



# Bayesian model inversion using stochastic spectral embedding

P.-R Wagner, S. Marelli, B. Sudret

## ► To cite this version:

P.-R Wagner, S. Marelli, B. Sudret. Bayesian model inversion using stochastic spectral embedding. 2020. hal-02612455

**HAL Id: hal-02612455**

**<https://hal.science/hal-02612455>**

Preprint submitted on 19 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# BAYESIAN MODEL INVERSION USING STOCHASTIC SPECTRAL EMBEDDING

P.-R. Wagner, S. Marelli and B. Sudret



## Data Sheet

---

<b>Journal:</b>	-
<b>Report Ref.:</b>	RSUQ-2020-005
<b>Arxiv Ref.:</b>	<a href="https://arxiv.org/abs/2005.07380">https://arxiv.org/abs/2005.07380</a> [stat.CO, stat.ME]
<b>DOI:</b>	-
<b>Date submitted:</b>	May 14th, 2020
<b>Date accepted:</b>	-

---

# Bayesian model inversion using stochastic spectral embedding

P.-R. Wagner, S. Marelli, B. Sudret

## Abstract

In this paper we propose a new sampling-free approach to solve Bayesian inverse problems that extends the recently introduced spectral likelihood expansions (SLE) method. The latter solves the inverse problem by expanding the likelihood function onto a global polynomial basis orthogonal w.r.t. the prior distribution. This gives rise to analytical expressions for key statistics of the Bayesian posterior distribution, such as evidence, posterior moments and posterior marginals by simple post-processing of the expansion coefficients. It is well known that in most practically relevant scenarios, likelihood functions have close-to-compact support, which causes the global SLE approach to fail due to the high polynomial degree required for an accurate spectral representation. To solve this problem, we herein replace the global polynomial expansion from SLE with a recently proposed method for local spectral expansion refinement called stochastic spectral embedding (SSE). This surrogate-modeling method was developed for functions with high local complexity. To increase the efficiency of SSE, we enhance it with an adaptive sample enrichment scheme. We show that SSE works well for likelihood approximations and retains the relevant spectral properties of SLE, thus preserving analytical expressions of posterior statistics. To assess the performance of our approach, we include three case studies ranging from low to high dimensional model inversion problems that showcase the superiority of the SSE approach compared to SLE and present the approach as a promising alternative to existing inversion frameworks.

**Keywords:** Bayesian model inversion, inverse problems, polynomial chaos expansions, spectral likelihood expansions, sampling-free inversion.

## 1 Introduction

Computational models are an invaluable tool for decision making, scientific advances and engineering breakthroughs. They establish a connection between a set of input parameters and output quantities with wide-ranging applications. *Model inversion* uses available experimental observations of the output to determine the set of input parameters that maximize the predictive potential of a model. The importance of efficient and reliable model inversion frameworks can hardly be overstated, considering that they establish a direct connection

between models and the real world. Without it, the most advanced model predictions might lack physical meaning and, consequently, be useless for their intended applications.

Bayesian model inversion is one way to formalize this problem (Jaynes, 2003; Gelman et al., 2014). It is based on Bayesian inference and poses the problem in a probabilistic setting by capitalizing on Bayes' theorem. In this setting a so-called *prior* (*i.e.*, before observations) probability distribution about the model parameters is updated to a so-called *posterior* (*i.e.*, after observations) distribution. The posterior distribution is the probability distribution of the input parameters conditioned on the available observations, and the main outcome of the Bayesian inversion process.

In Bayesian model inversion, the connection between the model output and the observations is established through a probabilistic *discrepancy model*. This model, which is a function of the input parameters, leads to the so-called *likelihood* function. The specific form of the likelihood function depends on the problem at hand, but typically it has a global maximum for the input parameters with the model output that is closest to the available observations (*w.r.t.* some metric), and rapidly goes to zero with increasing distance to those parameters.

Analytical expressions for the posterior distribution can only be found in few academic examples (*e.g.*, conjugate priors with a linear forward model, Bishop (2006); Gelman et al. (2014)). In general model inversion problems, such analytical solutions are not available though. Instead, it is common practice to resort to sampling methods to generate a sample distributed according to the posterior distribution. The family of *Markov chain Monte Carlo* (MCMC) algorithms are particularly suitable for generating such a *posterior sample* (Beck and Au, 2002; Robert and Casella, 2004).

While MCMC and its extensions are extensively used in model inversion, and new algorithms are continuously being developed (Haario et al., 2001; Ching and Chen, 2007; Goodman and Weare, 2010; Neal, 2011), it has a few notable shortcomings that hinder its application in many practical cases. It is well known that there are no robust convergence criteria for MCMC algorithms, and that their performance is particularly sensitive to their tuning parameters. Additionally, samples generated by MCMC algorithms are often highly correlated, thus requiring extensive heuristic post-processing and empirical rules (Gelman and Rubin, 1992; Brooks and Gelman, 1998). MCMC algorithms are also in general not well suited for sampling multimodal posterior distributions.

When considering complex engineering scenarios, the models subject to inversion are often computationally expensive. Because MCMC algorithms usually require a significant number of forward model evaluations, it has been proposed to accelerate the procedure by using *surrogate models* in lieu of the original models. These surrogate models are either constructed non-adaptively before sampling from the posterior distribution (Marzouk et al., 2007; Marzouk and Xiu, 2009) or adaptively during the sampling procedure (Li and Marzouk,

2014; Birolleau et al., 2014). Adaptive techniques can be of great benefit with posterior distributions that are concentrated in a small subspace of the prior domain, as the surrogate only needs to be accurate near high density areas of the posterior distribution.

*Polynomial chaos expansions* (PCE) are a widely used surrogate modelling technique based on expanding the forward model onto a suitable polynomial basis (Ghanem and Spanos, 1991; Xiu and Karniadakis, 2002). In other words, it provides a spectral representation of the computational forward model. Thanks to the introduction of sparse regression (see, e.g. Blatman and Sudret (2011)), its computation has become feasible even in the presence of complex and computationally expensive engineering models. This technique has been successfully used in conjunction with MCMC to reduce the total computational costs associated with sampling from the posterior distribution (Marzouk et al., 2007; Wagner et al., 2020).

Alternative approaches to compute the posterior distribution or its statistics include the *Laplace approximation* at a posterior mode (Tierney and Kadane, 1986; Tierney et al., 1989b,a), approximate Bayesian computations (ABC) (Marin et al., 2012; Sisson et al., 2018), optimal transport approaches (El Moselhy and Marzouk, 2012; Parno, 2015; Marzouk et al., 2016) and embarrassingly parallel quasi-Monte Carlo sampling (Dick et al., 2017; Gantner and Peters, 2018).

Recently, Nagel and Sudret (2016) proposed *spectral likelihood expansions* (SLE), a novel approach that is closely related to PCE and aims at finding a spectral expansion of the likelihood function in an orthogonal basis *w.r.t.* the prior density function. The advantage of this method is that it provides analytical expressions for the posterior marginals and general posterior moments by post-processing the expansion coefficients. However, the typically compact support of likelihood functions generally prevents a sparse functional representation.

*Stochastic spectral embedding* (SSE) is a metamodeling technique suitable for approximating functions with complex localized features recently developed in Marelli et al. (2020). In this paper we show how this technique, when extended with adaptive sample enrichment, is effective for the approximation of likelihood functions in Bayesian model inversion problems. Furthermore, we show that due to the construction of the basis polynomials, SSE preserves the most important post-processing properties of SLE.

The paper is organized as follows: In Section 2 we establish the basics of Bayesian inference and particularly Bayesian model inversion. We then give an introduction into spectral function decomposition with a focus on polynomial chaos expansions and their application to likelihood functions (SLE) in Section 3. In Section 4 we present the main contribution of the paper, namely the derivation of Bayesian posterior quantities of interest when SSE is applied to likelihood functions and the extension of the SSE algorithm with an adaptive sampling strategy. Finally, in Section 5 we showcase the performance of our approach on three case studies of varying complexity.

## 2 Model inversion

The problem of *model inversion* occurs whenever the predictions of a model are to be brought into agreement with available observations or *data*. This is achieved by properly adjusting a set of *input parameters* of the model. The goal of inversion can be twofold: on the one hand the inferred input parameters might be used to predict new realizations of the model output. On the other hand, the inferred input parameters might be the main interest. Model inversion is a common problem in many engineering disciplines, that in some cases is still routinely solved manually, *i.e.* by simply changing the input parameters until some, often qualitative, *goodness-of-fit* criterion is met. More quantitative inversion approaches aim at automatizing this process, by establishing a metric (*e.g.*,  $L^2$ -distance) between the data and the model response, which is then minimized through suitable optimization algorithms.

While such approaches can often be used in practical applications, they tend not to provide measures of the *uncertainties* associated with the inferred model input or predictions. These uncertainties are useful in judging the accuracy of the inversion, as well as indicating non-informative measurements. In fact, the lack of uncertainty quantification in the context of model inversion can lead to erroneous results that have far-reaching consequences in subsequent applications. One approach to consider uncertainties in inverse problems is the *Bayesian framework* for model inversion that will be presented hereinafter.

### 2.1 Bayesian inference

Consider some *non-observable* parameters  $\mathbf{X} \in \mathcal{D}_{\mathbf{X}}$  and the *observables*  $\mathbf{Y} \in \mathcal{D}_{\mathbf{Y}}$ . Furthermore, let  $\mathcal{Y} = \{\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(N)}\}$  be a set of  $N$  measurements, *i.e.*, noisy observations of a set of realizations of  $\mathbf{Y}$ . Statistical inference consists in drawing conclusions about  $\mathbf{X}$  using the information from  $\mathcal{Y}$  (Gelman et al., 2014). These measurements can be direct observations of the parameters ( $\mathbf{Y} = \mathbf{X}$ ) or some quantities indirectly related to  $\mathbf{X}$  through a function or *model*  $\mathcal{M} : \mathcal{D}_{\mathbf{X}} \rightarrow \mathcal{D}_{\mathbf{Y}}$ . One way to conduct this inference is through Bayes' theorem of conditional probabilities, a process known as *Bayesian inference*.

