Bayesian inference with subset simulation in spaces of varying dimension

Felipe Uribe\textsuperscript{a,*}, Iason Papaioannou\textsuperscript{a}, Jonas Latz\textsuperscript{b}, Wolfgang Betz\textsuperscript{a}, Elisabeth Ullmann\textsuperscript{b}, Daniel Straub\textsuperscript{a}

\textsuperscript{a}Engineering Risk Analysis Group, Technische Universität München. Arcisstraße 21, 80333 München, Germany
\textsuperscript{b}Chair of Numerical Analysis, Technische Universität München. Boltzmannstraße 3, 85748 Garching b.M., Germany

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, are based on functional representations that use series expansions 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 paper, we develop a sequential algorithm to handle such inference settings, and propose a prior distribution for the dimension parameter that penalizes increasing dimensionality. The method is a variable-dimensional extension of the BUS (Bayesian Updating with Structural reliability methods) approach, 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 parameter space. The performance of the proposed method with both MCMC variants is assessed numerically for two example problems: A 1D cantilever beam with spatially varying flexibility, and a 2D groundwater flow problem with uncertain hydraulic conductivity field.

Keywords: inverse problems, Bayesian model choice, trans-dimensional MCMC, subset simulation, random fields, Karhunen-Loève expansion.

1. Introduction

Numerical approximations to partial differential equations (PDEs) used in engineering and science 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 considering observations or measurements of the physical system. In the general case, this represents an inverse problem in which the objective is to identify the model parameters that are compatible with the available information. A more general type of inference arises in model choice situations, whereby the model itself is selected from a predefined collection of models. Each model in this set can have different parameters or can be based on different mathematical assumptions. The uncertainty in the model and its parameters can be treated in a unified manner within the Bayesian inference framework [1, 2, 3].

In the Bayesian approach to solve inverse problems, all uncertain parameters are modeled by 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

\*Corresponding author.

Email addresses: felipe.uribe@tum.de (Felipe Uribe), iason.papaioannou@tum.de (Iason Papaioannou), jonas.latz@tum.de (Jonas Latz), wolfgang.betz@tum.de (Wolfgang Betz), elisabeth.ullmann@ma.tum.de (Elisabeth Ullmann), straub@tum.de (Daniel Straub)

Preprint submitted to Journal of Computational Physics June 17, 2019
to estimating or characterizing this updated (posterior) distribution. In 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 one prior for the parameters conditioned on the model and a second prior for the model itself [4, 2].

Closed-form expressions of the posterior distribution are not available in most cases and approximate solutions are computed. One common approach in model choice situations is to use methods that are based on the likelihood function of individual models, including Akaike information, Bayesian information, Hannan-Quinn information (some applications are provided in [5, 6]). Another flexible way to approach the model choice problem is via sampling methods, such as Markov chain Monte Carlo (MCMC) [7]. In this case, the characterization of the posterior requires the exploration of a discrete-continuous space. This task can be performed by running several independent runs of an importance-sampling-type method for each particular model; an approach referred to as within-model simulation. For expensive PDE models this is computationally prohibitive. Alternatively, simultaneous inference on both, model and parameters, can be carried out using MCMC algorithms that are able to explore the disjoint union space by moving between different models; this approach is called across-model simulation. The standard algorithm to perform this type of inferences is the reversible jump MCMC algorithm [8, 9], which explores the parameter space using proposals that allow jumps between different models (see [10, 11] for an overview). Alternative MCMC approaches include: Carlin and Chib [12] adapt the Gibbs sampler to deal with model choice problems. Phillips and Smith [13] present an iterative jump-diffusion algorithm to perform model comparison; the approach uses two steps to construct a Markov process that generates samples from the joint posterior distribution, one jump component that makes discrete transitions between models, and a diffusion component that draws parameter samples between those jumps. Stephens [14] expresses the collection of models as a marked point process, with each point representing a model. Afterwards, an MCMC scheme implements a birth-death step to determine the stationary distribution of the process. Besag [15] modifies the reversible jump MCMC such that the parameter dimensions are all the same in a so-called product or saturated space defined over all possible models and parameters. In the same direction, Godsill [16] shows that the reversible jump MCMC can be derived as a classical Metropolis-Hastings algorithm applied to the product space.

All these methodologies can also be applied to cases with a single model that has parameters with variable dimension (in this context, they are also referred to as trans-dimensional MCMC algorithms [9]). Some examples include mixture models with an unknown number of components [17], polynomial regression where the degree of the polynomial is variable [18], object recognition where the image being analyzed is composed by an unknown number of shapes [13], 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 random fields. Random fields oftentimes increase the complexity of the inverse problem since the posterior distribution is defined over an infinite-dimensional space. Series representations are typically applied in order to project the random field to a finite-dimensional space. For instance, the Karhunen-Loève (K-L) [19, 20] expansion discretizes the field using the eigenvalues and eigenfunctions of its autocovariance operator to construct a series expansion with random coefficients [21]. It is common practice to truncate the K-L 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 [22].

Most of the disadvantages of standard MCMC samplers, such as convergence rate deterioration with increasing dimensions and burn-in/thinning periods requirements, are also present in the MCMC samplers used for Bayesian inference in varying dimensions [11]. In fixed-dimensional cases, specialized sequential algorithms that gradually approach the posterior distribution alleviate some of these issues [23, 18]. Some of these algorithms have been adapted to variable-dimensional settings, these include the sequential MCMC approach proposed in [24], which combines sequential importance sampling with reversible jump MCMC; and the population-based reversible jump MCMC presented in [25], which is able to estimate high- and trans-dimensional posterior distributions. Moreover, a sequential but within-model simulation scheme to perform model selection is proposed in [26]; the approach is based on a hybrid nested sampling algorithm and it is applied to the estimation of model evidences associated to different truncation orders in the K-L
Another algorithm used in fixed-dimensional problems is BUS (Bayesian updating with structural reliability methods), which expresses a Bayesian inverse problem as an equivalent rare event simulation task [27]. For an efficient and sequential solution of the inverse problem BUS is typically combined with subset simulation (SuS) [28] (this approach is termed BUS-SuS), although other rare event estimation methods can be employed (see e.g. [29]). In this paper, we propose an extension of the classical BUS methodology that is able to perform inference in parameter spaces of different dimension. The idea is to exploit the sequential structure of the BUS-SuS approach to develop an efficient algorithm for Bayesian model choice problems. The structure of the parameter space defined in BUS is modified by including the discrete dimension parameter. As a result, the sequence of intermediate levels is now associated to trans-dimensional distributions and the standard MCMC algorithms within BUS-SuS are no longer valid. Therefore, we investigate a class of trans-dimensional and dimension-independent MCMC algorithms that explore a saturated or composite parameter space in an alternating manner. Particularly, we discuss a Metropolis-within-Gibbs algorithm and present a step-wise sampler as a simplified reversible jump MCMC algorithm in the saturated space. In this space, the dimension is fixed to a maximum upper value, which is selected conservatively based on prior information. The core of these algorithms is given by the preconditioned Crank-Nicolson sampler [30] since the resulting saturated space is typically high-dimensional. Several parametric studies are carried out in order to find near-optimal tuning steps for the proposed trans-dimensional BUS-SuS algorithm. The efficiency and accuracy of the method is tested on random field models represented with the K-L expansion with different correlation lengths: 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 address the specification of the model/dimension prior by defining a discrete distribution that penalizes increasing dimensionality of the parameter space. In model choice problems, imposing a penalty to complicated models is necessary (see [1, 5, 31] for a discussion). A model with more parameters usually fits the data better than a model with less parameters. However, the actual improvement might be negligible and possible over-fitting of the data can arise [5, 26, 22]. The dimension prior is defined based on the geometry of the parameter space and prior information about the random fields. We also show that this prior has connections to penalized complexity priors [31].

The organization of the paper is as follows: In Section 2, we present fundamental concepts of random field modeling and the K-L 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. In Section 3, we present an overview of trans-dimensional MCMC algorithms from the viewpoint of the reversible jump MCMC and saturated space approaches. Thereafter, the two MCMC samplers used by our method are described and illustrated with a short example. The major contribution of this work is introduced in Section 4, where we explain the trans-dimensional BUS methodology, first from the general BUS framework and then from the combined BUS-SuS approach. Next, the proposed method is demonstrated by means of two numerical experiments in Section 5. The paper 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 $(\Omega, \mathcal{F}, P)$ be a probability space and $D \subseteq \mathbb{R}^d$ an index set representing a physical domain, and $L^2(\Omega, P)$ the Hilbert space of second-order random variables. A real-valued random field is a function $H(x, \omega) : D \times \Omega \to \mathbb{R}$, with arguments $x \in D$ a spatial coordinate and $\omega \in \Omega$ an outcome of the sample space [32, 33]. 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 [19, 20], 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 [32, 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 [34, p.248]. One can use such basis to represent a second-order random field as

\[ H(x, \omega) \approx \hat{H}(x; k) := \mu(x) + \sum_{i=1}^{\infty} 1(i \leq k) \sqrt{\lambda_i} \phi_i(x) \theta_i(\omega), \]  

(1)

where \( \hat{H}(x; k, \theta(\omega)) \) is the approximated field, \( 1(\cdot) \) denotes the indicator function, \( k \) is the truncation order of the expansion, \( \theta(\omega) : \Omega \to \mathbb{R} \) 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 \( \phi_i(x) : D \to \mathbb{R} \) are the eigenfunctions of the covariance operator, with \( \phi_i(x) \in L^2(D) \). For Gaussian random fields, the variables \( \theta(\omega) \) are independent standard Gaussian. In the general case, the distribution of \( \theta(\omega) \) is cumbersome to estimate. The series expansion in (1) follows from Mercer’s Theorem (details are provided in [35]) and it is referred to as the Karhunen-Loève (K-L) 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 [21], which can be solved using different approaches, including projection methods (collocation, Galerkin) [36], Nyström methods [37], among others.

The K-L expansion is often employed to reduce the dimensionality and parameterize random fields. Consider the square-integrable random vector \( \theta(\omega) \in \Theta_k := \mathbb{R}^k \) resulting from truncating the K-L series expansion (1) at the \( k \)-th term. This truncation yields an approximate field, which is optimal in the mean-squared-error sense as compared to any other spectral projection algorithm [21]. Since the eigenpairs associated to the covariance operator are deterministic quantities, the parameter \( \theta(\omega) \) characterizes the randomness of the field. Hence, the K-L construction only depends on the vector of random coefficients \( \theta \) and the truncation order \( k \).

2.2. Bayesian inverse problems in fixed dimensions

We begin by considering the forward problem \( y = G(H(x, \omega)) \), where \( G : L^2(D) \to L^2(D) \) is a solution operator expressing the relationship between the input parameters and the model response. We are interested in models where the operator \( G \) implies the solution of a PDE that has random fields as parameters. \( G \) operates on the function space \( L^2(D) \) since both, input and output, are random field realizations of two different quantities on the physical domain \( D \). The dimensionality 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 \( \Theta_k \) (with fixed dimension \( k \)). Since \( k \) is fixed, we write the approximated field as \( \hat{H}(x; \theta) \).

