;258:319-337.
29. Uribe F, Papaioannou I, Betz W, Straub D. Bayesian inference of random fields represented with the Karhunen-Loève expansion. *Comput Methods Appl Mech Eng*. 2020;358:112632.

30. Straub D, Papaioannou I. Bayesian updating with structural reliability methods. *J Eng Mech*. 2015;141(3):04014134.

31. Au SK, Beck JL. Estimation of small failure probabilities in high dimensions by subset simulation. *Probab Eng Mech*. 2001;16(4):263-277.

32. Straub D, Papaioannou I, Betz W. Bayesian analysis of rare events. *J Comput Phys*. 2016;314:538-556.

33. Cotter S, Roberts G, Stuart A, White A. MCMC methods for functions: modifying old algorithms to make them faster. *Stat Sci*. 2013;28(3):424-446.

34. Simpson D, Rue H, Martins TG, Riebler A, Sørbye S. Penalising model component complexity: a principled, practical approach to constructing priors. *Stat Sci*. 2017;32(1):1-28.

35. Abrahamsen P. *A Review of Gaussian Random Fields and Correlation Functions*. 2nd ed. Norway: Norwegian Computing Center; 1997.

36. Adler RJ. *The Geometry of Random Fields*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); 2010.

37. Kolmogorov AN, Fomin SV. *Introductory Real Analysis*. Mineola, New York: Dover Publications; 1975.

38. Alexanderian A. A brief note on the Karhunen-Loève expansion; 2015:1-14. arXiv:1509.07526v2 eprint.

39. Betz W, Papaioannou I, Straub D. Numerical methods for the discretization of random fields by means of the Karhunen-Loève expansion. *Comput Methods Appl Mech Eng*. 2014;271:109-129.

40. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in C: The Art of Scientific Computing*. 3rd ed. Cambridge, MA: Cambridge University Press; 2007.

41. Hastie DI, Green PJ. Model choice using reversible jump Markov chain Monte Carlo. *Stat Neerl*. 2012;66(3):309-338.

42. Fuglstad GA, Simpson D, Lindgren F, Rue H. Constructing priors that penalize the complexity of Gaussian random fields. *J Am Stat Assoc*. 2019;114(525):445-452.

43. Del Moral P, Doucet A, Jasra A. Sequential Monte Carlo samplers. *J R Stat Soc Ser B*. 2006;68(3):411-436.

44. Betz W, Papaioannou I, Beck JL, Straub D. Bayesian inference with subset simulation: strategies and improvements. *Comput Methods Appl Mech Eng*. 2018;331:72-93.

45. Papaioannou I, Betz W, Zwirglmaier K, Straub D. MCMC algorithms for subset simulation. *Probab Eng Mech*. 2015;41:89-103.

46. DiazDelaO F, Garbuno-Inigo A, Au S, Yoshida I. Bayesian updating and model class selection with subset simulation. *Comput Methods Appl Mech Eng*. 2017;317:1102-1121.

47. Brooks SP, Giudici P, Roberts GO. Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *J R Stat Soc Ser B Stat Methodol*. 2003;65(1):3-55.

48. Fan Y, Peters Q, Sisson S. Automating and evaluating reversible jump MCMC proposal distributions. *Stat Comput*. 2009;19:409-421.

49. Müller P. *A Generic Approach to Posterior Integration and Gibbs Sampling. Technical Report 91–09*. West Lafayette, IN: Department of Statistics, Purdue University; 1991.

50. Chen V, Dunlop MM, Papaspiliopoulos O, Stuart AM. Dimension-robust MCMC in Bayesian inverse problems; 2019:1-29. arXiv:1803.03344v2 eprint.

51. Cui T, Marzouk YM, Willcox KE. Data-driven model reduction for the Bayesian solution of inverse problems. *Int J Numer Methods Eng*. 2015;102(5):966-990.

52. Grigoriu MD. *Stochastic Systems: Uncertainty Quantification and Propagation*. New York, NY: Springer; 2012.

53. Gibbs AL, Su FE. On choosing and bounding probability metrics. *Int Stat Rev*. 2002;70(3):419-435.

54. Gelbrich M. On a formula for the L2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Math Nachr*. 1990;147(1):185-203.

APPENDIX A. PENALIZED COMPLEXITY PRIOR

The prior for the dimension in (7) can be seen as a penalized complexity (PC) prior[34] in a discrete variable setting. Let us consider a *base* prior model $\pi_{\text{pr}}(\theta \mid k = 1)$ and a *flexible* prior model $\pi_{\text{pr}}(\theta \mid k)$, where $k > 1$ is a flexibility parameter that accounts here for the dimension. A PC prior is defined by applying the following propositions:[34]

(i) Based on Occam's razor principle, a simpler model is preferred until there is enough information to support for a more complex model. The simplest model is defined in terms of the base prior $\pi_{\text{pr}}(\theta \mid k = 1)$.