Denoting by  $\pi(\cdot)$  a probability density function (PDF) and by  $\pi(\cdot|\mathbf{x})$  a PDF conditioned on  $\mathbf{x}$ , Bayes' theorem can be written as

$$\pi(\mathbf{x}|\mathcal{Y}) = \frac{\pi(\mathcal{Y}|\mathbf{x})\pi(\mathbf{x})}{\pi(\mathcal{Y})}, \quad (1)$$

where  $\pi(\mathbf{x})$  is known as the *prior* distribution of the parameters, *i.e.*, the distribution of  $\mathbf{X}$  before observing the data  $\mathcal{Y}$ . The conditional distribution  $\pi(\mathcal{Y}|\mathbf{x})$ , known as *likelihood*, establishes a connection between the observations  $\mathcal{Y}$  and a realization of the parameters  $\mathbf{X} = \mathbf{x}$ . For a given realization  $\mathbf{x}$ , it returns the probability density of observing the data  $\mathcal{Y}$ . Under the common assumption of *independence* between individual observations,  $\{\mathbf{y}^{(i)}\}_{i=1}^N$ ,

the likelihood function takes the form:

$$\mathcal{L} : \mathbf{x} \mapsto \mathcal{L}(\mathbf{x}; \mathcal{Y}) \stackrel{\text{def}}{=} \pi(\mathcal{Y}|\mathbf{x}) = \prod_{i=1}^N \pi(\mathbf{y}^{(i)}|\mathbf{x}). \quad (2)$$

The likelihood function is a map  $\mathcal{D}_{\mathbf{X}} \rightarrow \mathbb{R}_+$ , and it attains its maximum for the parameter set with the highest probability of yielding  $\mathcal{Y}$ . With this, Bayes' theorem from Eq. (1) can be rewritten as:

$$\pi(\mathbf{x}|\mathcal{Y}) = \frac{\mathcal{L}(\mathbf{x}; \mathcal{Y})\pi(\mathbf{x})}{Z}, \quad \text{with} \quad Z = \int_{\mathcal{D}_{\mathbf{X}}} \mathcal{L}(\mathbf{x}; \mathcal{Y})\pi(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \quad (3)$$

where  $Z$  is a normalizing constant often called *evidence* or *marginal likelihood*. On the left-hand side,  $\pi(\mathbf{x}|\mathcal{Y})$  is the *posterior* PDF, *i.e.*, the distribution of  $\mathbf{X}$  after observing data  $\mathcal{Y}$ . In this sense, Bayes' theorem establishes a general expression for updating the prior distribution using a likelihood function to incorporate information from the data.

## 2.2 Bayesian model inversion

*Bayesian model inversion* describes the application of the Bayesian inference framework to the problem of model inversion (Beck and Katafygiotis, 1998; Kennedy and O'Hagan, 2001; Jaynes, 2003; Tarantola, 2005). The two main ingredients needed to infer model parameters within the Bayesian framework are a prior distribution  $\pi(\mathbf{x})$  of the model parameters and a likelihood function  $\mathcal{L}$ . In practical applications, *prior* information about the model parameters is often readily available. Typical sources of such information are physical parameter constraints or expert knowledge. Additionally, prior inversion attempts can serve as guidelines to assign *informative* prior distributions. In cases where no prior information about the parameters is available, so-called *non-informative* or *invariant* prior distributions (Jeffreys, 1946; Harney, 2016) can also be assigned. The likelihood function serves instead as the link between model parameters  $\mathbf{X}$  and observations of the model output  $\mathcal{Y}$ . To connect these two quantities, it is necessary to choose a so-called *discrepancy model* that gives the relative probability that the model response to a realization of  $\mathbf{X} = \mathbf{x}$  describes the observations. One common assumption for this probabilistic model is that the measurements are perturbed by a Gaussian additive discrepancy term  $\mathbf{E} \sim \mathcal{N}(\boldsymbol{\varepsilon}|\mathbf{0}, \boldsymbol{\Sigma})$ , with covariance matrix  $\boldsymbol{\Sigma}$ . For a single measurement  $\mathbf{y}^{(i)}$  it reads:

$$\mathbf{y}^{(i)} = \mathcal{M}(\mathbf{x}) + \boldsymbol{\varepsilon}. \quad (4)$$

This discrepancy between the model output  $\mathcal{M}(\mathbf{X})$  and the observables  $\mathbf{Y}$  can result from *measurement error* or *model inadequacies*. By using this additive discrepancy model, the distribution of the observables conditioned on the parameters  $\mathbf{Y}|\mathbf{x}$  is written as:

$$\pi(\mathbf{y}^{(i)}|\mathbf{x}) = \mathcal{N}(\mathbf{y}^{(i)}|\mathcal{M}(\mathbf{x}), \boldsymbol{\Sigma}), \quad (5)$$

where  $\mathcal{N}(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes the multivariate Gaussian PDF with mean value  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . The likelihood function  $\mathcal{L}$  is then constructed using this probabilistic model



$\pi(\mathbf{y}^{(i)}|\mathbf{x})$  and Eq. (2). For a given set of measurements  $\mathcal{Y}$  it thus reads:

$$\mathcal{L}(\mathbf{x}; \mathcal{Y}) \stackrel{\text{def}}{=} \prod_{i=1}^N \mathcal{N}(\mathbf{y}^{(i)}|\mathcal{M}(\mathbf{x}), \Sigma). \quad (6)$$

With the fully specified Bayesian model inversion problem, Eq. (3) directly gives the posterior distribution of the model parameters  $\pi(\mathbf{x}|\mathcal{Y})$ . In the setting of model inversion, the posterior distribution represents therefore the state of belief about the true *data-generating* model parameters, considering all available information: computational forward model, discrepancy model and measurement data (Beck and Katafygiotis, 1998; Jaynes, 2003).

Often, the ultimate goal of model inversion is to provide a set of inferred parameters, with associated confidence measures/intervals. This is often achieved by computing posterior statistics (*e.g.*, moments, mode, etc.). Propagating the posterior through secondary models is also of interest. So-called *quantities of interest* (QoI) can be expressed by calculating the posterior expectation of suitable functions of the parameters  $h(\mathbf{x}) : \mathbb{R}^M \rightarrow \mathbb{R}$ , with  $\mathbf{X}|\mathcal{Y} \sim \pi(\mathbf{x}|\mathcal{Y})$ , as in:

$$\mathbb{E}[h(\mathbf{X})|\mathcal{Y}] = \int_{\mathcal{D}_{\mathbf{X}|\mathcal{Y}}} h(\mathbf{x})\pi(\mathbf{x}|\mathcal{Y}) d\mathbf{x}. \quad (7)$$

Depending on  $h$ , this formulation encompasses *posterior moments* ( $h(\mathbf{x}) = x_i$  or  $h(\mathbf{x}) = (x_i - \mathbb{E}[X_i])^2$  for the first and second moments, respectively), *posterior covariance* ( $h(\mathbf{x}) = x_i x_j - \mathbb{E}[X_i]\mathbb{E}[X_j]$ ) or expectations of secondary models ( $h(\mathbf{x}) = \mathcal{M}^*(\mathbf{x})$ ).

### 3 Spectral function decomposition

To pose a Bayesian inversion problem, the specification of a prior distribution and a likelihood function described in the previous section is sufficient. Its solution, however, is not available in closed form in the general case.

*Spectral likelihood expansion* (SLE) is a recently proposed method that aims at solving the Bayesian inversion problem by finding a *polynomial chaos expansion* (PCE) of the likelihood function in a basis orthogonal *w.r.t.* the prior distribution (Nagel and Sudret, 2016). This representation allows one to derive analytical expressions for the evidence  $Z$ , the posterior distribution, the posterior marginals, and many types of QoIs, including the posterior moments.

We offer here a brief introduction to *regression-based, sparse PCE* before introducing SLE, but refer the interested reader to more exhaustive resources on PCE (Ghanem and Spanos, 1991; Xiu and Karniadakis, 2002) and sparse PCE (Xiu, 2010; Blatman and Sudret, 2010, 2011).

Let us consider a random variable  $\mathbf{X}$  with *independent* components  $\{X_i, i = 1, \dots, M\}$  and associated probability density functions  $\pi_i(x_i)$  so that  $\pi(\mathbf{x}) = \prod_{i=1}^M \pi_i(x_i)$ . Assume further that  $\mathcal{M} : \mathcal{D}_{\mathbf{X}} = \prod_{i=1}^M \mathcal{D}_{X_i} \subseteq \mathbb{R}^M \rightarrow \mathbb{R}$  is a scalar function of  $\mathbf{X}$  which fulfills the *finite variance* condition ( $\mathbb{E}[\mathcal{M}(\mathbf{X})^2] < +\infty$ ). Then it is possible to find a so-called

truncated polynomial chaos approximation of  $\mathcal{M}$  such that

$$\mathcal{M}(\mathbf{X}) \approx \mathcal{M}_{\text{PCE}}(\mathbf{X}) \stackrel{\text{def}}{=} \sum_{\alpha \in \mathcal{A}} a_{\alpha} \Psi_{\alpha}(\mathbf{X}) \quad (8)$$

where  $\alpha$  is an  $M$ -tuple  $(\alpha_1, \dots, \alpha_M) \in \mathbb{N}^M$  and  $\mathcal{A} \subset \mathbb{N}^M$ . For most parametric distributions, well-known classical orthonormal polynomials  $\{\Psi_{\alpha}\}_{\alpha \in \mathbb{N}^M}$  satisfy the necessary orthonormality condition *w.r.t.*  $\pi(\mathbf{x})$  (Xiu and Karniadakis, 2002). For more general distributions, arbitrary orthonormal polynomials can be constructed numerically through the Stieltjes procedure (Gautschi, 2004). If additionally,  $\mathcal{A}$  is a sparse subset of  $\mathbb{N}^M$ , the truncated expansion in Eq. (8) is called a *sparse PCE*.

There exist different algorithms to produce a sparse PCE in practice, *i.e.* select a sparse basis  $\mathcal{A}$  and compute the corresponding coefficients. A powerful class of methods are *regression-based* approaches that rely on an initial input sample  $\mathcal{X}$ , called experimental design, and corresponding model evaluations  $\mathcal{M}(\mathcal{X})$  (See, *e.g.* Lüthen et al. (2020) for a recent survey). Additionally, it is possible to design adaptive algorithms that choose the truncated basis size (Blatman and Sudret, 2011; Jakeman et al., 2015).

To assess the accuracy of PCE, the so-called *generalization error*  $\mathbb{E}[(\mathcal{M}(\mathbf{X}) - \mathcal{M}_{\text{PCE}}(\mathbf{X}))^2]$  shall be evaluated. A robust generalization error estimator is given by the *leave-one-out* (LOO) cross validation technique. This estimator is obtained by

$$\varepsilon_{\text{LOO}} = \frac{1}{K} \sum_{i=1}^K \left( \mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}_{\text{PCE}}^i(\mathbf{x}^{(i)}) \right)^2, \quad (9)$$

where  $\mathcal{M}_{\text{PCE}}^i$  is constructed by leaving out the  $i$ -th point from the experimental design  $\mathcal{X}$ . For methods based on linear regression, it can be shown (Chapelle et al., 2002; Blatman and Sudret, 2010) that the LOO error is available analytically by post-processing the regressor matrix.

### 3.1 Spectral likelihood expansions

The idea of SLE is to use sparse PCE to find a spectral representation of the likelihood function  $\mathcal{L}$  occurring in Bayesian model inversion problems (see Eq. (2)). We present here a brief introduction to the method and the main results of Nagel and Sudret (2016).