In inverse problems, the aim is to infer the parameters \( \theta \in \Theta_k \) given noisy observations of the system response \( \tilde{y} \in Y := \mathbb{R}^m \), with \( m \) denoting the number of observations and \( Y \) the data space. Assuming an additive observation error, the objective is to find a set of parameters \( \theta \) such that the following relation is satisfied

\[ \tilde{y} = \mathcal{G}(\hat{H}(x; \theta)) + \eta, \]  

(2)

where \( \mathcal{G} = G \circ O : \Theta_k \to Y \) is the forward response operator, defined as the composition of the solution operator \( G : \Theta_k \to L^2(D) \) with an observation operator \( O : L^2(D) \to Y \) that maps the forward solution to the data space; and \( \eta \in \mathbb{R}^m \) is the observation noise which is typically assumed to be Gaussian distributed with mean zero and non-singular covariance matrix \( \Sigma_{\text{obs}} \in \mathbb{R}^{m \times m} \).

The inverse problem of finding the parameters \( \theta \) given the relation (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 \( \theta \) are modeled as random variables and are assumed to have an initial prior density \( f(\theta) \). The likelihood function \( L(\theta; \tilde{y}) = f(\tilde{y} \mid \theta) \) is a density on the data space and provides a link between the model and data. After including observations, the updated belief about \( \theta \) is represented by the posterior density \( f(\theta \mid \tilde{y}) \). Through
Bayes’ Theorem this conditional density is
\[
f(\theta \mid \tilde{y}) = \frac{1}{Z_\tilde{y}} f(\theta) L(\theta; \tilde{y}) \propto \exp \left( -\frac{1}{2} \left\| \Sigma_{pr}^{-1/2} (\theta - \mu_{pr}) \right\|^2 + \ln L(\theta; \tilde{y}) \right),
\]
where \( Z_\tilde{y} = \int_{\theta} f(\theta) L(\theta; \tilde{y}) \, d\theta \) is the normalizing constant of \( f(\theta \mid \tilde{y}) \), called the model evidence.

Remark (1). Since we employ the K-L expansion to represent random fields, the prior distribution is Gaussian \( \theta \sim \mathcal{N}(\mu_{pr}, \Sigma_{pr}) \), whose prior mean and covariance are given by \( \mu_{pr} = 0 \) and \( \Sigma_{pr} = I_k \) (\( I_k \in \mathbb{R}^{k \times k} \) denotes the identity matrix). This is reflected in the right-hand side of (3). The information about the second-order properties of the random field enters directly in the definition of the log-likelihood function via the forward operator.

2.3. Bayesian inverse problems in varying dimensions

Consider a more general inference setting for which the set of observed data \( \tilde{y} \) is associated not only with one, but with a finite collection of plausible models \( \{M_1, \ldots, M_k, \ldots, M_{k_{\text{max}}} \} \), where \( k \in \mathcal{K} \) is a model index, \( \mathcal{K} \) is the set of model indices under consideration, and \( k_{\text{max}} < \infty \) is a prescribed limit on the collection. The 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 \( \theta_k \in \Theta_k \) 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 \( M_k \).

Let \( f(\theta_k \mid k) \) be a first-level prior density imposed on the parameter \( \theta_k \) given the model \( k \), and \( f(k) \) a second-level discrete prior mass specified over the models \( k \). The joint posterior density over both, model and parameters, is computed based on Bayes’ Theorem as
\[
f(k, \theta_k \mid \tilde{y}) = \frac{1}{Z_{\tilde{y}}} f(k) f(\theta_k \mid k) L(k, \theta_k; \tilde{y}) \propto f(k) \exp \left( -\frac{1}{2} \left\| \Sigma_{pr,k}^{-1/2} (\theta_k - \mu_{pr,k}) \right\|^2 + \ln L(k, \theta_k; \tilde{y}) \right),
\]
wherein the evidence is given by the theorem of total probability [2]
\[
Z_{\tilde{y}} = \sum_{k' \in \mathcal{K}} f(k') Z_{\tilde{y}}(k') = \sum_{k' \in \mathcal{K}} f(k') \int_{\Theta_{k'}} f(\theta_{k'} \mid k') L(k', \theta_{k'}; \tilde{y}) \, d\theta_{k'}
\]
and \( Z_{\tilde{y}}(k) \) is the evidence of the model class \( k \). The posterior density of the models is obtained by integrating out the parameters in (4) as
\[
f(k \mid \tilde{y}) = \frac{f(k) \int_{\Theta_k} f(\theta_k \mid k) L(k, \theta_k; \tilde{y}) \, d\theta_k}{\sum_{k' \in \mathcal{K}} f(k') \int_{\Theta_{k'}} f(\theta_{k'} \mid k') L(k', \theta_{k'}; \tilde{y}) \, d\theta_{k'}} = f(k) \frac{Z_{\tilde{y}}(k)}{Z_{\tilde{y}}}.
\]

The model posterior in (6) can be used to either perform (i) model choice or selection, which requires the computation of the maximum a posteriori probability (MAP) estimator, \( k_{\text{best}} = \arg \max_{k \in \mathcal{K}} f(k \mid \tilde{y}) \), or (ii) model mixing or averaging which requires the consideration of the whole collection of parameters weighted by \( f(k \mid \tilde{y}) \). Model choice is used as indicator of model complexity, i.e., 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 avoids the underestimation of uncertainty resulting from choosing only a single model. 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 [10]. We are interested in the latter case since this corresponds to Bayesian inverse problems involving random fields represented by a series expansion whose number of terms is not fixed. In the K-L expansion the set of models is defined by
the model indicator indices \(\mathcal{K} = \{1, 2, \ldots, k, \ldots, k_{\text{max}}\}\), where each element defines a truncation order. Such truncation specifies the dimensionality of the standard Gaussian random coefficients of the K-L expansion, thus, each particular model/dimension \(k\) involves a vector of uncertain parameters \(\theta_k \in \Theta_k = \mathbb{R}^k\). The aim is to perform simultaneous inference on the discrete random variable \(k\) (dimension), and the associated random vector \(\theta_k\) (coefficients) of the K-L expansion. We note that the methods discussed here are applicable to general model updating problems whenever the variables of the different models have a nested structure (e.g. [15, 16]). In the following, we often use the terms model and dimension interchangeably.

2.4. Selection of the prior distribution for the dimension

In the K-L expansion, it has been shown that the information gained by the inclusion of additional terms becomes negligible once an optimal truncation is achieved [26, 22]. This is described by very small increments in the model evidence as the number of terms in the expansion increases. Based on this behavior, we define a prior for the dimension parameter \(k\) that penalizes increasing dimensionality. We employ a truncated geometric distribution

\[
f(k) = \frac{(1-p)^{k-p}}{1 - (1-p)^{k_{\text{max}}}} \quad k = 1, \ldots, k_{\text{max}},
\]

where \(k_{\text{max}}\) 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 at the tails, such that \(P[k \leq \tau_u] = \alpha\), where \(\tau_u\) is a prescribed upper bound, and \(\alpha\) is the probability associated to that event. Based on our experiments, \(\tau_u\) is chosen as the number of terms in the K-L expansion that retains 50\% of the variability in the prior random field, and we assign to that event a probability of \(\alpha = 0.10\). Furthermore, the truncation value \(k_{\text{max}}\) is selected as the number of terms in the expansion that retains 99\% of the variability. By doing so, 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 K-L terms.

The prior defined in (7) fits well into the concept of penalized complexity (PC) prior concept proposed in [31]. Let us consider a base prior model \(f(\theta | k = 1)\) and a flexible prior model \(f(\theta | k)\), where \(k > 1\) is a flexibility parameter that accounts here for the dimension. A PC prior is defined by applying the following propositions [31]:

(i) Based on Occam’s razor principle, a simpler model is preferred until there is enough support for a more complex model. The simplest model is defined in terms of the base prior \(f(\theta | k = 1)\).

(ii) A measure of increasing complexity is defined as the distance between the flexible and base models,

\[
d(k) = K(d_{\text{KL}}(f(\theta | k) || f(\theta | k = 1))),
\]

where \(d_{\text{KL}}(\cdot || \cdot)\) is the Kullback-Leibler divergence [38] 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,

\[
f(d(k)) = a \cdot \exp(-a \cdot d(k)).
\]

The prior of the flexibility parameter \(k\) is derived from \(f(d(k))\) as,

\[
f(k) \propto a \cdot \exp(-a \cdot d(k)) |\partial d(k)/\partial k|,
\]

where the last term corresponds to the determinant of the Jacobian of the transformation from \(d\) to \(k\).

(iv) The user is able to define the scaling of the prior by controlling its mass at the tail, such that \(a\) can be selected by considering \(P[k \leq \tau_u] = \alpha\).

In the Appendix, we present a way of connecting (7) with a PC prior in the context of the K-L expansion. We also remark that different PC priors have been derived in the context of random fields, e.g., for different values of a re-parameterized Whittle-Matérn kernel [39]; and an exponential prior with fixed decay rate for the truncation order in the K-L expansion in [30].
3. MCMC algorithms in spaces of varying dimension

In Bayesian inverse problems, MCMC-based algorithms are employed to generate samples from the target posterior distribution. 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 we use in combination with the proposed trans-dimensional BUS algorithm.

3.1. General remarks

In across-model simulation, the standard algorithm to sample from the joint posterior in (4) is the reversible jump Markov chain Monte Carlo (RJMCMC) method [8]. 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 subspace. In this case, the so-called detailed balance condition is guaranteed by a standard MCMC sampler [40]. 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 [10, 11] 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 a considerable amount of tuning step procedures, such as controlling the acceptance probabilities of the jumps and saturated space approaches [41], using additional Markov chains to improve the acceptance rate of RJMCMC [42], marginal density estimators to construct jumping functions [43], hierarchical centering [44], among others.

Another class of algorithms are the saturated (also referred to as product or composite) space approaches ([12, 15, 16, 41]). 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. The joint posterior in the saturated space is expressed as [16]

$$f(k, \theta_k | \tilde{y}) = \frac{1}{Z_{\tilde{y}}} f(k) f(\theta_k | k) f(\theta_{\sim k} | k, \theta_k) L(k, \theta_k; \tilde{y}).$$  

(8)

The additional component is the so-called linking density or “pseudo-prior” $f(\theta_{\sim k} | k, \theta_k)$, where the $\theta_{\sim k}$ denotes the parameters that are not used by the model $k$. This formulation allows us to employ standard MCMC procedures to variable-dimensional problems. Most of these techniques are based on Gibbs and Metropolis-Hastings updates, or combination of both to improve efficiency.