(ii) A measure of increasing complexity is defined as the distance between the flexible and base models, $d(k) = K(d_{\text{KL}}(\pi_{\text{pr}}(\theta \mid k)||\pi_{\text{pr}}(\theta \mid k = 1)))$, where $d_{\text{KL}}(\cdot||\cdot)$ is the Kullback–Leibler divergence (KLD)[53] from the base to the flexible model and K is an increasing function, with $K(0) = 0$.

(iii) A constant rate penalization is applied to deviations from the base model. This implies an exponential prior distribution on the distance scale, $\overline{\pi}_{pr}(d(k)) = a \, \exp(-a \, d(k))$. The prior of the flexibility parameter k is derived from $\overline{\pi}_{pr}(d(k))$ as, $\overline{\pi}_{pr}(k) \propto a \, \exp(-a \, d(k)) \, |\partial d(k)/\partial k|$, where the last term corresponds to the determinant of the Jacobian of the transformation from d to k.

(iv) The user defines the scaling of the prior by controlling its mass at the tail; a is selected by considering $\mathbb{P}\left[k \leq k_{\min}\right] = \alpha$.

We apply such propositions in our variable-dimensional setting. Let $k = 1, \ldots, k_{\max}$ and $\pi_{pr}(\boldsymbol{\theta} \mid k) := \mathcal{N}(\mathbf{0}, \mathbf{I}_{1:k})$ be a (possibly degenerate) zero mean Gaussian distribution on $\mathbb{R}^{k_{\max}}$, with covariance

$$\mathbf{I}_{1:k} := \operatorname{diag}(\underbrace{1, \ldots, 1}_{k}, \overbrace{0, \ldots, 0}^{k_{\max}-k}). \tag{A1}$$

We define the base model as the Gaussian distribution $\pi_{pr}(\boldsymbol{\theta} \mid k = 1)$. Note that $\pi_{pr}(\boldsymbol{\theta} \mid k)$ has no Lebesgue density on $\mathbb{R}^{k_{\max}}$, for $k < k_{\max}$. It is possible to generalize the KLD to allow to compute distances between $\pi_{pr}(\boldsymbol{\theta} \mid k)$ and $\pi_{pr}(\boldsymbol{\theta} \mid k')$, for $k, k' = 1, \ldots, k_{\max}, k' \leq k$. However, in this case we obtain

$$d_{KL}\left(\pi_{pr}(\boldsymbol{\theta} \mid k) || \pi_{pr}(\boldsymbol{\theta} \mid k')\right) \tag{A2}$$

$$= \int_{\Theta} \ln(\mathcal{N}(\boldsymbol{\theta}; \mathbf{0}, \mathbf{I}_{1:k})) - \ln(\mathcal{N}(\boldsymbol{\theta}; \mathbf{0}, \mathbf{I}_{1:k'})) \, d\mathcal{N}(\mathbf{0}, \mathbf{I}_{1:k})(\boldsymbol{\theta}) \tag{A3}$$

$$= \int_{\Theta} \ln(\mathcal{N}(\boldsymbol{\theta}; \mathbf{0}, \mathbf{I}_{1:k})) - \ln\left(\left\{\begin{matrix} \prod_{i=1}^{k'} \mathcal{N}(\theta_i; 0, 1), & \text{if } \theta_{k'+1:k_{\max}} = 0 \\ 0, & \text{otherwise} \end{matrix}\right\}\right) d\mathcal{N}(\mathbf{0}, \mathbf{I}_{1:k})(\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } k' = k, \\ \infty, & \text{otherwise.} \end{cases} \tag{A4}$$

Hence, we cannot use the KLD as a metric to construct a PC prior in this particular setting. Instead, we use the Wasserstein-2 distance. Let $f, g : \mathbb{R}^{k_{\max}} \to \mathbb{R}$ be two probability density functions. Moreover, let

$$\operatorname{Cop}(f, g) := \left\{ h : \mathbb{R}^{2k_{\max}} \to \mathbb{R} : \int h(\cdot, \boldsymbol{\theta}^{(2)}) d\boldsymbol{\theta}^{(2)} \equiv f, \int h(\boldsymbol{\theta}^{(1)}, \cdot) d\boldsymbol{\theta}^{(1)} \equiv g \right\} \tag{A5}$$

be the set of couplings between f and g. That is, the set of joint probability density functions on $\mathbb{R}^{2k_{\max}}$ that have f and g as marginals. The *Wasserstein-2 distance* between f and g is given by

$$d_{Was(2)}(f || g) := \left(\inf_{h \in \operatorname{Cop}(f,g)} \iint \|\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}\|^2 h(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}) d\boldsymbol{\theta}^{(1)} d\boldsymbol{\theta}^{(2)} \right)^{1/2}. \tag{A6}$$