Likelihood functions can be seen as scalar functions of the input random vector  $\mathbf{X} \sim \pi(\mathbf{x})$ . In this work we assume priors of the type  $\pi(\mathbf{x}) = \prod_{i=1}^M \pi_i(x_i)$ , *i.e.* with independent marginals. Additionally, likelihood functions fulfill the finite variance condition (Nagel and Sudret, 2016) and therefore admit a spectral decomposition:

$$\mathcal{L}(\mathbf{X}) \approx \mathcal{L}_{\text{SLE}}(\mathbf{X}) \stackrel{\text{def}}{=} \sum_{\alpha \in \mathcal{A}} a_{\alpha} \Psi_{\alpha}(\mathbf{X}), \quad (10)$$

where the explicit dependence on  $\mathcal{Y}$  was dropped for notational simplicity.

Upon computing the basis and coefficients in Eq. (8), the solution to the inverse problem is converted to merely post-processing the coefficients  $a_{\alpha}$ . The following expressions can be derived for the individual quantities:

**Evidence** The evidence emerges as the constant polynomial's coefficient  $a_0$

$$Z = \int_{\mathcal{D}_{\mathbf{X}}} \mathcal{L}(\mathbf{x}) \pi(\mathbf{x}) d\mathbf{x} \approx \langle \mathcal{L}_{\text{SLE}}, 1 \rangle_{\pi} = a_0. \quad (11)$$

**Posterior** Upon computing the evidence  $Z$ , the posterior can be evaluated directly through

$$\pi(\mathbf{x}|\mathcal{Y}) \approx \frac{\mathcal{L}_{\text{SLE}}(\mathbf{x}) \pi(\mathbf{x})}{Z} = \frac{\pi(\mathbf{x})}{a_0} \sum_{\alpha \in \mathcal{A}} a_{\alpha} \Psi_{\alpha}(\mathbf{x}). \quad (12)$$

**Posterior marginals** Let  $\mathbf{u}$  and  $\mathbf{v}$  be two non-empty disjoint index sets such that  $\mathbf{u} \cup \mathbf{v} = \{1, \dots, M\}$ . We split the random vector  $\mathbf{X}$  into two vectors  $\mathbf{X}_{\mathbf{u}}$  with components  $\{X_i\}_{i \in \mathbf{u}} \in \mathcal{D}_{\mathbf{X}_{\mathbf{u}}}$  and  $\mathbf{X}_{\mathbf{v}}$  with components  $\{X_i\}_{i \in \mathbf{v}} \in \mathcal{D}_{\mathbf{X}_{\mathbf{v}}}$ . Denote further by  $\pi_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) \stackrel{\text{def}}{=} \prod_{i \in \mathbf{u}} \pi_i(x_i)$  and  $\pi_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}}) \stackrel{\text{def}}{=} \prod_{i \in \mathbf{v}} \pi_i(x_i)$  the prior marginal density functions of  $\mathbf{X}_{\mathbf{u}}$  and  $\mathbf{X}_{\mathbf{v}}$  respectively. The posterior marginals then read:

$$\pi_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}|\mathcal{Y}) = \int_{\mathcal{D}_{\mathbf{X}_{\mathbf{v}}}} \pi(\mathbf{x}|\mathcal{Y}) d\mathbf{x}_{\mathbf{v}} \approx \frac{\pi_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})}{a_0} \sum_{\alpha \in \mathcal{A}_{\mathbf{v}=0}} a_{\alpha} \Psi_{\alpha}(\mathbf{x}_{\mathbf{u}}), \quad (13)$$

where  $\mathcal{A}_{\mathbf{v}=0} = \{\alpha \in \mathcal{A} : \alpha_i = 0 \Leftrightarrow i \in \mathbf{v}\}$ . The series in the above equation constitutes a subexpansion that contains non-constant polynomials only in the directions  $i \in \mathbf{u}$ .

**Quantities of interest** Finally, it is also possible to analytically compute posterior expectations of functions that admit a polynomial chaos expansion in the same basis of the form  $h(\mathbf{X}) \approx \sum_{\alpha \in \mathcal{A}} b_{\alpha} \Psi_{\alpha}(\mathbf{X})$ . Eq. (7) then reduces to the spectral product:

$$\mathbb{E}[h(\mathbf{X})|\mathcal{Y}] = \frac{1}{a_0} \sum_{\alpha \in \mathcal{A}} a_{\alpha} b_{\alpha}. \quad (14)$$

The quality of these results depends only on the approximation error introduced in Eq. (10). The latter, in turn, depends mainly on the chosen PCE truncation strategy (Blatman and Sudret, 2011; Nagel and Sudret, 2016) and the number of points used to compute the coefficients (*i.e.*, the experimental design). It is known that likelihood functions typically have quasi-compact supports (*i.e.*,  $\mathcal{L}(\mathbf{X}) \approx 0$  on a majority of  $\mathcal{D}_{\mathbf{X}}$ ). Such functions require a very high polynomial degree to be approximated accurately, which in turn can lead to the need for prohibitively large experimental designs.

## 4 Stochastic spectral embedding

Stochastic spectral embedding (SSE) is a multi-level approach to surrogate modeling originally proposed in Marelli et al. (2020). It attempts to approximate a given square-integrable function  $\mathcal{M}$  through

$$\mathcal{M} \approx \mathcal{M}_{\text{SSE}}(\mathbf{X}) = \sum_{k \in \mathcal{K}} \mathbf{1}_{\mathcal{D}_{\mathbf{X}}^k}(\mathbf{X}) \hat{\mathcal{R}}_S^k(\mathbf{X}), \quad (15)$$

where  $\mathcal{K} \subseteq \mathbb{N}^2$  is a set of multi-indices with elements  $k = (\ell, p)$  for which  $\ell = 0, \dots, L$  and  $p = 1, \dots, P_{\ell}$  where  $L$  is the number of levels and  $P_{\ell}$  is the number of subdomains at a

specific level  $\ell$ . We call  $\hat{\mathcal{R}}_S^k(\mathbf{X})$  a *residual expansion* given by

$$\hat{\mathcal{R}}_S^k(\mathbf{X}) = \sum_{\alpha \in \mathcal{A}^k} a_\alpha^k \Psi_\alpha^k(\mathbf{X}). \quad (16)$$

In the present paper the term  $\sum_{j \in \mathcal{A}^k} a_j^k \Psi_j^k(\mathbf{X})$  denotes a polynomial chaos expansion (see Eq. (8)) constructed in the subdomain  $\mathcal{D}_{\mathbf{X}}^k$ , but in principle it can refer to any spectral expansion (*e.g.*, Fourier series). A schematic representation of the summation in Eq. (15) is given in Figure 2. The detailed notation and the algorithm to sequentially construct an SSE are given in the sequel.

#### 4.1 SSE for Bayesian model inversion

Viewing the likelihood as a function of a random variable  $\mathbf{X}$ , we can directly use Eq. (15) to write down its SSE representation

$$\mathcal{L}(\mathbf{X}) \approx \mathcal{L}_{\text{SSE}}(\mathbf{X}) \stackrel{\text{def}}{=} \sum_{k \in \mathcal{K}} \mathbf{1}_{\mathcal{D}_{\mathbf{X}}^k}(\mathbf{X}) \hat{\mathcal{R}}_S^k(\mathbf{X}), \quad (17)$$

where the variable  $\mathbf{X}$  is distributed according to the prior distribution  $\pi(\mathbf{x})$  and, consequently, the local basis used to compute  $\hat{\mathcal{R}}_S^k(\mathbf{X})$  is orthonormal *w.r.t.* that distribution.

Due to the local spectral properties of the residual expansions, the SSE representation of the likelihood function retains all of the post-processing properties of SLE (Section 3.1):

**Evidence** The normalization constant  $Z$  emerges as the sum of the constant polynomial coefficients weighted by the prior mass:

$$Z = \sum_{k \in \mathcal{K}} \sum_{\alpha \in \mathcal{A}^k} a_\alpha^k \int_{\mathcal{D}_{\mathbf{X}}^k} \Psi_\alpha^k(\mathbf{x}) \pi(\mathbf{x}) d\mathbf{x} = \sum_{k \in \mathcal{K}} \mathcal{V}^k a_{\mathbf{0}}^k, \quad \text{where} \quad \mathcal{V}^k = \int_{\mathcal{D}_{\mathbf{X}}^k} \pi(\mathbf{x}) d\mathbf{x}. \quad (18)$$

**Posterior** This allows us to write the posterior as

$$\pi(\mathbf{x}|\mathcal{Y}) \approx \frac{\mathcal{L}_{\text{SSE}}(\mathbf{x})\pi(\mathbf{x})}{Z} = \frac{\pi(\mathbf{x})}{\sum_{k \in \mathcal{K}} \mathcal{V}^k a_{\mathbf{0}}^k} \sum_{k \in \mathcal{K}} \mathbf{1}_{\mathcal{D}_{\mathbf{X}}^k}(\mathbf{x}) \hat{\mathcal{R}}_S^k(\mathbf{x}). \quad (19)$$

**Posterior marginal** Utilizing again the disjoint sets  $\mathbf{u}$  and  $\mathbf{v}$  from Eq. (13) it is also possible to analytically derive posterior marginal PDFs as

$$\pi_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}|\mathcal{Y}) = \int_{\mathcal{D}_{\mathbf{X}_{\mathbf{v}}}} \pi(\mathbf{x}|\mathcal{Y}) d\mathbf{x}_{\mathbf{v}} \approx \frac{\pi_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})}{\sum_{k \in \mathcal{K}} \mathcal{V}^k a_{\mathbf{0}}^k} \sum_{k \in \mathcal{K}} \mathbf{1}_{\mathcal{D}_{\mathbf{X}_{\mathbf{u}}}^k}(\mathbf{x}_{\mathbf{u}}) \hat{\mathcal{R}}_{S,\mathbf{u}}^k(\mathbf{x}_{\mathbf{u}}) \mathcal{V}_{\mathbf{v}}^k \quad (20)$$

where

$$\hat{\mathcal{R}}_{S,\mathbf{u}}^k(\mathbf{x}_{\mathbf{u}}) = \sum_{\alpha \in \mathcal{A}_{\mathbf{v}=0}^k} a_\alpha^k \Psi_\alpha^k(\mathbf{x}_{\mathbf{u}}) \quad \text{and} \quad \mathcal{V}_{\mathbf{v}}^k = \int_{\mathcal{D}_{\mathbf{X}_{\mathbf{v}}}^k} \pi_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}}) d\mathbf{x}_{\mathbf{v}}. \quad (21)$$

$\hat{\mathcal{R}}_{S,\mathbf{u}}^k(\mathbf{x}_{\mathbf{u}})$  is a subexpansion of  $\hat{\mathcal{R}}_S^k(\mathbf{x})$  that contains only non-constant polynomials in the directions  $i \in \mathbf{u}$ . Note that, as we assumed that the prior distribution has independent components, the constants  $\mathcal{V}^k$  and  $\mathcal{V}_{\mathbf{v}}^k$  are obtained as products of univariate integrals which are available analytically from the prior marginal cumulative distribution functions (CDFs).