Note that for nested models as in the K-L expansion, the variable $k$ has the effect of switching on and off coefficients in the series. Under the expansion (1) the likelihood in (4) is independent of $\theta_i$ when $i > k$. Moreover, the prior of the parameters $\theta$ follows the standard Gaussian distribution in the saturated space and it is independent of $k$. Therefore, for this case the joint posterior in (8) can be written as

$$f(k, \theta | \tilde{y}) = \frac{1}{Z_{\tilde{y}}} f(k) f(\theta | k, \theta_k; \tilde{y}),$$  

(9)

with $f(\theta)$ denoting the parameter prior in the saturated space. In the random field context, the resulting saturated space is typically high-dimensional. In order to avoid convergence deterioration with increasing $k$, dimension-independent MCMC algorithms are applied. These samplers are based on numerical discretizations to SDEs 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, Cotter et al. [30] derived a sampling algorithm by discretizing a preconditioned overdamped Langevin dynamic SDE [45, 46] using a Crank-Nicolson scheme. The Crank-Nicolson approximation is applied on the linear part of the drift, and a tuning parameter is included in the nonlinear part to distinguish between random walk and Langevin...
proposals [47]. If the preconditioner is set equal to the prior covariance, the preconditioned Crank-Nicolson (pCN) proposal arises [30, 48]
\[ \theta^* = \sqrt{1 - \beta^2} \theta + \beta \xi, \]
where \( \theta \) is the current state, \( \theta^* \) is the proposed state, \( \beta \in (0, 1] \) is the proposal scaling, and \( \xi \sim \mathcal{N}(0, \Sigma_{\text{pr}}) \).

Given the high-dimensional nature of random fields and the structure of the K-L expansion, we focus on saturated space approaches that are based on dimension-independent MCMC algorithms, namely: the Metropolis-within-Gibbs and step-wise samplers.

3.2. Metropolis-within-Gibbs sampler

The Metropolis-within-Gibbs (MwG) algorithm [49, 50] updates the K-L coefficients \( \theta \) and the dimension \( k \) in an alternating manner. In the saturated space context, MwG explores the joint posterior (8) with a Gibbs sampling version of the Carlin and Chib [12] algorithm after including Metropolis-Hastings steps (details are provided in [16]). The algorithm can also be derived by expressing the joint posterior (9) as the product of the model posterior and the model specific parameter posterior [8]
\[ f(k, \theta \mid \tilde{y}) = f(k \mid \tilde{y})f(\theta \mid k, \tilde{y}). \]

In general Bayesian inference cases, the densities \( f(\theta \mid k) \) may differ abruptly for small changes in the variable \( k \), and thus, the chain can always remain in some state. However, under the formulation (1), the coefficients and the dimension are independent a priori. This property alleviates potential poor mixing properties in MwG [51]. The idea of MwG is to sample each conditional density in (11) by applying two different steps. In the first step, we fix the parameter \( \theta \) and sample the conditional distribution \( f(k \mid \cdot) \) using a standard Metropolis-Hastings sampler to propose a candidate dimension \( k^* \). In the second step, we fix the variable \( k \) (accepted in the first step), and sample the conditional distribution \( f(\theta \mid \cdot) \) to obtain a candidate parameter \( \theta^* \) using the pCN proposal in (10). The state update in MwG is formally described in Algorithm 1. Observe that this approach requires two PDE model evaluations for the generation of one state of the chain.

Since \( k \) is a discrete variable, the proposal distribution in the standard Metropolis-Hastings algorithm is replaced by a proposal matrix \( Q \in \mathbb{R}^{k_{\max} \times k_{\max}} \). This proposal is a right stochastic matrix containing the probabilities of the moves [7]. Such probabilities are assigned using a discrete probability law that is controlled by a parameter \( 1 \leq \rho \leq k_{\max} \) defining the spread or jump lengths of the proposal.

3.3. Step-wise sampler

We now construct a step-wise sampler 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 [15, 16]. Following the posterior (9), we can define a proposal distribution \( q(k, \theta) = q_1(k)q_2(\theta) \). This constitutes a joint proposal across the full state space \( X \) that accounts for the proposal of the dimension move \( q_1(k) = Q(k, \cdot) \) and the proposal of the parameters in the saturated space \( q_2(\theta) \). Under these assumptions the acceptance probability of the standard Metropolis-Hastings algorithm becomes [16]
\[ \alpha(k, \theta; k^*, \theta^*) = \min \left\{ 1, \frac{q_1(k^*)q_2(\theta^*) f(k^*, \theta^* \mid \tilde{y})}{q_1(k)q_2(\theta) f(k, \theta \mid y)} \right\}. \]

We can employ the pCN proposal for the parameter vector \( \theta \) in the saturated space. In this case, the proposal \( q_2(\theta) \) cancels out with the saturated parameter prior in the target posterior. The resulting acceptance probability simplifies to
\[ \alpha(k, \theta; k^*, \theta^*) = \min \left\{ 1, \frac{L(k^*, \theta^*; \tilde{y})}{L(k, \theta; \tilde{y})} \cdot \frac{f(k^*) Q(k^*, k)}{f(k) Q(k, k^*)} \right\}. \]

This Metropolis-Hastings implementation on the composite space proceeds in a “step-wise” fashion as follows: In the first step of the simulation, a candidate dimension \( k^* \) is proposed according to the proposal matrix \( Q \). In the second step, a candidate parameter \( \theta^* \) is proposed using the pCN proposal in (10).
Algorithm 1: State update in the Metropolis-within-Gibbs sampler

1. Let \((k, \theta)\) be the current state of the Markov chain
2. // Step 1: For fixed \(\theta\), sample the conditional distribution \(\pi_1(k) = f(k|\theta)\)
3. Sample the dimension, \(k^* \sim Q(k,.)\)
4. Compute the acceptance probability
   \[ \alpha_k \leftarrow \min \left\{ 1, \frac{L(k^*, \theta; \tilde{y})}{L(k, \theta; \tilde{y})} \frac{\pi_1(k^*) Q(k^*, k)}{\pi_1(k) Q(k, k^*)} \right\} \]
5. Sample, \(U_k \sim \text{Unif}(0, 1)\)
6. if \(U_k < \alpha_k\) then
7. \(k_{\text{next}} \leftarrow k^*\)
8. else
9. \(k_{\text{next}} \leftarrow k\)
10. end
11. // Step 2: For fixed \(k_{\text{next}}\), sample the conditional distribution \(\pi_2(\theta) = f(\theta|k_{\text{next}})\)
12. Compute the acceptance probability \(\alpha_{\theta} \leftarrow \min \{ 1, \frac{L(k_{\text{next}}, \theta^*; \tilde{y})}{L(k_{\text{next}}, \theta; \tilde{y})} \} \)
13. Sample, \(U_{\theta} \sim \text{Unif}(0, 1)\)
14. if \(U_{\theta} < \alpha_{\theta}\) then
15. \(\theta_{\text{next}} \leftarrow \theta^*\)
16. else
17. \(\theta_{\text{next}} \leftarrow \theta\)
18. end
19. output: \((k_{\text{next}}, \theta_{\text{next}})\)

Afterwards, the candidate pair \((k^*, \theta^*)\) is rejected or accepted jointly according to the probability (13). Algorithm 2 describes the procedure in detail.

We remark that from the RJMCMC viewpoint, the jumps in the step-wise sampler take place between nested models differing in dimension according to the proposal matrix \(Q\). Because of the nested structure of the K-L expansion, a natural jumping function linking the parameter spaces is the identity; this makes the determinant of the Jacobian of the jumping function in RJMCMC equal to one [41]. This approach is discussed in [16].

Algorithm 2: State update in the step-wise sampler

1. Let \((k, \theta)\) be the current state of the Markov chain
2. // Step 1: Sample the dimension
3. Draw a candidate dimension \(k^* \sim Q(k,.)\)
4. // Step 2: Sample the coefficients using pCN
5. Draw candidate coefficients \(\theta^* \leftarrow \sqrt{1 - \beta^2} \theta + \beta \xi\), where \(\xi \sim N(0, \Sigma_{pr})\)
6. Compute the acceptance probability
   \[ \alpha_{k, \theta} \leftarrow \min \left\{ 1, \frac{L(k^*, \theta^*; \tilde{y})}{L(k, \theta; \tilde{y})} \frac{f(k^*) Q(k^*, k)}{f(k) Q(k, k^*)} \right\} \]
7. Sample, \(U_{k, \theta} \sim \text{Unif}(0, 1)\)
8. if \(U_{k, \theta} < \alpha_{k, \theta}\) then
9. \(k_{\text{next}} \leftarrow k^*\) and \(\theta_{\text{next}} \leftarrow \theta^*\)
10. else
11. \(k_{\text{next}} \leftarrow k\) and \(\theta_{\text{next}} \leftarrow \theta\)
12. end
13. output: \((k_{\text{next}}, \theta_{\text{next}})\)

Illustration 1 (Polynomial regression). We conclude this section by illustrating the performance of the MCMC samplers with a simple example (a modified version of Example 6.10 in [18]). The problem setup corresponds to inferring the polynomial degree and coefficients of a regression model using observations of a
model response. The relationship between the input \( \tilde{x} \) and output \( \tilde{y} \) is modeled as

\[
\tilde{y}_i = \sum_{j=0}^{k} \theta_j \tilde{x}_i^j + \eta_i, \quad i = 1, \ldots, m,
\]

where \( m \) is the number of data points, and each \( \eta_i \) is a Gaussian distributed error with mean zero and variance \( \sigma_{\text{obs}}^2 = 1 \). The model \( 14 \) can be expressed in matrix form as, \( \tilde{y} = X\theta + \eta \), where \( \tilde{y} \) is the data vector, \( X \) is the design (Vandermonde) matrix, \( \theta \) the parameter vector, and \( \eta \) the vector of random errors. A set of \( m = 101 \) data points \( \tilde{y} \) is generated from the relation (14) assuming a quadratic polynomial model with coefficients \( \theta = [1.0, 0.3, -0.2]^T \). The design matrix is constructed at points \( \tilde{x}_i = (i - 1)/20, i = 1, 2, \ldots, m \).

The dimension of the problem is controlled by the polynomial degree in the regression. In order to evaluate which degree better fits the data, we employ the MCMC algorithms discussed in this section. The following settings are considered: (i) The dimension \( k \) is supported on the set \{1, 2, 3\}, thus we impose on \( k \) a discrete uniform distribution between 1 and 3. (ii) The parameters \( \theta \) are assumed to be Gaussian distributed with mean zero and covariance \( \Sigma_{\text{pr}} = \sigma_{\text{pr}}^2 I_3 \); the prior variance is \( \sigma_{\text{pr}}^2 = 4 \). (iii) The jumping proposal matrix \( Q \) is constructed such that all the dimensions are equally possible (i.e., with probability 1/3).

We run \( N_{\text{sim}} = 50 \) independent simulations of the MwG and step-wise algorithms. At each simulation, the length of the Markov chains is \( N = 1 \times 10^5 \) and a burn-in period of \( N_b = 0.2N \) is applied. For this illustration, the scaling of the pCN proposal is fixed as \( \beta = 0.25 \) for both algorithms. In MwG, this choice of \( \beta \) yields an average acceptance probability of 0.334 for the dimension step and \( 9.949 \times 10^{-3} \) for the parameter step. In the step-wise sampler, the average joint acceptance probability is \( 3.019 \times 10^{-3} \).