Standard results about the Wasserstein-2 distance between two Gaussian measures (the base and all other models) indicate[54]

$$d_{Was(2)}(\pi_{pr}(\boldsymbol{\theta} \mid k) || \pi_{pr}(\boldsymbol{\theta} \mid 1)) = \|\mathbf{I}_{1:1} - \mathbf{I}_{1:k}\|_F^2 = k - 1; \tag{A7}$$

by choosing $K(x) := x$ in proposition (ii), the distance becomes $d(k) = k - 1$. Using this information, we can now construct the PC prior. Proposition (iii) induces a prior of the form

$$\overline{\pi}_{pr}(k) \propto \exp(-a \cdot d(k)) = \exp(-a(k - 1)), \quad \text{for some } a > 0. \tag{A8}$$

We obtain the truncated geometric prior (up to a normalizing constant) proposed in (7) by setting $a := \ln(1/(1 - p))$. The parameter p can be chosen via prior information as specified in proposition (iv).

APPENDIX B. EFFICIENCY METRICS

Consider a collection of samples of a random variable Q representing a monitored QoI, $\{Q_i^{(j)}, j = 1, \ldots, N_{\text{sim}}, i = 1, \ldots, N\}$, where N_{sim} is the number of independent simulation runs. Note that each element $Q^{(j)}$ contains N samples of the QoI. The MCMC steps within tBUS-SuS induce some correlation in the resulting posterior samples. We estimate the effective number of independent samples of Q in the set of $(N \cdot N_{\text{sim}})$ samples as[16]

$$N_{\text{eff}}(Q) = N_{\text{sim}} \frac{\mathbb{V}[Q]}{\widehat{\mathbb{V}}[\hat{\mu}_{Q^{(j)}}]}, \tag{B1}$$

where $\mathbb{V}[Q]$ is the true variance of the QoI, $\hat{\mu}_{Q^{(j)}}$ denotes the sample mean of the QoI obtained at the jth simulation, and $\widehat{\mathbb{V}}[\cdot]$ denotes the sample variance.

In addition to N_{eff}, we also take into account the computational cost. The number of model calls in BUS-SuS, is equal to $N_{\text{call}}^{(j)} = N + (N - N_s)(N_{\text{lv}}^{(j)} - 1)$, where $N_s = p_0 N$ is the number of seeds used in the MCMC steps, and $N_{\text{lv}}^{(j)}$ is the number of levels obtained at the jth simulation. The total number of calls is then given by $\overline{N}_{\text{call,BUS}}^{(k)} = \sum_{j=1}^{N_{\text{sim}}} N_{\text{call}}^{(j)}$. These values are obtained per dimension since the N_{sim} runs of BUS-SuS are performed at each specific $k = 1, \ldots, k_{\max}$. In the case of tBUS-SuS, the number of model calls depends on the trans-dimensional MCMC algorithm. If one employs the step-wise sampler, the number of model calls per simulation run is $N_{\text{call,a}}^{(j)} = N + (N - N_s)(N_{\text{lv}}^{(j)} - 1)$. If one utilizes the MwG algorithm, the number of model calls is $N_{\text{call,b}}^{(j)} = N + 2(N - N_s)(N_{\text{lv}}^{(j)} - 1)$ since the likelihood function is evaluated twice at the MCMC step. Hence, the total number of calls using the step-wise sampler is $\overline{N}_{\text{call,tBUS}} = \sum_{j=1}^{N_{\text{sim}}} N_{\text{call,a}}^{(j)}$, and with MwG is $\overline{N}_{\text{call,tBUS}} = \sum_{j=1}^{N_{\text{sim}}} N_{\text{call,b}}^{(j)}$.

As a result, we define the efficiency metrics for within-model BUS-SuS and across-model tBUS-SuS as follows:

$$\text{eff}_{\text{BUS}}(Q) = \frac{N_{\text{eff,BUS}}^{(k)}(Q)}{\overline{N}_{\text{call,BUS}}^{(k)}} \quad \text{and} \quad \text{eff}_{\text{tBUS}}(Q) = \frac{N_{\text{eff,tBUS}}^{(k)}(Q)}{\overline{N}_{\text{call,tBUS}}\hat{\bar{\pi}}_{\text{pos}}(k \mid \tilde{\mathbf{y}})}, \tag{B2}$$

where $N_{\text{eff,BUS}}^{(k)}(Q)$ and $N_{\text{eff,tBUS}}^{(k)}(Q)$ are the effective number of independent samples of the QoI in (B1), obtained at each dimension k and from each method. Note that since the cost of tBUS-SuS is associated to the whole dimension spectrum and not to each individual k, the value is approximately distributed among dimensions by multiplying it with the estimated dimension posterior $\hat{\bar{\pi}}_{\text{pos}}(k \mid \tilde{\mathbf{y}})$.