**Quantities of interest** Expected values of function  $h(\mathbf{x}) = \sum_{\alpha \in \mathcal{A}^k} b_{\alpha}^k \Psi_{\alpha}^k(\mathbf{x})$  for  $k \in \mathcal{K}$  under the posterior can be approximated by:

$$\begin{aligned} \mathbb{E}[h(\mathbf{X})|\mathcal{Y}] &= \int_{\mathcal{D}_{\mathbf{X}}} h(\mathbf{x}) \pi(\mathbf{x}|\mathcal{Y}) d\mathbf{x} \\ &= \frac{1}{Z} \sum_{k \in \mathcal{K}} \sum_{\alpha \in \mathcal{A}^k} a_{\alpha}^k \int_{\mathcal{D}_{\mathbf{X}}} h(\mathbf{x}) \Psi_{\alpha}^k(\mathbf{x}) \pi(\mathbf{x}) d\mathbf{x} \\ &= \frac{1}{Z} \sum_{k \in \mathcal{K}} \sum_{\alpha \in \mathcal{A}^k} a_{\alpha}^k b_{\alpha}^k, \end{aligned} \quad (22)$$

where  $b_{\alpha}^k$  are the coefficients of the PCE of  $h$  in the  $\text{card}(\mathcal{K})$  bases  $\{\Psi_{\alpha}^k\}_{\alpha \in \mathcal{A}^k}$ . This can also be used for computing posterior moments.

These expressions can be seen as a generalization of the ones for SLE detailed in Section 3.1. For a single-level global expansion (*i.e.*,  $\text{card}(\mathcal{K}) = 1$ ) and consequently  $\mathcal{V}^{(0,1)} = 1$ , they are identical.

## 4.2 Modifications to the original algorithm

The original algorithm for computing an SSE was presented in Marelli et al. (2020). It recursively partitions the input domain  $\mathcal{D}_{\mathbf{X}}$  and constructs truncated expansions of the residual. We reproduce it below for reference but replace the model  $\mathcal{M}$  with the likelihood function  $\mathcal{L}$ . We further simplify the algorithm by choosing a partitioning strategy with  $N_S = 2$ .

1. **Initialization:**
  - (a)  $\ell = 0, p = 1$
  - (b)  $\mathcal{D}_{\mathbf{X}}^{\ell,p} = \mathcal{D}_{\mathbf{X}}$
  - (c)  $\mathcal{R}^{\ell}(\mathbf{X}) = \mathcal{L}(\mathbf{X})$
2. **For each subdomain  $\mathcal{D}_{\mathbf{X}}^{\ell,p}, p = 1, \dots, P_{\mathbf{X}}^{\ell}$ :**
  - (a) Calculate the truncated expansion  $\widehat{\mathcal{R}}_S^{\ell,p}(\mathbf{X}^{\ell,p})$  of the residual  $\mathcal{R}^{\ell}(\mathbf{X}^{\ell,p})$  in the current subdomain
  - (b) Update the residual in the current subdomain  $\mathcal{R}^{\ell+1}(\mathbf{X}^{\ell,p}) = \mathcal{R}^{\ell}(\mathbf{X}^{\ell,p}) - \widehat{\mathcal{R}}_S^{\ell,p}(\mathbf{X}^{\ell,p})$
  - (c) Split the current subdomain  $\mathcal{D}_{\mathbf{X}}^{\ell,p}$  in 2 subdomains  $\mathcal{D}_{\mathbf{X}}^{\ell+1,\{s_1,s_2\}}$  based on a partitioning strategy
  - (d) If  $\ell < L$ ,  $\ell \leftarrow \ell + 1$ , go back to 2a, otherwise terminate the algorithm
3. **Termination**
  - (a) Return the full sequence of  $\mathcal{D}_{\mathbf{X}}^{\ell,p}$  and  $\widehat{\mathcal{R}}_S^{\ell,p}(\mathbf{X}^{\ell,p})$  needed to compute Eq. (15).

In practice, the residual expansions  $\widehat{\mathcal{R}}_S^{\ell,p}(\mathbf{X}^{\ell,p})$  are computed using a *fixed experimental design*  $\mathcal{X}$  and corresponding model evaluations  $\mathcal{L}(\mathcal{X})$ . The algorithm then only requires the specification of a partitioning strategy and a termination criterion, as detailed in Marelli et al. (2020).

Likelihood functions are typically characterized by a localized behaviour: Close to the data-generating parameters they peak while quickly decaying to 0 in the remainder of the

prior domain. This means that in a majority of the domain the likelihood evaluation is non-informative. Directly applying the original algorithm is then expected to *waste* many likelihood evaluations.

We therefore modify the original algorithm by adding an adaptive sampling scheme (Section 4.2.1) that includes the termination criterion and introducing an improved partitioning strategy (Section 4.2.2) that is especially suitable for finding compact support features. The rationale for these modifications is presented next.

#### 4.2.1 Adaptive sampling scheme

The proposed algorithm has two parameters: the experimental design size for the residual expansions  $N_{\text{ref}}$  and the final experimental design size corresponding to the available computational budget  $N_{\text{ED}}$ . At the initialization of the algorithm  $N_{\text{ref}}$  points are sampled as a first experimental design. At every further iteration, additional points are then sampled from the prior distribution. These samples are generated in a space-filling way (*e.g.* through latin hypercube sampling) in the newly created subdomains  $\mathcal{D}_{\mathbf{X}}^{\ell+1,s}$  to have always *exactly*  $N_{\text{ref}}$  points available for constructing the residual expansions. The algorithm is terminated, once the computational budget  $N_{\text{ED}}$  has been exhausted.

At every step, the proposed algorithm chooses a single *refinement domain* from the set of unsplit, *i.e.* *terminal domains*, creates two new subdomains by splitting the refinement domain and constructs residual expansions after enriching the experimental design. The selection of this refinement domain is based on the error estimator  $\mathcal{E}^k$  that is defined by

$$\mathcal{E}^{\ell+1,s} = \begin{cases} E_{\text{LOO}}^{\ell+1,s} \mathcal{V}^{\ell+1,s}, & \text{if } \exists \hat{\mathcal{R}}_S^{\ell+1,s}, \\ E_{\text{LOO}}^{\ell,s} \mathcal{V}^{\ell+1,s}, & \text{otherwise.} \end{cases} \quad (23)$$

This estimator incorporates the subdomain size through the prior mass  $\mathcal{V}^{\ell+1,s}$ , and the approximation accuracy, through the leave-one-out estimator. The distinction is necessary to assign an error estimator also to domains that have too few points to construct a residual expansion, in which case the error estimator of the previous level  $E_{\text{LOO}}^{\ell,s}$  is reused.

The algorithm sequentially splits and refines subdomains with large approximation errors. Because likelihood functions typically have the highest complexity close to their peak, these regions tend to have larger approximation errors and are therefore predominantly picked for refinement. The proposed way of adaptive sampling then ends up placing more points near the likelihood peak, thereby reducing the number of non-informative likelihood evaluations.

The choice of a constant  $N_{\text{ref}}$  is simple and could in principle be replaced by a more elaborate strategy (*e.g.*, based on the approximation error of the current subdomain relative to the total approximation error). A benefit of this enrichment criterion is that all residual expansions are computed with experimental designs of the same size. Upon choosing the domain with the maximum approximation error among the terminal domains, the error estimators then have a more comparable estimation accuracy.

### 4.2.2 Partitioning strategy

The partitioning strategy determines how a selected refinement domain is split. As described in Marelli et al. (2020), it is easy to define the splitting in the uniformly distributed quantile space  $\mathbf{U}$  and map the resulting split domains  $\mathcal{D}_{\mathbf{U}}^{\ell,p}$  to the (possibly unbounded) real space  $\mathbf{X}$  through an appropriate *isoprobabilistic transform* (e.g., the Rosenblatt transform (Rosenblatt, 1952)).

Similar to the original SSE algorithm presented in Marelli et al. (2020), we split the refinement domain in half *w.r.t.* its prior mass. The original algorithm chooses the splitting direction based on the partial variance in the refinement domain. This approach is well suited for generic function approximation problems. For the approximation of likelihood functions, however, we propose a partitioning strategy that is more apt for dealing with their *compact support*.

We propose to pick the split direction along which a split yields a maximum difference in the residual empirical variance between the two candidate subdomains created by the split. This can easily be visualized with an example given by the  $M = 2$  dimensional domain  $\mathcal{D}_{\mathbf{X}}^{\ell,p}$  in Figure 1(a). Assume this subdomain was selected as the refinement domain. To decide along which dimension to split, we construct the  $M$  candidate subdomain pairs  $\{\mathcal{D}_{\text{split}}^{i,1}, \mathcal{D}_{\text{split}}^{i,2}\}_{i=1,\dots,M}$  and estimate the corresponding  $\{E_{\text{split}}^i\}_{i=1,\dots,M}$  in those subdomains defined by

$$E_{\text{split}}^i \stackrel{\text{def}}{=} \left| \text{Var} \left[ \mathcal{R}^{\ell+1}(\mathcal{X}_{\text{split}}^{i,1}) \right] - \text{Var} \left[ \mathcal{R}^{\ell+1}(\mathcal{X}_{\text{split}}^{i,2}) \right] \right|. \quad (24)$$

In this expression,  $\mathcal{X}_{\text{split}}^{i,1}$  and  $\mathcal{X}_{\text{split}}^{i,2}$  denote subsets of the experimental design  $\mathcal{X}$  inside the subdomains  $\mathcal{D}_{\text{split}}^{i,1}$  and  $\mathcal{D}_{\text{split}}^{i,2}$  respectively. The occurring variances can be easily estimated with the empirical variance of the residuals in the respective candidate subdomains.

After computing the residual variance differences, the split is carried out along the dimension

$$d = \arg \max_{i \in \{1,\dots,M\}} E_{\text{split}}^i, \quad (25)$$

*i.e.*, to keep the subdomains  $\mathcal{D}_{\text{split}}^{d,1}$  and  $\mathcal{D}_{\text{split}}^{d,2}$  that introduce the largest difference in variance. For  $d = 1$ , the resulting split can be seen in Figure 1(d).

The choice of this partitioning strategy can be justified heuristically with the goal of approximating compact support functions. Assume that the likelihood function has compact support, this criterion will avoid *cutting through* its support and instead identify a split direction that results in one subdomain with large variance (expected to contain the likelihood support) and a subdomain with small variance. In subsequent steps, the algorithm will proceed by *cutting away* low variance subdomains, until the likelihood support is isolated.