The results of the simulation are shown in Figure 1 for MwG in the first row and the step-wise algorithm in the second row. The values of the data-misfit function are plotted for the \( N \) samples of a single simulation; the sharp jumps in the values are due to the change in dimension. The regression solution is shown in terms of the mean and 95\% credible intervals (shaded areas) for the two dimensions that give the highest model posterior (linear and quadratic models). The mean value of the parameters computed by MwG is \([1.652, -0.650]\) for the linear, and \([0.916, 0.247, -0.180]\) for the quadratic model; in the step-wise approach these values are respectively, \([1.651, -0.650]\) and \([0.924, 0.243, -0.179]\). The model posterior computed as the average from the \( N_{\text{sim}} \) independent simulations indicates that the quadratic model gives the best fit. We also show the mean and 95\% CIs of the cumulative mean \( \hat{\mu}(k_t) = E[k_t | \tilde{y}] \), computed from \( N_{\text{sim}} \) independent simulations. Here, \( k_{1,t} \) indicates the first \( t \) states of the dimension chain.

![Figure 1: Comparison between MwG (1st row) and step-wise sampler (2nd row) algorithms. The evolution of the negative log-likelihood function is shown for a single simulation. The mean and quantile estimators in the remaining plots are computed from 50 independent simulations.](image-url)
4. Bayesian inference with subset simulation in spaces of varying dimensions

The characterization of the posterior distribution using standard MCMC algorithms can be inefficient not only because several iterations are required to compute accurate statistics, but also because tuning and post-processing steps need to be implemented (e.g., burn-in and lag periods). The task is even more complicated 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 [52, 53]. The idea is to explore the posterior on-the-fly by constructing a set of intermediate measures that converge to the full posterior. An alternative approach is provided by the BUS (Bayesian Updating with Structural reliability methods) framework, which reformulates the Bayesian inverse problem as a classical reliability analysis (rare event estimation) problem. This construction allows one to employ efficient sampling-based reliability estimation algorithms to sample from the posterior. In this section, we discuss the combination of BUS with subset simulation [28] and its extension to variable-dimensional inference problems. Since BUS can be interpreted as an extension of classical rejection sampling, the next section starts by reviewing the latter.

4.1. The BUS formulation

When the computation of the model evidence is intractable, the posterior density is only known up to its scaling constant, that is \( f(\theta | y) \propto f(\theta) L(\theta; y) = \pi(\theta) \). We refer to \( \pi(\theta) \) as the target function. The posterior density can be characterized by drawing samples from this un-normalized density. Particularly, the rejection sampling algorithm generates samples from \( \pi(\theta) \) using a proposal density \( q(\theta) \). The proposal is selected such that it dominates the target function. This means that \( q(\theta) \) must have equal or heavier tails than those of \( \pi(\theta) \). Generally, it has been shown that the prior must have sufficiently heavy tails to achieve stability of the Bayesian inverse problem solution; the conditions are discussed in [54]. Therefore, the proposal satisfies the relation

\[
\sup_{\theta} \left( \frac{\pi(\theta)}{q(\theta)} \right) \leq \bar{c} < \infty \quad \text{for some covering constant } \bar{c} \in \mathbb{R}_{>0}
\]

and \( \text{sup} (\pi(\theta)) \subseteq \text{sup} (q(\theta)) \). Thereafter, samples drawn from \( q(\theta) \) are rejected strategically to make the resulting accepted samples distributed according to \( \pi(\theta) \).

A simple choice for the proposal density is the prior distribution \( f(\theta) \). In this case, the acceptance probability \( \alpha \) in rejection sampling [23] becomes

\[
\alpha = \frac{\pi(\theta)}{\bar{c} \cdot q(\theta)} = \frac{f(\theta) L(\theta; y)}{\bar{c} \cdot f(\theta)} = c \cdot L(\theta; y),
\]

where \( c = 1/\bar{c} \in \mathbb{R}_{>0} \) and the covering constant is selected such that \( \bar{c} \geq L_{\text{max}} = \max(L(\theta; y)) \). Rejection sampling then amounts to (i) drawing a standard uniform random number \( v \sim \text{Unif}[0, 1] \), (ii) sampling a candidate from the prior \( \theta \sim f(\theta) \), and (iii) accepting the candidate if \( v \leq \alpha = c \cdot L(\theta; y) \). This particular acceptance-rejection mechanism allows us to generate the space

\[
\mathcal{H} = \{ (\theta, v) \in \Theta : h(\theta, v) \leq 0 \}, \quad \text{where } h(\theta, v) = v - c \cdot L(\theta; y)
\]

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

In the context of reliability analysis and rare event estimation, the space \( \mathcal{H} \) defines a failure domain with limit-state function (LSF) \( h(\theta, v) \). Samples drawn from the prior that fall into \( \mathcal{H} \) are distributed according to the posterior. 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 \( \Theta \).

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 \( \mathcal{H} \), i.e.,
The Markov chains are initialized from where θ vector in (17) yields (56) in order to guarantee a smooth transition between the intermediate levels, as well as for numerical stability, the BUS formulation, this also translates to the Bayesian inversion. Hence, a new standard Gaussian parameter evidence via (18).

Note that the application of BUS requires the knowledge of the constant $c = 1/\bar{c}$. From (16), it is seen that the covering constant is optimally chosen as the maximum of the likelihood function $\bar{c} = L_{\text{max}}$. If $\bar{c} > L_{\text{max}}$, the efficiency of BUS decreases because the value of $p_H$ will be small and more samples are required for its estimation. Since in many cases $L_{\text{max}}$ is not known in advance and its computation poses an additional cost, we employ the strategy introduced in [55] for which the constant $c$ is adaptively computed at each step of the simulation. We discuss this method in the variable-dimensional context in Subsection 4.3.

4.2. 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) [28]. The combination of BUS with SuS (termed BUS-SuS), performs Bayesian inversion sequentially. This is because SuS transforms the task of estimating the rare event $\{h(\theta, \nu) \leq 0\}$ into a sequence of problems involving more frequent events.

In BUS-SuS, the parameter space $\Theta$ 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, i.e., $\Theta = H_0 \supset H_1 \supset \cdots \supset H_{N_{\nu}} = \mathcal{H}$, where $N_{\nu}$ is the number of intermediate levels. Based on the general product rule of probability, the probability that the prior samples fall into the posterior space, $p_H$ is given by

$$p_H = P\left[ \bigcap_{j=0}^{N_{\nu}} H_j \right] = P\left[ H_{N_{\nu}} \mid \bigcap_{j=0}^{N_{\nu}-1} H_j \right] P\left[ \bigcap_{j=0}^{N_{\nu}-1} H_j \right] = \prod_{j=1}^{N_{\nu}} P[H_j \mid H_{j-1}],$$

where $P[H_j \mid H_{j-1}]$ represents the conditional probability at level $(j-1)$. Each intermediate level is defined as the set $H_j = \{ (\theta, \nu) \in \Theta : h(\theta, \nu) \leq \tau_j \}$, where $\infty = \tau_0 > \tau_1 > \cdots > \tau_j > \cdots > \tau_{N_{\nu}} = 0$, is a decreasing sequence of threshold levels. In practice, it is not possible to make an optimal a priori selection of the sequence $\{\tau_j\}_{j=0}^{N_{\nu}}$. Therefore, they are adaptively selected as the $p_0$-percentile of the LSF values of the samples simulated at intermediate level $H_{j-1}$ [28]. This implies fixing the conditional probabilities to a common value $p_0 = P[H_j \mid H_{j-1}]$ (with $p_0 \in [0.1, 0.3]$ [28]).

At the first level, $H_0$ samples are generated using standard Monte Carlo simulation. Thereafter, BUS-SuS employs a modified MCMC algorithm to draw samples from each intermediate conditional density $f(\theta, \nu \mid H_j)$. The Markov chains are initialized from $N_{s} = N \cdot p_0$ samples conditional on $H_{j-1}$ for which $h(\theta, \nu) \leq \tau_j$. The process is repeated until the target posterior domain is reached (see [48] for further details). At the last level, the probability $p_H$ in (19) is estimated as $\tilde{p}_H = p_0^{N_{\nu}-1} \cdot \hat{p}_{N_{\nu}}$, where $\hat{p}_{N_{\nu}}$ 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 $\hat{p}_{N_{\nu}} \cdot N$ samples that lie in $\mathcal{H}$ are used as seeds to generate the final batch of $N$ samples conditional on $\mathcal{H}$. The resulting samples are uniformly weighted but correlated samples of the posterior distribution. Furthermore, the probability estimate $\tilde{p}_H$ can be used to compute the model evidence via (18).

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 $\vartheta = [\theta, \theta_v]^T \in \mathbb{R}^{k+1}$ can be created by combining the K-L coefficients $\theta$ and the transformed auxiliary variable $\theta_v = \Phi^{-1}(\nu)$, 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 (17) is expressed in terms of the log-likelihood. Applying the natural logarithm to each term of $h(\theta, \nu)$ in (17) yields [56]

$$h_{\ln}(\vartheta) = \ln(\Phi(\theta_v)) - \ln(c \cdot L(\theta; \hat{y})) = \ln(\Phi(\theta_v)) + \ln(\bar{c}) - \ln L(\theta; \hat{y}).$$

12
4.3. BUS with subset simulation in varying dimensions

We now extend the concepts of Subsections 4.1 and 4.2 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 rejection sampling algorithm in Subsection 4.1 can be reformulated accordingly. The joint posterior distribution is characterized by a target function in a discrete-continuous space, $\pi(k, \theta) = f(k) f(\theta_k \mid k) L(k, \theta_k; \hat{y}) \propto f(k) f(\theta_k \mid \hat{y})$. Again, we choose the proposal distribution to be equal to the prior $q(k, \theta_k) = f(k) f(\theta_k \mid k)$. The acceptance probability $\alpha$ in rejection sampling becomes

$$\alpha = \frac{\pi(k, \theta)}{r \cdot q(k, \theta)} = \frac{f(k) f(\theta_k \mid k) L(k, \theta_k; \hat{y})}{r \cdot f(k) f(\theta_k \mid k)} = r \cdot L(k, \theta_k; \hat{y}),$$

where $r = 1/\bar{r} \in \mathbb{R}_{>0}$. By analogy with the fixed-dimensional setting, the covering constant $\bar{r}$ can be optimally chosen as $L_{\text{max,all}} = \max(K(k, \theta_k; \hat{y}))$, i.e., as the maximum of the likelihood function across different dimensions.