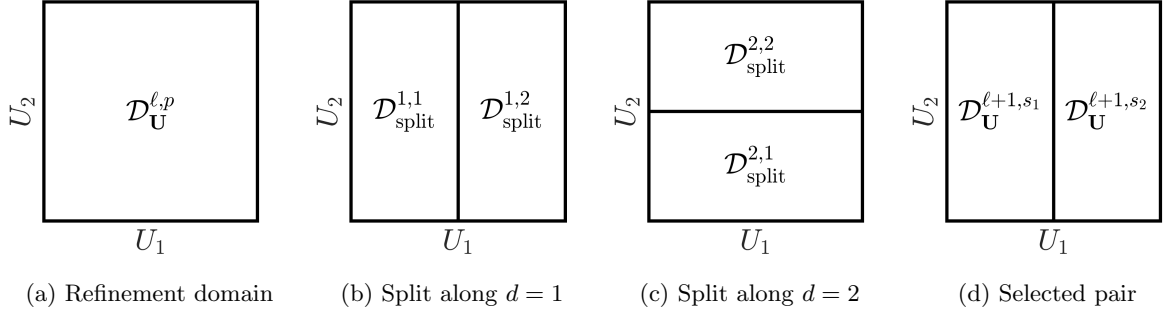


Figure 1: Partitioning strategy for a 2D example visualized in the quantile space  $\mathbf{U}$ . The refinement domain  $\mathcal{D}_{\mathbf{U}}^{\ell,p}$  is split into two subdomains  $\mathcal{D}_{\mathbf{U}}^{\ell+1,s_1}$  and  $\mathcal{D}_{\mathbf{U}}^{\ell+1,s_2}$ .

#### 4.2.3 The adaptive SSE algorithm

The algorithm is presented here with its two parameters  $N_{\text{ref}}$ , the minimum experimental design size needed to expand a residual, and  $N_{\text{ED}}$ , the final experimental design size. The sample  $\mathcal{X}^{\ell,p}$  refers to  $\mathcal{X} \cap \mathcal{D}_{\mathbf{X}}^{\ell,p}$ , *i.e.* the subset of  $\mathcal{X}$  inside  $\mathcal{D}_{\mathbf{X}}^{\ell,p}$ . Further, the multi-index set  $\mathcal{T} \in \mathbb{N}^2$  at each step of the algorithm gathers all indices  $(\ell, p)$  of unsplit subdomains. It thus denotes the terminal domains:  $\mathcal{D}_{\mathbf{X}}^k, k \in \mathcal{T}$ . For visualization purposes we show the first iterations of the algorithm for a two-dimensional example in Figure 2.

**1. Initialization:**

- (a)  $\mathcal{D}_{\mathbf{X}}^{0,1} = \mathcal{D}_{\mathbf{X}}$
- (b) Sample from prior distribution  $\mathcal{X} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_{\text{ref}})}\}$
- (c) Calculate the truncated expansion  $\widehat{\mathcal{R}}_S^{0,1}(\mathbf{X})$  of  $\mathcal{L}(\mathbf{X})$  in the full domain  $\mathcal{X}^{0,1}$ , retrieve its approximation error  $\mathcal{E}^{0,1}$  and initialize  $\mathcal{T} = \{(0, 1)\}$
- (d)  $\mathcal{R}^1(\mathbf{X}) = \mathcal{L}(\mathbf{X}) - \widehat{\mathcal{R}}_S^{0,1}(\mathbf{X})$

**2. For  $(\ell, p) = \arg \max_{k \in \mathcal{T}} \mathcal{E}^k$ :**

- (a) Split the current subdomain  $\mathcal{D}_{\mathbf{X}}^{\ell,p}$  in 2 sub-parts  $\mathcal{D}_{\mathbf{X}}^{\ell+1,\{s_1,s_2\}}$  and update  $\mathcal{T}$
- (b) **For each split  $s = \{s_1, s_2\}$** 
  - i. **If**  $|\mathcal{X}^{\ell+1,s}| < N_{\text{ref}}$  and  $N_{\text{ref}} - |\mathcal{X}^{\ell+1,s}| < N_{\text{ED}} - |\mathcal{X}|$ 
    - A. Enrich sample  $\mathcal{X}$  with  $N_{\text{ref}} - |\mathcal{X}^{\ell+1,s}|$  new points inside  $\mathcal{D}_{\mathbf{X}}^{\ell+1,s}$
  - ii. **If**  $|\mathcal{X}^{\ell+1,s}| = N_{\text{ref}}$ 
    - A. Create the truncated expansion  $\widehat{\mathcal{R}}_S^{\ell+1,s}(\mathbf{X}^{\ell+1,s})$  of the residual  $\mathcal{R}^{\ell+1}(\mathbf{X}^{\ell+1,s})$  in the current subdomain using  $\mathcal{X}^{\ell+1,s}$
    - B. Update the residual in the current subdomain  $\mathcal{R}^{\ell+2}(\mathbf{X}^{\ell+1,s}) = \mathcal{R}^{\ell+1}(\mathbf{X}^{\ell+1,s}) - \widehat{\mathcal{R}}_S^{\ell+1,s}(\mathbf{X}^{\ell+1,s})$
  - iii. Retrieve the approximation error  $\mathcal{E}^{\ell+1,s}$  from Eq. (23)
- (c) **If** no new expansions were created, terminate the algorithm, otherwise go back to 2

**3. Termination**



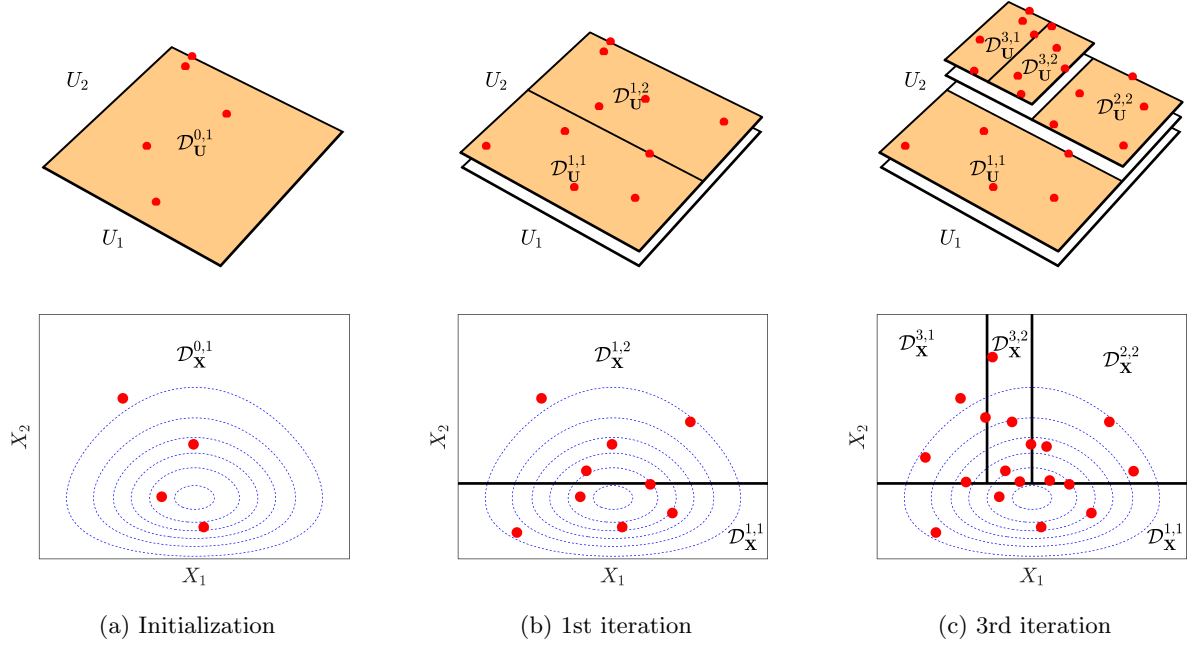


Figure 2: Graphical representation of the first steps of the adaptive SSE algorithm described in Section 4.2.3 for a two-dimensional problem with independent prior distribution. Upper row: partitioning in the quantile space; Lower row: partitioning in the unbounded real space with  $\pi(\mathbf{x})$  contour lines in dashed blue. Red dots show the adaptive experimental design that has a constant size of  $N_{\text{ref}} = 5$  in each created subdomain. The terminal domains  $\mathcal{T}$  are highlighted in orange. The splitting direction in each subdomain is determined randomly in this example.

(a) Return the full sequence of  $\mathcal{D}_{\mathbf{X}}^{\ell,p}$  and  $\widehat{\mathcal{R}}_S^{\ell,p}(\mathbf{X}^{\ell,p})$  needed to compute Eq. (15)

The updating of the multi-index set in Step 2a refers to removing the current index  $(\ell, p)$  from the set and adding to it the newly created indices  $(\ell + 1, s_1)$  and  $(\ell + 1, s_2)$ .

## 5 Case studies

To showcase the effectiveness of the proposed adaptive SSE approach, we present three case studies with increasing dimensionality and different model complexity: (i) a one-dimensional vibration problem for illustrative purposes, (ii) a six-dimensional heat transfer problem that describes the steady-state heat evolution in a solid body with inclusions and (iii) a 62-dimensional diffusion problem modeling the concentration-driven diffusion in a one-dimensional domain.

For all case studies, we adopt the adaptive sparse-PCE based on LARS approach developed in Blatman and Sudret (2011) through its numerical implementation in UQLab (Marelli and Sudret, 2014, 2019). Each  $\widehat{\mathcal{R}}_S^k$  is therefore a degree- and  $q$ -norm-adaptive polynomial

chaos expansion. We further introduce a rank truncation of  $r = 2$ , *i.e.* we limit the maximum number of input interactions (Marelli and Sudret, 2019) to two variables at a time. The truncation set for each spectral expansion (Eq. (8)) thus reads:

$$\mathcal{A}^{M,p,q,r} = \{\boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_q \leq p, \|\boldsymbol{\alpha}\|_0 \leq r\}. \quad (26)$$

where

$$\|\boldsymbol{\alpha}\|_q = \left( \sum_{i=1}^M \alpha_i^q \right)^{\frac{1}{q}}, q \in (0, 1]; \quad \|\boldsymbol{\alpha}\|_0 = \sum_{i=1}^M 1_{\{\alpha_i > 0\}}. \quad (27)$$

The  $q$ -norm is adaptively increased between  $q = \{0.5, \dots, 0.8\}$  while the maximum polynomial degree is adaptively increased in the interval  $p = \{0, 1, \dots, p\}$ , where the maximum degree  $p = 20$  for case study (i) and (ii) and  $p = 3$  for case study (iii) due to its high dimensionality.

In case study (ii) and (iii), the performance of SLE (Nagel and Sudret, 2016), the original non-adaptive SSE (Marelli et al., 2020) and the proposed adaptive SSE approach presented in Section 4.2 is compared. The comparison was omitted for case study (i), because only adaptive SSE succeeded in solving the problem. For clarity, we henceforth abbreviate the adaptive SSE algorithm to adSSE.

To simplify the comparison, the same partitioning strategy employed for adSSE (Section 4.2.2) was employed for the non-adaptive SSE approach. Also, the same experimental designs were used for the non-adaptive SSE and the SLE approaches. Finally, the same parameter  $N_{\text{ref}}$  was used to define the enrichment samples in adSSE and the termination criterion in non-adaptive SSE.

To assess the performance of the three algorithms considered, we define an error measure that allows to quantitatively compare the similarity of the SSE, adSSE and SLE solution with the reference MCMC solution. This comparison is inherently difficult, as a sampling-based approach (MCMC) needs to be compared to a functional approximation (SSE, adSSE, SLE). We proceed to compare the *univariate posterior marginals*, available analytically in SSE, adSSE and SLE (See Eq. (13) and Eq. (20)), to the reference posterior marginals estimated with kernel density estimation (KDE, Wand and Jones (1995)) from the MCMC sample. Denoting by  $\hat{\pi}_i(\kappa_i|\mathcal{Y})$  the SSE, adSSE or SLE approximations and by  $\pi_i(\kappa_i|\mathcal{Y})$  the reference solution, we define the following error measure

$$\eta \stackrel{\text{def}}{=} \frac{1}{M} \sum_{i=1}^M \text{JSD}(\hat{\pi}_i(\kappa_i|\mathcal{Y}) || \pi_i(\kappa_i|\mathcal{Y})) \quad (28)$$

where  $M$  is the dimensionality of the problem and JSD is the *Jensen-Shannon divergence* (Lin, 1991), a symmetric and bounded ( $[0, \log(2)]$ ) distance measure for probability distributions based on the Kullback-Leibler divergence.

The purpose of the error measure  $\eta$  is to allow for a fair comparison between the different methods investigated. It is not a practical measure for engineering applications because it relies on the availability of a reference solution, and its magnitude does not have a clear

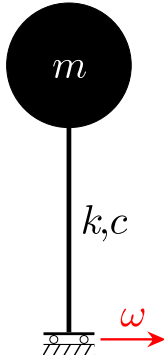


Figure 3: *1-dimensional vibration problem*: Sketch of the linear oscillator

quantitative interpretation. However, it is considerably more comprehensive than a pure moment-based error measure. Because it is averaged over all  $M$  marginals, it encapsulates the approximation accuracy of all univariate posterior marginals in a scalar value.

As all algorithms (SSE, adSSE, SLE) depend on a randomly chosen experimental designs, we produce 20 replications for case study (ii) and 5 replications for case study (iii) by running them multiple times.

### 5.1 1-dimensional vibration problem

This first case study serves as an illustrative example of how the proposed adaptive algorithm constructs an adSSE. In the presented problem the goal is the inference of a single unknown parameter with a bi-modal posterior distribution. This problem is very difficult to solve with standard MCMC methods due to the probability valley, *i.e.* low probability region, between the posterior peaks. The considered problem is fabricated and uses synthetic data, but is presented in the context of a relevant engineering problem.

Consider the oscillator displayed in Figure 3 subject to harmonic (*i.e.*, sinusoidal) excitation. Assume the prior information about its stiffness  $X \stackrel{\text{def}}{=} k$  is that it follows a lognormal distribution with  $\mu = 0.8 \text{ N/m}$  and  $\sigma = 0.1 \text{ N/m}$ . Its true value shall be determined using measurements of the oscillation amplitude at the location of the mass  $m$ . The known properties of the oscillator system are the oscillator mass  $m = 1 \text{ kg}$ , the excitation frequency  $\omega = 1 \text{ rad/s}$  and the viscous damping coefficient  $c = 0.1 \text{ Ns/m}$ . The oscillation amplitude is measured in five independent oscillation events and normalized by the forcing amplitude yielding the measured amplitude ratios  $\mathcal{Y} = \{9.01, 8.67, 8.84, 9.22, 8.54\}$ .

This problem is well known in mechanics and in the linear case (*i.e.*, assuming small deformations and linear material behavior) can be solved analytically with the *amplitude of the frequency response function*. This function returns the ratio between the steady state amplitude of a linear oscillator and the amplitude of its excitation. It is given by

$$\mathcal{M}(X) = \frac{m\omega^2}{\sqrt{(X - m\omega^2)^2 + (c\omega)^2}}. \quad (29)$$

We assume a discrepancy model with known discrepancy standard deviation  $\sigma$ . In conjunction with the available measurements  $\mathcal{Y}$ , this leads to the following likelihood function:

$$\mathcal{L}(\mathbf{x}; \mathcal{Y}) = \prod_{i=1}^5 \mathcal{N}(y^{(i)} | \mathcal{M}(x), \sigma^2). \quad (30)$$

We employ the adSSE algorithm to approximate this likelihood function with  $N_{\text{ref}} = 10$ . A few iterations from the solution process are shown in Figure 4. The top plots show the subdomains  $\mathcal{D}_X^{\ell,p}$  constructed at each refinement step, highlighting the terminal domains  $\mathcal{T}$ . The middle plots display the residual between the true likelihood and the approximation at the current iteration, as well as the adaptively chosen experimental design  $\mathcal{X}$ . The bottom plots display the target likelihood function and its current approximation.

The initial global approximation of the first iteration in Figure 4(a) is a constant polynomial based on the initial experimental design. By the third iteration, the algorithm has identified the subdomain  $\mathcal{D}_X^{2,2}$  as the one of interest and proceeds to refine it in subsequent steps. By the 8th iteration both likelihood peaks have been identified. Finally, by the 10th iteration in Figure 4(d), both likelihood peaks are approximated well by the adSSE approach.

The last iteration shows how the algorithm splits domains and adds new sample points. There is a clear clustering of subdomains and sample points near the likelihood peaks at  $X = 0.95$  and  $X = 1.05$ .

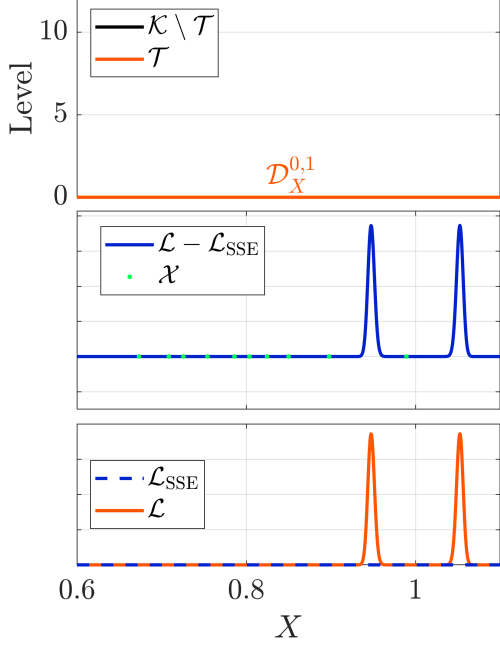
The results from Eq. (22) show that without further computations it would be possible to directly extract the posterior moments by post-processing the SSE coefficients. In the present bi-modal case, however, the posterior moments are not very meaningful. Instead, the available posterior approximation gives a full picture of the inferred parameter  $X | \mathcal{Y}$ . It is shown together with the true posterior and the original prior distribution in Figure 5.

For this case study, non-adaptive experimental design approaches like the standard SSE (Marelli et al., 2020) and the original SLE algorithm (Nagel and Sudret, 2016) will almost surely fail for the considered experimental design of  $N_{\text{ED}} = 100$ . In numerous trial runs these approaches did not manage to accurately reconstruct the likelihood function due to a lack of *informative samples* near the likelihood peaks.

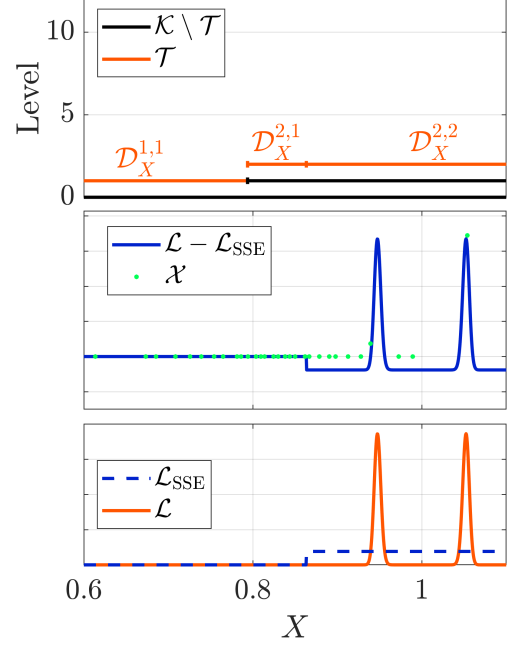
## 5.2 Moderate-dimensional heat transfer problem

This case study was originally presented in Nagel and Sudret (2016) and solved there using SLE. We again solve the same problem with SSE and compare the performance of SLE (Nagel and Sudret, 2016), the original non-adaptive SSE (Marelli et al., 2020) and the proposed adSSE approach presented in Section 4.2.

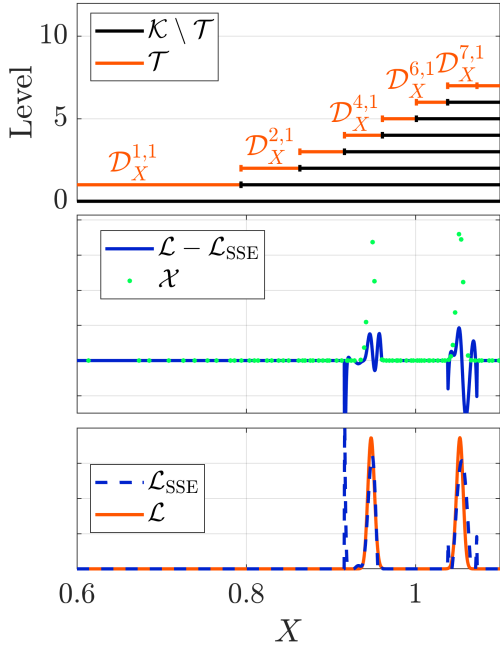
Consider the diffusion-driven stationary heat transfer problem sketched in Figure 6(a). It models a 2D plate with a background matrix of constant thermal conductivity  $\kappa_0$  and 6 inclusions with conductivities  $\boldsymbol{\kappa} \stackrel{\text{def}}{=} (\kappa_1, \dots, \kappa_6)^\top$ . The diffusion driven steady state heat distribution is described by a heat equation in Euclidean coordinates  $\mathbf{r} \stackrel{\text{def}}{=} (r_1, r_2)^\top$  of the



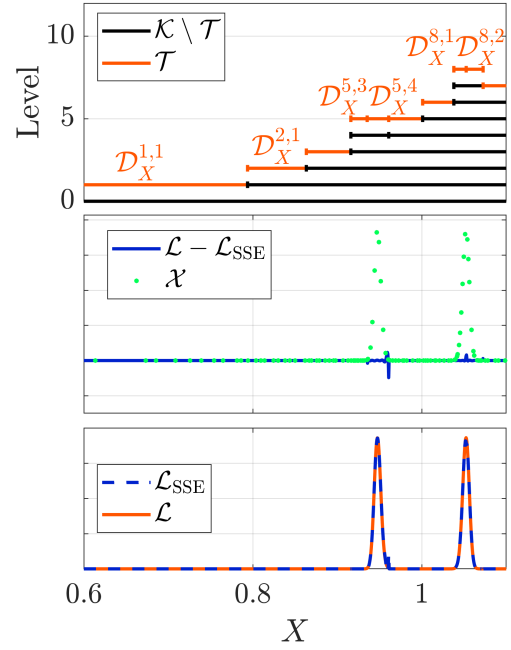
(a) 1st iteration,  $N_{\text{ED}} = 10$



(b) 3rd iteration,  $N_{\text{ED}} = 30$



(c) 8th iteration,  $N_{\text{ED}} = 80$



(d) 10th iteration,  $N_{\text{ED}} = 100$

Figure 4: *One-dimensional vibration problem*: Illustration of the adSSE algorithm approximating the likelihood function  $\mathcal{L}$ .