Samples drawn from the priors $k \sim f(k)$ and $\theta \sim f(\theta_k \mid k)$ are accepted if $r \leq \alpha = r \cdot L(k, \theta_k; \hat{y})$, otherwise they are rejected. In this case, the $H$-space and the LSF are now expressed as

$$H = \{(k, \theta_k, v) : h(k, \theta_k, v) \leq 0\}, \text{ where } h(k, \theta_k, v) = r - r \cdot L(k, \theta_k; \hat{y}),$$

where $\Theta^* = [K, \Theta_k, \Upsilon]$ is the re-augmented discrete-continuous parameter space ($k \in K \subseteq \mathbb{Z}^+, \theta_k \in \Theta_k \subseteq \mathbb{R}^k$ and $v \in \Upsilon \subseteq \mathbb{R}_{[0,1]}$). As in Section 3, we employ trans-dimensional MCMC algorithms that work in a saturated space for which $\theta \in \Theta \subseteq \mathbb{R}^{k_{\text{max}}}$, and thus $\Theta^* = [K, \Theta, \Upsilon]$.

The BUS-SuS algorithm can be extended analogously to solve the Bayesian inverse problem (22). Each intermediate domain is now defined as the set $H_j = \{(k, \theta, v) \in \Theta^* : h(k, \theta, v) \leq \tau_j\}$, with the threshold level sequence $\{\tau_j\}_{j=0}^{N_0}$ adaptively selected as in the fixed-dimensional case. Under the LSF (22), the standard MCMC algorithms used within BUS-SuS are no longer suitable for sampling the intermediate densities $f(k, \theta, v \mid H_j)$ and trans-dimensional MCMC methods are required. We modify these algorithms to sample the intermediate densities conditional on events defined by the sequence of levels $\{\tau_j\}$, as is typically carried out in the SuS context [48]. Moreover, instead of the LSF in (20) we employ its variable dimensional extension

$$h_{\text{un}}(k, \theta) = \ln(f(\theta, v)) + \bar{\nu} - \ln L(k, \theta; \hat{y})$$

where $\bar{\nu} = \ln(\bar{r})$ is optimally the maximum of the log-likelihood function. The tBUS-SuS method is summarized in Algorithm 3.

Since finding the constant $\bar{\nu} = \ln(\bar{r})$ poses an additional computational cost, it is convenient to introduce a tBUS-SuS algorithm for which the covering constant $\bar{r}$ is not required as an input. We employ the adaptive BUS-SuS methodology proposed in [55] for the trans-dimensional setting. In this case, the covering constant is updated at each level, leading to a set of values $\{\tau_j\}_{j=0}^{N_0}$. In order to guarantee the nestedness of the intermediate domains, the threshold levels $\tau_j$ are corrected after updating the value $\tau_j$. First note that from (23), a $j$-th intermediate domain is defined as the set

$$H_j = \{(k, \theta) : \ln(f(\theta, v)) + \bar{\nu} - \ln L(k, \theta; \hat{y}) \leq \tau_j\}$$

where $\bar{\nu} = \ln(\bar{r})$ is the maximum of the log-likelihood function observed at the $j$-th sampling level. The idea of [55] is that the event $H_j$ associated with the log-scaling constant $\bar{\nu}$ and the threshold $\tau_j$, can be equivalently expressed by a scaling $\tau_j^*$ and a modified threshold $\tau_j^*$ selected as $\tau_j^* = \tau_j - \bar{\nu} + \bar{\nu}$. This allows one to sequentially update the covering constant $\tau_j$ to a new value $\bar{\nu}^*$, without compromising the distribution of the samples: Adjusting the threshold value from $\tau_j$ to $\tau_j^*$ (after updating the scaling from $\bar{\nu}$ for $\bar{\nu}^*$ still defines the same intermediate domain, as $H_j = \{(k, \theta) : \ln(f(\theta, v)) \leq \tau_j^* - \bar{\nu} + \ln L(k, \theta; \hat{y})\}$ is equivalent to (24). In principle, $\tau_j^*$ corrects the level $\tau_j$ using the residual of maximum log-likelihood values observed at different levels. At the last simulation level when $\bar{\nu} = \bar{\nu}_j$, the covering reaches a value that is close or
Remark (3). We apply an adaptive version of the pCN algorithm used within the trans-dimensional samplers. The idea is to control the scaling parameter $\beta$ 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 [48, 55].

Remark (4). When MwG and step-wise samplers are applied within tBUS-SuS, the acceptance probabilities need to be modified to sample intermediate measures defined by the sequence of threshold levels $\{\tau_j\}_{j=0}^{N_{lv}}$. In this case, the LSF (23) needs to be checked at each proposed state. Therefore, at each intermediate level we employ Algorithms 1 and 2 with target density set to the density conditional on $H_j$, $f(k, \theta, v \mid H_j) = \mathbf{1}_{H_j}(k, \theta, v) f(k) f(\theta \mid k)$. This implies substituting the indicator function for the likelihood function in the acceptance probabilities of the two algorithms. For instance, in the step-wise sampler the modification reads equal to the actual maximum log-likelihood, i.e., $\tilde{r}_{N_{lv}} \leq \ln(L_{\text{max,all}})$. In the limit $N \to \infty$, the value $\tilde{r}_{N_{lv}}$ converges to $\ln(L_{\text{max,all}})$. Despite the fact that $\tilde{r}_{N_{lv}}$ is likely to be smaller than $\ln(L_{\text{max,all}})$, the samples generated by the algorithm follow the posterior distribution, as shown in [55]. The adaptive tBUS-SuS method is summarized in Algorithm 4.

Algorithm 3: tBUS-SuS.

<table>
<thead>
<tr>
<th>Input</th>
<th>number of samples per level $N$, conditional probability $p_0$, covering constant $\tilde{s}$, maximum dimension $k_{\text{max}}$, log-likelihood function $\ln L(\cdot, \cdot; \tilde{y})$, dimension prior $f(k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Draw $N$ samples from the dimension prior, $k_0 \sim f(k)$</td>
<td></td>
</tr>
<tr>
<td>2 Draw $N$ samples from the standard Gaussian, $\theta_0 = [\theta_0, \theta_{c,0}] \sim \mathcal{N}(0, \mathbf{I}<em>{k</em>{\text{max}}+1})$</td>
<td></td>
</tr>
<tr>
<td>3 Compute initial log-likelihood function values, $L_{\text{eval}} \leftarrow \ln L(k_0, \theta_0; \tilde{y})$</td>
<td></td>
</tr>
<tr>
<td>4 Set $j \leftarrow 0$ and $\tau_0 \leftarrow \infty$</td>
<td></td>
</tr>
<tr>
<td>5 while $\tau_j &gt; 0$ do</td>
<td></td>
</tr>
<tr>
<td>6 Increase intermediate level counter, $j \leftarrow j + 1$</td>
<td></td>
</tr>
<tr>
<td>7 Compute LSF values, $h_{\text{eval}} \leftarrow \ln(\Phi(\theta_{c,j-1})) + \tilde{s} - L_{\text{eval}}$</td>
<td></td>
</tr>
<tr>
<td>8 Sort $h_{\text{eval}}$ in ascending order and create a variable $id$ to store the indices of this sorting</td>
<td></td>
</tr>
<tr>
<td>9 Create $k_{\text{sort}}$, $\theta_{\text{sort}}$ as the dimension and parameter samples $k_{j-1}, \theta_{j-1}$ sorted according to $id$</td>
<td></td>
</tr>
<tr>
<td>10 Set the intermediate threshold level $\tau_j$ as the $p_0$-percentile of the values in $h_{\text{eval}}$</td>
<td></td>
</tr>
<tr>
<td>11 Compute the number of samples in the $j$-th intermediate level, $N_{H_j} \leftarrow \sum_{i=1}^{N}(h_{\text{eval}} \leq \max(0, \tau_j))$</td>
<td></td>
</tr>
<tr>
<td>12 if $\tau_j &gt; 0$ then</td>
<td></td>
</tr>
<tr>
<td>13 $\quad$ Select seeds for the MCMC step, $(k_{\text{seed}}, \theta_{\text{seed}}) \leftarrow \left( k_{\text{sort}}, \theta_{\text{sort}} \right)<em>{i=1}^{N</em>{H_j}}$</td>
<td></td>
</tr>
<tr>
<td>14 $\quad$ Randomize the ordering of the seeds</td>
<td></td>
</tr>
<tr>
<td>15 $\quad$ Generate next level values $(k_j, \theta_j, L_{\text{eval}})<em>{i=1}^{N}$ from the randomized seeds $(k</em>{\text{seed}}, \theta_{\text{seed}})$ and intermediate level $\tau_j$ using a trans-dimensional MCMC algorithm (e.g. Algorithms 1 and 2 with likelihood set equal to the indicator function as in (25)). Here, each seed is used to construct a chain with $N_s = \lfloor N_{H_j} \rfloor$ states, where $N_s = N_{H_j}$ is the number of seeds</td>
<td></td>
</tr>
<tr>
<td>16 end</td>
<td></td>
</tr>
<tr>
<td>17 Find the number of posterior samples that lie in dimension $k$, $N_k \leftarrow \sum_{i=1}^{N}(k_{\text{pos}}^{(i)} = k)$</td>
<td></td>
</tr>
<tr>
<td>18 Estimate the model posterior, $\tilde{f}^{(k)} \leftarrow N_k / N$</td>
<td></td>
</tr>
<tr>
<td>19 end</td>
<td></td>
</tr>
<tr>
<td>Output: sequence of posterior dimension samples ${k_{\text{pos}}^{(i)}}<em>{i=1}^{N}$ and parameter samples ${\theta</em>{\text{pos}}^{(i)}}_{i=1}^{N}$, and model posterior $\tilde{f}$.</td>
<td></td>
</tr>
</tbody>
</table>
as follows:

\[
\alpha_{k, \theta} \leftarrow \min \left\{ 1, \mathbb{1}_{H_j}(k^*, \theta^*) \frac{f(k^*)}{f(k^*)} \mathcal{Q}(k^*, k) \right\} = \mathbb{1}_{H_j}(k^*, \theta^*) \min \left\{ 1, \frac{f(k^*)}{f(k^*)} \mathcal{Q}(k^*, k) \right\}.
\]  

(25)

We implement the right term in (25): First, a model \( k^* \) is proposed which is accepted with probability (\( * \)); then \( \theta^* \) is drawn from a pCN proposal; finally the pair \( (k^*, \theta^*) \) is accepted if it lies in the domain \( H_j \). The modification is analogous for each acceptance rate in MwG.

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 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 \( p_0 = 0.1 \).