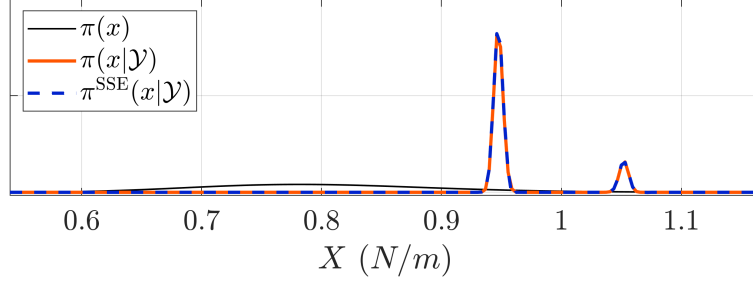


Figure 5: *One-dimensional vibration problem*: Comparison of the true multimodal posterior and its adSSE based approximation.

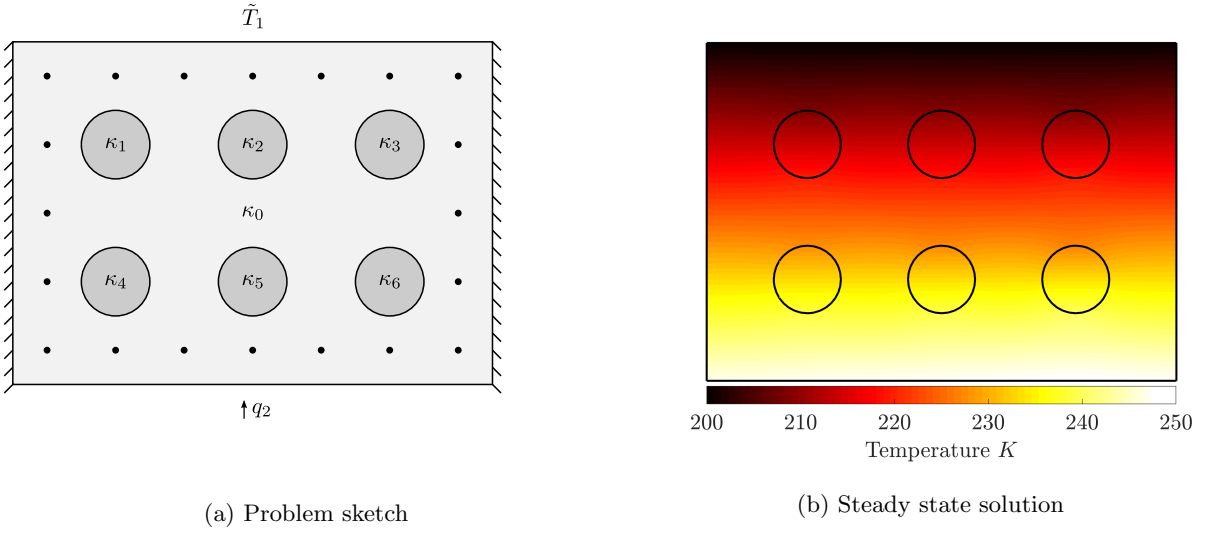


Figure 6: *Moderate-dimensional heat transfer problem*: Model setup and exemplary solution.

form

$$\nabla \cdot (\kappa(\mathbf{r}) \nabla \tilde{T}(\mathbf{r})) = 0, \quad (31)$$

where the thermal conductivity field is denoted by  $\kappa$  and the temperature field by  $\tilde{T}$ . The boundary conditions of the plate are given by a no-heat-flux Neumann boundary conditions on the *left* and *right* sides ( $\partial \tilde{T} / r_1 = 0$ ), a Neumann boundary condition on the *bottom* ( $\kappa_0 \partial \tilde{T} / r_2 = q_2$ ) and a temperature  $\tilde{T}_1$  Dirichlet boundary condition on the *top*.

We employ a finite element (FE) solver to solve the weak form of Eq. (31) by discretizing the domain into approximately  $10^5$  triangular elements. A sample solution returned by the FE-solver is shown in Figure 6b.

In this example we intend to infer the thermal conductivities  $\kappa$  of the inclusions. We assume the same problem constants as in Nagel and Sudret (2016) (*i.e.*,  $q_2 = 2,000 \text{ W/m}^2$ ,  $\tilde{T}_1 = 200 \text{ K}$ ,  $\kappa_0 = 45 \text{ W/m/K}$ ). The forward model  $\mathcal{M}$  takes as an input the conductivities of the inclusions  $\kappa$ , solves the finite element problem and returns the steady state temperature  $\tilde{\mathbf{T}} \stackrel{\text{def}}{=} (\tilde{T}_1, \dots, \tilde{T}_{20})^\top$  at the measurement points, *i.e.*,  $\mathcal{M} : \kappa \mapsto \tilde{\mathbf{T}}$ .

To solve the inverse problem, we assume a multivariate lognormal prior distribution with independent marginals on the inclusion conductivities, *i.e.*  $\pi(\boldsymbol{\kappa}) = \prod_{i=1}^6 \mathcal{LN}(\kappa_i | \mu = 30 \text{ W/m/K}, \sigma = 6 \text{ W/m/K})$ . We further assume an additive Gaussian discrepancy model, which yields the likelihood function

$$\mathcal{L}(\boldsymbol{\kappa}; \mathcal{Y}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{T} - \mathcal{M}(\boldsymbol{\kappa}))^\top (\mathbf{T} - \mathcal{M}(\boldsymbol{\kappa}))\right), \quad (32)$$

with a discrepancy standard deviation of  $\sigma$ .

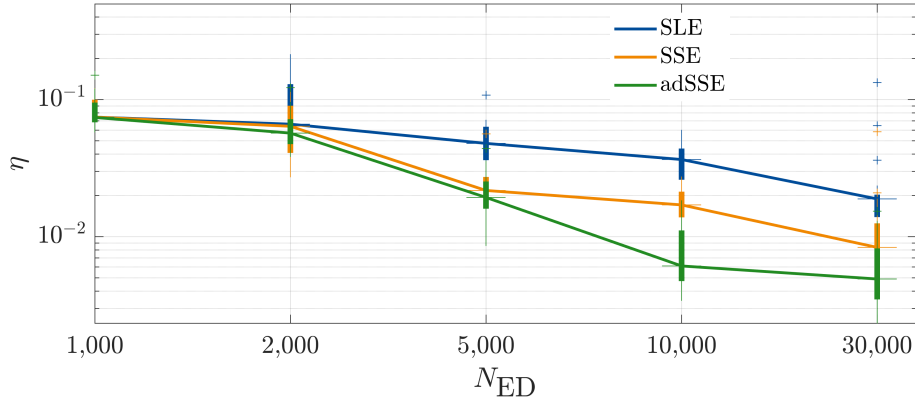
As measurements, we generate one temperature field with  $\hat{\boldsymbol{\kappa}} \stackrel{\text{def}}{=} (32, 36, 20, 24, 40, 28)^\top \text{ W/m/K}$  and collect its values at 20 points indicated by black dots in Figure 6(a). We then perturb these temperature values with additive Gaussian noise and use them as the inversion data  $\mathcal{Y} \stackrel{\text{def}}{=} \mathbf{T} = (T_1, \dots, T_{20})^\top$ .

We look at two instances of this problem that differ only by the discrepancy parameter  $\sigma$  from Eq. (32). The prior model response has a standard deviation of approximately 0.3 K, depending on the measurement point  $T_i$ . We therefore solve the problem first with a *large* value  $\sigma = 0.25 \text{ K}$  and second with a *small* value  $\sigma = 0.1 \text{ K}$ . As the discrepancy standard deviation determines how peaked the likelihood function is, the first problem has a likelihood function with a much wider support and in turn is significantly easier to solve than the second one. It is noted here that in practice, the peakedness of the likelihood function is either increased by a smaller discrepancy standard deviation, or the inclusion of additional experimental data.

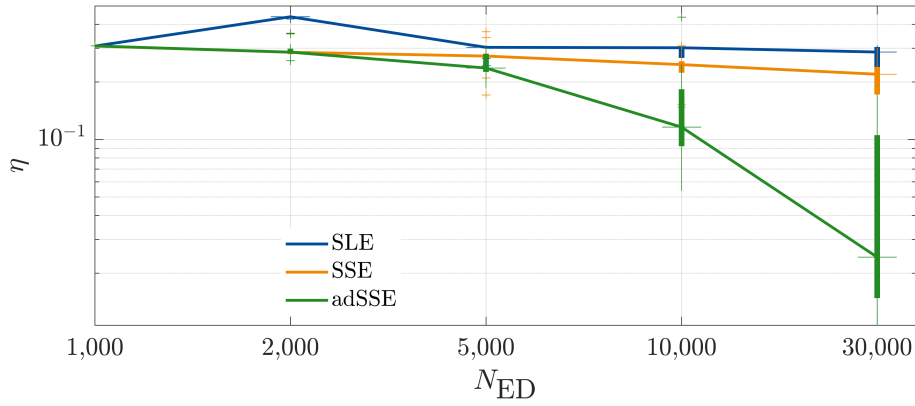
To monitor the dependence of the algorithms on the number of likelihood evaluations, we solve both problems with a set of maximum likelihood evaluations  $N_{\text{ED}} = \{1,000; 2,000; 5,000; 10,000; 30,000\}$ . The number of refinement samples is set to  $N_{\text{ref}} = 1,000$ .

As a benchmark, we use reference posterior samples generated by the affine-invariant ensemble sampler MCMC algorithm (Goodman and Weare, 2010) with 30,000 steps and 50 parallel chains, requiring a total of  $N_{\text{ED}} = 1.5 \cdot 10^6$  likelihood evaluations. Based on numerous heuristic convergence tests and due to the large number of MCMC steps, the resulting samples can be considered to accurately represent the true posterior distributions.