5.1. 1D cantilever beam problem

The first example is an inverse problem involving 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 [22]. 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 \( F(x) = (I E(x))^{-1} \) 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 [57]:

\[
M(x) = -F^{-1}(x) \frac{d^2 w(x)}{dx^2} \quad \Rightarrow \quad w(x) = -P \int_0^x \int_0^s (L - t) F(t) \, dt \, ds,
\]  

(26)

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

The flexibility is modeled by a Gaussian random field prior \( \mathcal{N}(\mu_{pr}, \Sigma_{pr}) \), with constant mean vector \( \mu_{pr} = 1 \times 10^{-4} \) (kN\(^{-1}\)m\(^{-2}\)) and covariance matrix \( \Sigma_{pr} \) defined through a Matérn kernel with smoothing parameter \( \nu = 0.5 \), which yields the following exponential autocovariance function:

\[
C(x, x') = \sigma_{pr}^2 \exp \left( -\frac{||x - x'||}{\ell} \right) \quad \text{for} \ x, x' \in D.
\]  

(27)
We set the prior standard deviation to $\sigma_{pr} = 3.5 \times 10^{-4} \text{ (kN}^{-1}\text{m}^{-2})$ and perform a parameter study on the correlation length $\ell$. Note that for this type of covariance operator, the K-L eigenvalue problem has an analytical solution [21].

The true flexibility field is a realization from the prior random field (with correlation length $\ell_{\text{true}} = 2$ m). Partial observations of the deflection field are generated by simulating the ODE (26) using this underlying realization. The data is collected at $m = 10$ equally-spaced points of the domain $D$ (Figure 2). This generates a measurement vector $\tilde{y} \in \mathbb{R}^{m \times 1}$ with additive and spatially correlated error described by a Gaussian PDF, $\eta \sim N(0, \Sigma_{\text{obs}})$, where the covariance structure of the error is constructed from an exponential kernel with standard deviation $\sigma_{\text{obs}} = 1 \times 10^{-3}$ and correlation length $\ell_{\eta} = 1$ m.

For this example, closed-form expressions of the model evidence for each dimension $k$ are available (see [22]), 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 also defines different dimension priors as follows:

- for $\ell = 0.1$, the truncation parameter is $k_{\text{max}} = 1014$ and the $k_{\text{min}} = 17$. This yields $p = 6.166 \times 10^{-3}$.
- for $\ell = 0.5$, the truncation parameter is $k_{\text{max}} = 204$ and the $k_{\text{min}} = 4$. This yields $p = 2.586 \times 10^{-2}$.
- for $\ell = 0.9$, the truncation parameter is $k_{\text{max}} = 114$ and the $k_{\text{min}} = 2$. This yields $p = 5.118 \times 10^{-2}$.

The priors together with the analytical model posterior and evidence are shown in Figure 3. We employ these closed-form solutions as reference to test the performance of the proposed tBUS-SuS approach.

Figure 3: Closed-form solutions for different correlation lengths in the prior flexibility random field. Left: Model evidence. Center: Model/dimension prior. Right: Model/dimension posterior.
We initially perform parameter studies on the proposal scalings of the trans-dimensional MCMC methods. This allows us to find a set of near-optimal parameter settings that are employed to solve the Bayesian model choice problem. Later, we perform a comparison between within-model BUS-SuS runs and across-model tBUS-SuS. This is carried out by computing an efficiency metric that is equivalent for both approaches; the metric is defined in the Appendix. At the end of this section, we use tBUS-SuS to estimate the posteriors of the dimension and random fields.

5.1.1 Influence of the proposal scaling

We first evaluate the performance of the tBUS-SuS algorithm for different proposal scalings, this includes the parameter $\rho$ of the jump proposal $Q$ and the parameter $\beta$ of the pCN proposal. The numerical studies are performed for a fixed prior correlation length $\ell = 0.5$ and for each MCMC sampler presented in Section 3. The results are shown for an average of $N_{\text{sim}} = 100$ independent simulations using $N = 10^4$ samples per level.

In this example, the constant $\hat{k}$ can be computed by optimization or from initial log-likelihood function evaluations at the largest dimension $k_{\text{max}}$. Therefore, the parameter studies are performed using standard tBUS-SuS without adaptation of $\hat{k}$. The performance of the algorithm is measured in terms of the root-mean-squared error (RMSE) in the estimation of a quantity of interest (QoI)

$$\text{RMSE}(k) = \sqrt{\mathbb{E} \left[ (\hat{Q}(k) - Q)^2 \right]},$$

(28)

where $\hat{Q}(k)$ is the approximated QoI using a K-L expansion with $k$ terms, and $Q$ denotes the reference QoI value. We monitor three posterior QoIs, namely, the dimension parameter $k$, the flexibility random field at the middle of the beam $F_{\text{mid}}$, and the deflection random field at the tip of the beam $w_{\text{tip}}$. For the evaluation of (28), we employ the closed-form model posterior as reference value for dimension QoI, and for the remaining random fields QoIs we use closed-form posterior means, $\mu_{F_{\text{mid}}} = 1.1114 \times 10^{-4}$ and $\mu_{w_{\text{tip}}} = -6.9346 \times 10^{-2}$, as reference values.

Jump length $\rho$: The pCN proposal scaling is fixed to $\beta = 0.2$ and we estimate the RMSE of the QoIs for different jump lengths $\rho = \{1, [0.1, 0.25, 0.5, 1.0] \cdot k_{\text{max}}\} = \{1, 20, 51, 102, 204\}$. The jump proposal matrix $Q$ is constructed from a discrete triangular distribution with mode equal to a given sample $k$, and lower and upper limits defined by $k \pm \rho$. Alternatively, we also evaluate an independent-sampler-like setting in the dimension sampling steps, that is, instead of drawing samples from $Q$, the variable $k$ is sampled directly from the dimension prior $f(k)$.

![Figure 4: Parameter study on the jump length $\rho$. Acceptance rates in the MCMC algorithms.](image)

Figure 4 shows the acceptance rates obtained in the initial and last levels of tBUS-SuS with step-wise (jointly for $k$ and $\theta$) and MwG (separately for $\theta$ and $k$). The acceptance rate in the step-wise sampler is almost insensitive to the jump length $\rho$, since the parameters are jointly accepted or rejected with the dimension, and the pCN proposal is dominating the acceptance rate. Similar values are obtained for the acceptance rate of the parameter $\theta$ in MwG. From the first to the last level of tBUS-SuS, there is a decrease of the acceptance rate since the samples in the posterior region are more concentrated and the scaling $\beta$...
remains fixed. Moreover, it is seen that for increasing values of the jump length \( \rho \), the acceptance rate of \( k \) decreases since the spread of the jump proposal is larger.

In Figure 5, we plot the evolution of the RMSE with the number of terms in the K-L expansion, for each investigated QoI. The MwG sampler yields slightly better approximation of the model posterior as compared to the step-wise sampler. Note that the errors are larger around the model index that yields the largest value of the model posterior; the tBUS-SuS estimates become more accurate after exceeding this truncation level. There are no considerable differences between model posterior approximations obtained by sampling from the dimension prior or using the jumping proposal with an appropriate jump length \( \rho \). Still, we observe that sampling from the prior produces smaller errors in the estimation of the model posterior. If we employ the jumping proposal \( Q \), it is seen that \( \rho = 0.25 \cdot k_{\text{max}} = 51 \) is preferable for the MwG sampler, whereas intermediate to smaller jumps are more suitable for the step-wise sampler (\( \rho \in [0.1, 0.25] \cdot k_{\text{max}} = [20, 51] \)). For this example, taking dimension samples from the prior appears to be the most appropriate choice since the model prior and posterior are close to each other (Figure 3). Based on these results, we henceforth choose to sample directly from the prior \( f(k) \) at the dimension proposal step of both MCMC samplers.

**Scaling of pCN \( \beta \):** The dimension prior is used to sample \( k \) and we estimate the RMSE for different pCN proposal scalings \( \beta = \{0.01, 0.1, 0.2, 0.3, 0.4\} \).

As in the previous study, we initially show in Figure 6 the acceptance rates obtained at the initial and last levels of tBUS-SuS using step-wise (jointly for \( k \) and \( \theta \)) and MwG (separately for \( \theta \) and \( k \)) methods. The acceptance rates obtained for the parameter \( \theta \) are comparable in both MCMC samplers. In this case, the scaling \( \beta \) has no effect on the acceptance of the dimension samples \( k \) of the MwG method.

The evolution of the RMSE with the truncation of the K-L expansion for the investigated QoIs and for different \( \beta \) is shown in Figure 7. The MwG sampler yields a better estimation of the model posterior than the step-wise sampler. In this case, the scaling \( \beta \) has less effect on the approximation as compared to the jump length \( \rho \). There are only minor differences in the RMSE values generated from the investigated proposal scalings; only for \( \beta = 0.01 \) the errors associated to the model posterior estimate increase. Note however that for the step-wise sampler, different \( \beta \) values generate different error levels in the estimation of the model posterior. In both methods, the errors in the QoIs corresponding to the random fields are almost
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 8 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, i.e., the one-dimensional K-L 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 (Figure 3), 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 \( \kappa_{\text{max}} = 204 \) is \( c_{204} = 68.170 \), and at dimension \( k = 1 \) it is \( c_1 = 56.529 \). Due to the nested structure
The results are computed for an average of within-model runs of BUS-SuS. The objective is twofold, (i) identify the number of effective independent samples in standard tBUS-SuS, as shown in Table 1. In this case, the efficiency is not a function of the random fields, we compare the approximation of the model posterior between standard tBUS-SuS (with pre-defined constant \( r \)) and adaptive tBUS-SuS (Algorithm 4). The adaptive version is a more general method since it is not always possible to define the constant \( r \) a priori. Moreover, it can produce similar or better results than the standard tBUS-SuS depending on the accuracy of the constant \( r \) in standard tBUS-SuS, as shown in Table 1. In this case, the efficiency is not a function of \( k \) because we

Figure 8: tBUS-SuS samples of the K-L 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.

For \( k = 2 \), Figure 8 plots the components of the two-dimensional K-L 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 (Figure 3).

5.1.2. Computational efficiency

We now investigate the computational gain of using the across-model tBUS-SuS, compared to 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. The results are computed for an average of \( N_{\text{sim}} = 100 \) independent simulations using \( N = 5 \times 10^3 \) samples per level in BUS-SuS and \( N = 10^4 \) samples per level in tBUS-SuS.

For each correlation length \( \sigma_k^2 = \{0.1, 0.5, 0.9\} \), we use the reference variances of the QoIs for the estimation of the effective number of samples in (36):

- \( \sigma_k^2 = \{24161.22, 1250.80, 328.56\} \) for the dimension,
- \( \sigma_{\text{flex}}^2 = \{1.143 \times 10^{-9}, 7.484 \times 10^{-10}, 5.364 \times 10^{-10}\} \) for the flexibility, and
- \( \sigma_{\text{def}}^2 = \{8.446 \times 10^{-7}, 9.042 \times 10^{-7}, 9.159 \times 10^{-7}\} \) for the deflection.

Standard tBUS-SuS vs. Adaptive tBUS-SuS: Before computing the efficiencies in the estimation of the QoIs related to the random fields, we compare the approximation of the model posterior between standard tBUS-SuS (with pre-defined constant \( \tilde{r} \), Algorithm 3) and adaptive tBUS-SuS (Algorithm 4). The adaptive version is a more general method since it is not always possible to define the constant \( \tilde{r} \) a priori. Moreover, it can produce similar or better results than the standard tBUS-SuS depending on the accuracy of the constant \( \tilde{r} \) in standard tBUS-SuS, as shown in Table 1. In this case, the efficiency is not a function of \( k \) because we
are estimating the efficiency of the quantity $k$ itself. The results show that overall the efficiency of adaptive tBUS-SuS is larger or comparable to the efficiency provided by standard tBUS-SuS. This is related to the way we estimate $\bar{r} = \ln(\bar{r})$ in tBUS-SuS. The constant is chosen as the maximum of $10^3$ independent log-likelihood evaluations; this value is also increased by 25% such that $\bar{r} \approx L_{\text{max,all}}$. However, it is possible this way of selecting $\bar{r}$ is too conservative and more levels in tBUS-SuS are required to estimate the solution. This is not an issue of the method, but of course will lead to a reduced efficiency since more model evaluations are required. Therefore, we employ the adaptive tBUS-SuS algorithm for the solution of the Bayesian model choice problem in the remainder of the paper.

\begin{table}[h]
\centering
\begin{tabular}{c|cc|cc}
\hline
$\ell$ & adaptive: eff$_{\text{tBUS}}(k)$ & standard: eff$_{\text{tBUS}}(k)$ \\
& MwG & step-wise & MwG & step-wise \\
\hline
0.1 & $8.01 \times 10^{-2}$ & $4.24 \times 10^{-2}$ & $4.00 \times 10^{-2}$ & $1.79 \times 10^{-2}$ \\
0.5 & $5.49 \times 10^{-2}$ & $2.38 \times 10^{-2}$ & $6.13 \times 10^{-3}$ & $1.85 \times 10^{-2}$ \\
0.9 & $3.59 \times 10^{-2}$ & $2.41 \times 10^{-2}$ & $4.03 \times 10^{-3}$ & $2.08 \times 10^{-3}$ \\
\hline
\end{tabular}
\caption{Efficiency metric (37) of the dimension parameter $k$ for adaptive and standard tBUS-SuS.}
\end{table}

\begin{figure*}[h]
\centering
\includegraphics[width=\textwidth]{figure9.png}
\caption{Within-model BUS-SuS runs vs. tBUS-SuS: We compare the efficiencies in the estimation of the QoIs related to the random fields produced by within-model runs of adaptive BUS-SuS and adaptive tBUS-SuS. Figure 9 shows the efficiencies computed with individual adaptive BUS-SuS runs (1st column) and adaptive tBUS-SuS using the two trans-dimensional MCMC algorithms (2nd and 3rd columns). In the first row, we plot the total number of calls per dimension. In the fixed-dimensional BUS-SuS, the cost increases with the dimension since more intermediate levels are required to reach the posterior. Conversely, the cost is a single value for all the dimensions in across-model tBUS-SuS, and thus we need to distribute it according to the model posterior. Note 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. Moreover, in the second and third rows of Figure 9, the efficiencies of the random field QoIs are shown per dimension, since different K-L truncation orders yield different random field approximations. We remark that 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 (37), and MwG yields almost twice the cost of the step-wise sampler. Finally, we clearly see the advantage of employing across-model simulation algorithms for the solution of Bayesian model choice problems, as compared to single model runs.}
\end{figure*}

\subsection{Approximation of the dimension and random field posteriors}

Based on the settings discussed in the previous subsections, we employ the adaptive version of tBUS-SuS for the estimation of the model and random field posteriors. Again, the solution is computed as an average of $N_{\text{sim}} = 100$ simulations, using $N = 10^4$ samples per level.

The approximated model posteriors are shown in Figure 10 using the MwG and step-wise samplers. In this case, we plot the mean and standard deviation bounds of the approximation. 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. The differences are larger for smaller correlation lengths. The tBUS-SuS simulations require in average $N_{\text{av}} = 5$ 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. We use as reference the closed-form expressions of the posterior random fields [22]. Figure 11 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 (termed \emph{differential} deflection), in order to distinguish the approximations.
\[
\bar{N}_{\text{call}, \text{BUS}}(k) = f(k | \tilde{y}) \times 10^5
\]

Figure 9: Comparison between within-model BUS-SuS and across-model tBUS-SuS: Number of model calls and efficiency metrics \((37)\) 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).

The model choice estimate is given by the truncation order that yields the maximum model posterior (Figure 10), in this case \(k_{\text{best}} = \{10, 3, 3\}\) for the correlation lengths \{0.1, 0.5, 0.9\}, respectively. The model mixing estimate takes into account the whole dimension spectrum (up to \(k_{\text{max}}\)). Note that the reference CIs agree closely with the model mixing estimates since we use all the K-L expansions associated to the model posterior for the random field representation. For the larger correlation length, the model choice solution fails to capture the assumed true flexibility in different intervals of the domain.

5.2. 2D groundwater flow problem

We consider inference of the hydraulic conductivity field of an aquifer using observations of the hydraulic head measured at specific boreholes (see [58, 59]). We define an aquifer on the square domain \(D = [0, 1] \times [0, 1] \text{ km}^2\) with boundary \(\partial D\). Spatial coordinates are denoted by \(x = [x_1, x_2] \in D\). 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 \(\kappa(x, \omega)\) and sink or source terms \(J(x)\), the hydraulic head \(u(x)\) follows the elliptic PDE

\[
-\nabla \cdot [\kappa(x, \omega) \nabla u(x)] = J(x),
\]

with Dirichlet boundary condition, \(u(x) = 0\) for \(x \in \partial D\). 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
Figure 10: 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 11: 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.

unitary strengths, and are centered at locations $\mu_j = [0.25 \cdot i, 0.25 \cdot j]$ with $i, j = 1, \ldots, 3$, that is

$$ J(x) = \sum_{i=1}^{9} N(x; \mu_j^{(i)}, \sigma_j^2 I_2). $$

(30)
We employ the finite element method to solve the PDE (29) and $2 \times 80^2 = 1.28 \times 10^4$ three-noded triangular elements are used for the discretization.

The prior hydraulic conductivity field is modeled as a log-normal random field, $\kappa(x) := \exp(\kappa'(x))$. The underlying Gaussian field $\kappa'(x)$ has mean zero and standard deviation $\sigma_{\kappa'} = 1$ km/day. The covariance operator of the Gaussian field is constructed from a Whittle-Matérn kernel

$$C_{\nu}(d) = \sigma_{\kappa'}^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{\sqrt{2\nu d}}{\ell} \right)^\nu K_{\nu} \left( \frac{\sqrt{2\nu d}}{\ell} \right),$$

where $d = \|x - x'\|_2$, $\Gamma(\cdot)$ is the gamma function, $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind. We set the smoothing parameter to $\nu = 1.5$. The solution of the eigenvalue problem is computed with the Nyström method using 100 Gauss-Legendre points in each direction [37].

The true conductivity $\kappa(x)$ 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 K-L expansion used to generate this realization is 312, which captures 99% of the prior variance. The hydraulic head observations $\hat{y}$ 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 $\Sigma_{\text{obs}} = \sigma_{\text{obs}}^2 I_m$. The variance of the measurement noise is prescribed such that the observations have a signal-to-noise ratio $\sqrt{V[\hat{y}]}/\sigma_{\text{obs}} = 120$. The true hydraulic conductivity and hydraulic head fields, together with the source terms and the synthetic data are shown in Figure 12.

In this example, we evaluate the posterior for different correlation lengths $\ell = \{0.1, 0.2, 0.3\}$. Each correlation length defines a dimension prior as follows:

- for $\ell = 0.1$, the truncation parameter is $k_{\text{max}} = 512$ and the $k_{\text{min}} = 16$. This yields $p = 6.2914 \times 10^{-3}$.
- for $\ell = 0.2$, the truncation parameter is $k_{\text{max}} = 138$ and the $k_{\text{min}} = 5$. This yields $p = 1.9398 \times 10^{-2}$.
- for $\ell = 0.3$, the truncation parameter is $k_{\text{max}} = 65$ and the $k_{\text{min}} = 3$. This yields $p = 2.9396 \times 10^{-2}$.

We employ the same proposal scaling settings investigated in the previous example. However, in this case the dimension posterior differs significantly from the dimension prior. Therefore, additional to sampling $k$ directly from the prior, the proposal $Q$ is used. The jump proposal matrix is constructed from a discrete triangular distribution with jump length $\rho = 0.25 \cdot k_{\text{max}}$, which appears to be a good choice for both MCMC algorithms (cf. Subsection 5.1.1).

### 5.2.1. Approximation of the posterior for the dimension and the random field

The 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 $N = 5 \times 10^4$ 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: $k_{\text{snap}} = \{45, 45, 55, 60, 70\}$ for $\ell = 0.1$, $k_{\text{snap}} = \{20, 30, 35, 40, 50, 60\}$ for $\ell = 0.2$, and $k_{\text{snap}} = \{20, 25, 30, 35, 40\}$ for $\ell = 0.3$. Since the reference solutions are given in terms of the model evidences $Z_y(k_{\text{snap}})$, 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

$$f(k_{\text{snap}} \mid \hat{y}) = \frac{f(k_{\text{snap}})Z_y(k_{\text{snap}})}{\sum_{k \in k_{\text{snap}}} f(k)Z_y(k)} \sum_{k \in k_{\text{snap}}} \hat{f}(k \mid \hat{y}),$$

such that the sum of the reference $f(k_{\text{snap}} \mid \hat{y})$ match the sum of the estimated model posteriors $\hat{f}(k_{\text{snap}} \mid \hat{y})$ at the given snapshots. Using such an approach, the reference solution is limited to the tBUS-SuS solution. Nevertheless it allow us to check that the algorithm is following the correct shape of the dimension posterior.
Figure 12: Groundwater flow problem: True hydraulic conductivity, true hydrostatic pressure with measurement locations, source term, measured and true (noise-free) hydraulic head.

The dimension posteriors estimated by adaptive tBUS-SuS with MwG are shown in Figure 13. 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 since there are no significant differences between the posterior approximations. In general, we observe an increase in the variability around the MAP estimate. Note also that in this example, 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 K-L eigenfunctions. For instance, when using $\ell = 0.3$ there is a jump in the values of the probability mass after the 15-th dimension, every 5 dimensions until the MAP estimate. Furthermore, for the correlation length $\ell = 0.1$ and sampling from the prior, the method requires on average $N_{lv} = 12$ intermediate levels and the proposal scaling $\beta$ changes from 0.75 in the first level to 0.09 in the last level.

The approximated dimension posteriors using adaptive tBUS-SuS with the step-wise sampler are shown in Figure 14. 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 $\beta$ 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 2 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 most cases, the estimation is very close to the reference mean value.

Furthermore, we estimate the posterior hydraulic conductivity and hydraulic head random fields for the
The proposed method is a variable-dimensional extension of the BUS formulation for the solution of Bayesian model choice problems. This extension defines an across-model simulation algorithm that allows one to jointly infer the uncertain parameters and their dimension in a sequential manner. Such type of inferences are common in random field updating tasks, when the number of terms in the random field series investigated correlation lengths. Figure 15 shows the model choice and model mixing solutions in terms of the posterior mean and standard deviation of the hydraulic conductivity. The model choice estimate is given by the truncation order that yields the maximum model posterior, in this case $k_{\text{best}} = \{43, 43, 33\}$ for the correlation lengths $\{0.1, 0.2, 0.3\}$ (Figure 14, 2nd row). The model mixing estimate takes into account the whole dimension spectrum. Note that the values in the posterior mean are smaller than those of the assumed truth. In this case, most of the measurements are concentrated in the lower left corner of the aquifer. In this area, the values of the true hydraulic conductivity are very small, and are influencing the posterior solution. Nevertheless, the statistics are revealing the locations of lower and higher permeability values. The field modeled with the smaller correlation length is able to represent the small fluctuations better than those with larger correlation lengths, for which the resulting random field realizations are smoother. In contrast to the hydraulic conductivity field, the differences in the model choice and mixing solutions for the hydraulic head random field are minimal. This quantity is computed by integrating the PDE model, this can be seen as an averaging operation that reduces the effect of the spatial variability, similar to Example 1.

6. Discussion of results

The proposed method is a variable-dimensional extension of the BUS formulation for the solution of Bayesian model choice problems. This extension defines an across-model simulation algorithm that allows one to jointly infer the uncertain parameters and their dimension in a sequential manner. Such type of inferences are common in random field updating tasks, when the number of terms in the random field series

Figure 15: 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.
expansion is modeled probabilistically.

The main findings of this contribution are summarized as follows: (i) If one employs a uniform prior for the dimension, the model evidences define the model posterior. In this case, visits to models with high dimension have the same probability of occurrence, e.g., in the beam example for the correlation length \( \ell = 0.5 \) a model with 10 terms in the K-L 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 3). Thus, the inclusion of a prior that penalizes models with increased number of parameters is beneficial in the context of the K-L 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) At present, it is common to perform inference at some fixed dimensions of interest in the random field discretization. This provides one with information of the dimension posterior at those specific interest points. With the proposed across-model simulation algorithm, the whole dimension posterior can be estimated at a much lower cost and with higher efficiency compared to individual runs for each dimension. This is verified for selected quantities of interest in Example 1 (cf. Figure 9). The tBUS-SuS algorithm provides a sample efficiency of around two orders of magnitude higher compared to individual within-model BUS-SuS runs. In order to populate the discrete-continuous parameter spaces, the algorithm requires a significant number of samples per level \( N \) to achieve an accurate solution; as shown in Example 2 the variability of the simulations results improve for larger \( N \). However, other sequential and trans-dimensional algorithms require a large amount of samples as well, see for instance [24, 25]. (iii) In the context of random fields, 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

Figure 16: 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.
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. (iv) 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 (apart from 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 (e.g., if the models correspond to different PDEs). This is beyond the scope of this paper and can be investigated in a future study. (v) The choice of the dimension proposal within the trans-dimensional MCMC algorithms plays an important role in the estimation of the model posterior. 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 itself 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. (vi) 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.

Finally we remark that one possible improvement of the method is to use the coverings \( \bar{c}_k | k \leq k_{max} \) for each particular dimension, instead of choosing the constant \( \bar{r} = \max(\bar{c}_k | k_{max} = 1) \). As discussed at the end of Subsection 5.1.1, 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 (23) 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. The method (termed tBUS-SuS) is a variable-dimensional extension of the BUS formulation when used in combination with subset simulation. 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. 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 MwG sampler and derive a step-wise sampler based on a simplification of the reversible jump MCMC algorithm in the saturated space. The results 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 samplers with tBUS-SuS, we perform different parameter studies to find optimal tuning settings for the proposal distributions. Both algorithms yield similar performance in terms of efficiency and quality 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 problems where the Karhunen-Loève expansion is used to represent random fields. The results compared to reference solutions indicate that the tBUS-SuS algorithm with the two
MCMC variants perform well in high- and trans-dimensional parameter spaces, i.e., it is able to estimate not only the posterior random fields but also the posterior distribution of the dimension in the discretization. Finally, in random field problems it is recommended to employ the model mixing solution instead of the single model choice solution, since the latter can fail to represent the spatial variability of the underlying random field.

Acknowledgements

This research has been supported by the *Deutsche Forschungsgemeinschaft* (DFG) through the TUM International Graduate School of Science and Engineering (IGSSE) within the project 10.02 BAYES. The numerical experiments were performed on the Linux clusters of the *Leibniz-Rechenzentrum* at the *Bayerische Akademie der Wissenschaften*.

Appendix

**PC prior**

The prior for the dimension in (7) can be seen as a PC prior in a discrete variable setting. Let $k = 1, \ldots, k_{\text{max}}$ and $f(\theta \mid k) := N(0, I_{1:k})$ be a (possibly degenerate) zero mean Gaussian measure on $\mathbb{R}^{k_{\text{max}}}$, with covariance

$$I_{1:k} := \text{diag}(1, \ldots, 1, 0, \ldots, 0).$$

The base model is defined as the Gaussian measure $f(\theta \mid k = 1)$. Note that the Gaussian measure $f(\theta \mid k)$ has no Lebesgue density on $\mathbb{R}^{k_{\text{max}}}$, for $k < k_{\text{max}}$. It is possible to generalize the Kullback-Leibler divergence to allow to compute distances between $f(\theta \mid k)$ and $f(\theta \mid k')$, for $k, k' = 1, \ldots, k_{\text{max}}$, $k' \leq k$. However, in this case we obtain

$$d_{\text{KL}}(f(\theta \mid k) \mid \mid f(\theta \mid k')) = \int \ln(N(\theta; 0, I_{1:k})) - \ln(N(\theta; 0, I_{1:k'})) \, dN(0, I_{1:k})(\theta)$$

$$= \int \ln(N(\theta; 0, I_{1:k})) - \ln \left( \left\{ \prod_{i=1}^{k'} N(\theta_i; 0, 1), \text{ if } \theta_{k'+1:k_{\text{max}}} = 0 \right\} \right) \, dN(0, I_{1:k})(\theta)$$

$$= \begin{cases} 0, & \text{if } k' = k, \\ \infty, & \text{otherwise}. \end{cases}$$

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

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

be the set of couplings between $f$ and $g$. That is, the set of joint probability density functions on $\mathbb{R}^{2k_{\text{max}}}$ that have $f$ and $g$ as marginals. Let $p \geq 1$. The *Wasserstein-\(p\) distance* between $f$ and $g$ is given by

$$d_{\text{Was}(p)}(f, g) := \left( \inf_{h \in \text{Cop}(f, g)} \int \|\theta^{(1)} - \theta^{(2)}\|^p h(\theta^{(1)}, \theta^{(2)}) d\theta^{(1)} d\theta^{(2)} \right)^{1/p}$$

(33)

between base and all the other models. Standard results about the Wasserstein-2 distance between two Gaussian measures indicate that

$$d_{\text{Was}(2)}(f(\theta \mid k), f(\theta \mid 1))) = \|I_{1:1} - I_{1:k}\|_F^2 = k - 1$$

(34)
and by choosing $K(x) := x$, the distance becomes $d(k) = k - 1$. Using this information, we can now construct the PC prior. Axiom (iii) in Subsection 2.4 induces a prior of the form

$$f(k) \propto \exp(-a \cdot d(k)) = \exp(-a(k-1)), \quad \text{for some } a > 0. \quad (35)$$

We obtain the truncated geometric prior (up to a normalizing constant) proposed in (7) by setting $a := \ln(1/(1-p))$.

**Efficiency metrics**

Consider a collection of samples of a random variable $Q$ representing a monitored QoI, $\{Q^{(j)}_i, j = 1, \ldots, N_{\text{sim}}, i = 1, \ldots, N\}$, where $N_{\text{sim}}$ is the number of independent simulation runs. Note that each element $Q^{(j)}_i$ 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 all $N_{\text{sim}}N$ samples as $[23]$

$$N_{\text{eff}}(Q) = N_{\text{sim}} \frac{\text{V}[Q]}{\text{V}[\hat{\mu}_Q^{(j)}]}, \quad (36)$$

where $\text{V}[Q]$ is the true variance of the QoI, and $\hat{\mu}_Q^{(j)}$ denotes the sample mean of the QoI obtained at the $j$-th simulation.

In addition to the value $N_{\text{eff}}$, we also take into account the computational cost. The number of model calls in BUS-SuS, is equal to $N_{\text{call}}^{(j)} = N + (N - N_a)(N_{N_v}^{(j)} - 1)$, where $N_a = p_0 \cdot N$ is the number of seeds used in the MCMC steps, and $N_{N_v}^{(j)}$ is the number of levels obtained at the $j$-th simulation. In the case of tBUS-SuS, the number of model calls depends on the trans-dimensional MCMC algorithm that is employed. If we use the step-wise sampler, the number of model calls per simulation run is, $N_{\text{call},a}^{(j)} = N + (N - N_a)(N_{N_v}^{(j)} - 1)$. If we employ the MwG algorithm, the number of model calls is $N_{\text{call},b}^{(j)} = N + 2(N - N_a)(N_{N_v}^{(j)} - 1)$ since the likelihood function is evaluated twice at the MCMC steps.

We define the efficiency metrics as follows:

$$\text{eff}_{\text{BUS}}(Q) = \frac{N_{\text{call},\text{BUS}}^{(k)}}{N_{\text{call},\text{BUS}}^{(k)}}, \quad \text{and} \quad \text{eff}_{\text{tBUS}}(Q) = \frac{N_{\text{call},\text{tBUS}}^{(k)}}{f(k | \tilde{y})}, \quad (37)$$

where $N_{\text{call},\text{BUS}}^{(k)}$ (or $N_{\text{call},\text{tBUS}}^{(k)}$) is the effective number of independent samples of the QoI obtained at each dimension $k$. Furthermore, in BUS-SuS the total number of calls is given by $N_{\text{call},\text{BUS}}^{(k)} = \sum_{j=1}^{N_{\text{sim}}} N_{\text{call}}^{(j)}$. These values are obtained per dimension since $N_{\text{sim}}$ runs of BUS-SuS are performed at each specific $k = 1, \ldots, k_{\text{max}}$. In tBUS-SuS, the total number of calls using the step-wise sampler is $N_{\text{call},\text{tBUS}}^{(k)} = \sum_{j=1}^{N_{\text{sim}}} N_{\text{call},a}^{(j)}$, or with MwG is $N_{\text{call},\text{tBUS}}^{(k)} = \sum_{j=1}^{N_{\text{sim}}} N_{\text{call},b}^{(j)}$. This cost is associated to the whole dimension spectrum and not to each individual $k$, thus the value is re-distributed among dimensions by applying the model posterior.

**References**

reversible jump MCMC algorithm for random effects models, Computational Statistics Data Analysis 98 (2016) 79–90.