The results of the analyses are summarized in Figure 7, where the error measure  $\eta$  is plotted against the number of likelihood evaluations for the *large* and *small* standard deviation case. For the *large* discrepancy standard deviation case, both SSE approaches clearly outperform standard SLE *w.r.t.* the error measure  $\eta$ . This is most significant at mid-range experimental designs ( $N_{\text{ED}} = 5,000; 10,000$ ), where SLE does not reach the required high degrees and fails to accurately approximate the likelihood function. At larger experimental designs SLE catches up to non-adaptive SSE but is still outperformed by the proposed adSSE approach. The real strength of the adaptive algorithm shows for the case of a *small* discrepancy standard deviation, where the limitations of fixed experimental designs become obvious. When the likelihood function is nonzero in a small subdomain of the prior, the global SLE and non-adaptive SSE approach will fail in practice because of the insufficient



(a) Large discrepancy,  $\sigma = 0.25$  K



(b) Small discrepancy,  $\sigma = 0.1$  K

Figure 7: *Moderate-dimensional heat transfer problem*: Convergence of the  $\eta$  error measure (Eq. (28)) as a function of the experimental design size  $N_{ED}$  in 20 replications for SLE, SSE with a static experimental design and the proposed adSSE approach. We display the two discrepancy standard deviation cases  $\sigma = \{0.25, 0.1\}$  K.

number of samples placed in the informative regions. The adSSE approach, however, works very well in these types of problems. It manages to identify the regions of interest and produces a likelihood approximation that accurately reproduces the posterior marginals.

Tables 1 and 2 show the convergence of the adSSE method's moment estimate (mean and variance) to the reference solution for a single run. In brackets next to the moment estimates  $\xi$ , the relative error  $\epsilon \stackrel{\text{def}}{=} |\xi_{\text{MCMC}} - \xi_{\text{SSE}}|/\xi_{\text{MCMC}}$  is also shown. Due to the non-strict positivity of the SSE estimate, one variance estimate computed with Eq. (22) is negative and is therefore omitted from Table 2.

The full posterior marginals obtained from one run of adSSE with  $N_{ED} = 30,000$  are also compared to those of the reference MCMC and displayed in Figure 8. The individual plots show the univariate posterior marginals (*i.e.*  $\pi(x_i|\mathcal{Y})$ ) on the main diagonal and the bivariate posterior marginals (*i.e.*  $\pi(\mathbf{x}_{ij}|\mathcal{Y})$ ) in the  $i$ -th row and  $j$ -th column. It can be clearly seen that



Table 1: *Moderate-dimensional heat transfer problem:* adSSE results with *large* discrepancy standard deviation  $\sigma = 0.25$  K. Relative errors *w.r.t.* MCMC reference solution are shown in brackets.

$\mathbb{E}[\kappa_i \mathcal{Y}]$	$\kappa_1$	$\kappa_2$	$\kappa_3$	$\kappa_4$	$\kappa_5$	$\kappa_6$
<b>adSSE</b> $N_{\text{ED}} = 1,000$	30.00(0.7%)	30.00(7.1%)	19.51(5.8%)	30.01(7.1%)	35.39(2.9%)	30.00(14.6%)
$N_{\text{ED}} = 2,000$	31.74(6.5%)	29.80(7.7%)	20.89(0.9%)	30.00(7.1%)	35.56(2.4%)	28.30(8.1%)
$N_{\text{ED}} = 5,000$	29.70(0.4%)	31.74(1.7%)	20.21(2.4%)	30.21(6.5%)	36.45(0.0%)	27.58(5.4%)
$N_{\text{ED}} = 10,000$	29.96(0.5%)	32.41(0.3%)	19.90(3.9%)	31.55(2.3%)	36.57(0.3%)	26.20(0.1%)
$N_{\text{ED}} = 30,000$	29.92(0.4%)	32.40(0.3%)	19.94(3.7%)	31.96(1.1%)	36.53(0.2%)	26.15(0.1%)
<b>MCMC</b>	<b>29.81</b>	<b>32.30</b>	<b>20.71</b>	<b>32.31</b>	<b>36.45</b>	<b>26.18</b>
$\sqrt{\text{Var}[\kappa_i \mathcal{Y}]}$						
<b>adSSE</b> $N_{\text{ED}} = 1,000$	6.00(86.5%)	6.01(41.7%)	6.00(142.0%)	5.99(18.1%)	4.44(15.9%)	5.99(97.1%)
$N_{\text{ED}} = 2,000$	6.09(89.3%)	4.78(12.9%)	3.44(38.6%)	6.00(18.1%)	4.34(13.5%)	5.38(76.9%)
$N_{\text{ED}} = 5,000$	2.19(32.0%)	5.61(32.4%)	3.09(24.4%)	6.06(19.4%)	3.62(5.5%)	4.78(57.2%)
$N_{\text{ED}} = 10,000$	2.62(18.7%)	4.40(3.7%)	2.82(13.6%)	5.15(1.5%)	3.96(3.6%)	3.14(3.4%)
$N_{\text{ED}} = 30,000$	2.40(25.5%)	4.38(3.4%)	2.70(8.8%)	5.23(3.1%)	3.87(1.2%)	3.19(5.0%)
<b>MCMC</b>	<b>3.22</b>	<b>4.24</b>	<b>2.48</b>	<b>5.08</b>	<b>3.83</b>	<b>3.04</b>

Table 2: *Moderate-dimensional heat transfer problem:* adSSE results with *small* discrepancy standard deviation  $\sigma = 0.1$  K. Relative errors *w.r.t.* MCMC reference solution are shown in brackets. Field with an asterisk (\*) indicates negative variance estimate.

$\mathbb{E}[\kappa_i \mathcal{Y}]$	$\kappa_1$	$\kappa_2$	$\kappa_3$	$\kappa_4$	$\kappa_5$	$\kappa_6$
<b>adSSE</b> $N_{\text{ED}} = 1,000$	30.00(3.8%)	30.00(10.7%)	30.00(62.2%)	30.00(5.7%)	30.00(22.8%)	30.00(17.6%)
$N_{\text{ED}} = 2,000$	30.00(3.8%)	34.58(2.9%)	30.00(62.2%)	30.00(5.7%)	30.00(22.8%)	30.00(17.6%)
$N_{\text{ED}} = 5,000$	30.00(3.8%)	34.70(3.2%)	25.30(36.8%)	30.00(5.7%)	37.28(4.0%)	30.00(17.6%)
$N_{\text{ED}} = 10,000$	30.00(3.8%)	34.71(3.3%)	18.92(2.3%)	30.00(5.7%)	37.87(2.5%)	25.44(0.3%)
$N_{\text{ED}} = 30,000$	31.16(0.0%)	34.28(2.0%)	18.57(0.4%)	31.37(1.4%)	38.68(0.4%)	25.56(0.2%)
<b>MCMC</b>	<b>31.17</b>	<b>33.61</b>	<b>18.49</b>	<b>31.83</b>	<b>38.84</b>	<b>25.51</b>
$\sqrt{\text{Var}[\kappa_i \mathcal{Y}]}$						
<b>adSSE</b> $N_{\text{ED}} = 1,000$	6.00(268.3%)	6.00(149.7%)	6.00(362.1%)	6.00(69.4%)	5.99(202.8%)	5.99(264.7%)
$N_{\text{ED}} = 2,000$	6.00(268.5%)	4.53(88.9%)	6.00(362.2%)	6.00(69.5%)	6.00(203.1%)	6.00(265.2%)
$N_{\text{ED}} = 5,000$	6.00(268.5%)	4.43(84.3%)	2.87(120.8%)	6.00(69.5%)	4.29(116.8%)	6.00(265.2%)
$N_{\text{ED}} = 10,000$	4.49(176.0%)	*(*)	1.84(42.1%)	2.58(27.2%)	3.89(96.6%)	2.66(61.8%)
$N_{\text{ED}} = 30,000$	1.20(26.6%)	3.60(49.8%)	1.62(25.2%)	2.66(24.7%)	3.06(54.6%)	1.74(5.9%)
<b>MCMC</b>	<b>1.63</b>	<b>2.40</b>	<b>1.30</b>	<b>3.54</b>	<b>1.98</b>	<b>1.64</b>

the posterior characteristics are very well captured. However, the adSSE approach sometimes fails to accurately represent the tails of the distribution. This is especially obvious in the small discrepancy case in Figure 8(c) where the tail is sometimes cut off. We emphasize here that the SSE marginals are obtained analytically as 1D and 2D surfaces for the univariate and bivariate marginals respectively. For the reference MCMC approach, on the other hand, they need to be approximated with histograms based on the available posterior sample.

### 5.2.1 Convergence of the posterior moments

In practical inference applications, posterior moments are often one of the main quantities of interest. An estimator of these moments is readily available at every refinement step of adSSE through Eq. (22).

Tracking the evolution of the posterior moments throughout the adSSE iterations can be used as a heuristic estimator of the convergence of the adSSE algorithm. However, only the stability of the solution can be assessed, without guarantees on the bias. As an example, we now consider the *large discrepancy problem* and plot the evolution of the posterior mean and standard deviation for every  $X_i$  as a function of the number of likelihood evaluations in Figure 9. It can be seen that after  $\sim 10,000$  likelihood evaluations, most moment estimators achieve convergence to a value close to the reference solution. This plot also reveals a small bias of the  $\mathbb{E}[X_3|\mathcal{Y}]$  and  $\sqrt{\text{Var}[X_3|\mathcal{Y}]}$  estimators, that was previously highlighted in Table 1.

### 5.2.2 Influence of $N_{\text{ref}}$

The main hyperparameter of the proposed adSSE algorithm is the number  $N_{\text{ref}}$ , which corresponds to the number of sample points that are required at each PCE construction step (see Section 4.2). In Figure 10 we display the effect of different  $N_{\text{ref}}$  values on the convergence in the small and large discrepancy problems.

$N_{\text{ref}}$  influences the accuracy of the two error estimators used inside the adSSE algorithm. They are: (i) the residual expansion accuracy  $E_{\text{LOO}}$  in Eq. (23) and (ii) the splitting error  $E_{\text{split}}$  in Eq. (24).

Small values of  $N_{\text{ref}}$  allow to quickly obtain a crude likelihood approximation with limited experimental design sizes  $N_{\text{ED}}$ , but this comes at the cost of lower convergence rates at larger  $N_{\text{ED}}$ . This behaviour can be partially attributed to the deterioration of residual expansion error  $E_{\text{LOO}}$  in Eq. (23). At small experimental design sizes, the overall number of terminal domains is relatively small and this effect is not as pronounced. At larger experimental designs and higher numbers of subdomains, however, the error estimators high variances can lead to difficulties in identifying the true high error subdomains.

Large values of  $N_{\text{ref}}$  lead to slower initial convergence rates because of the smaller number of overall subdomains. The algorithm stability, however, is increased because both error estimators have lower variance and thereby allow the algorithm to more reliably identify the

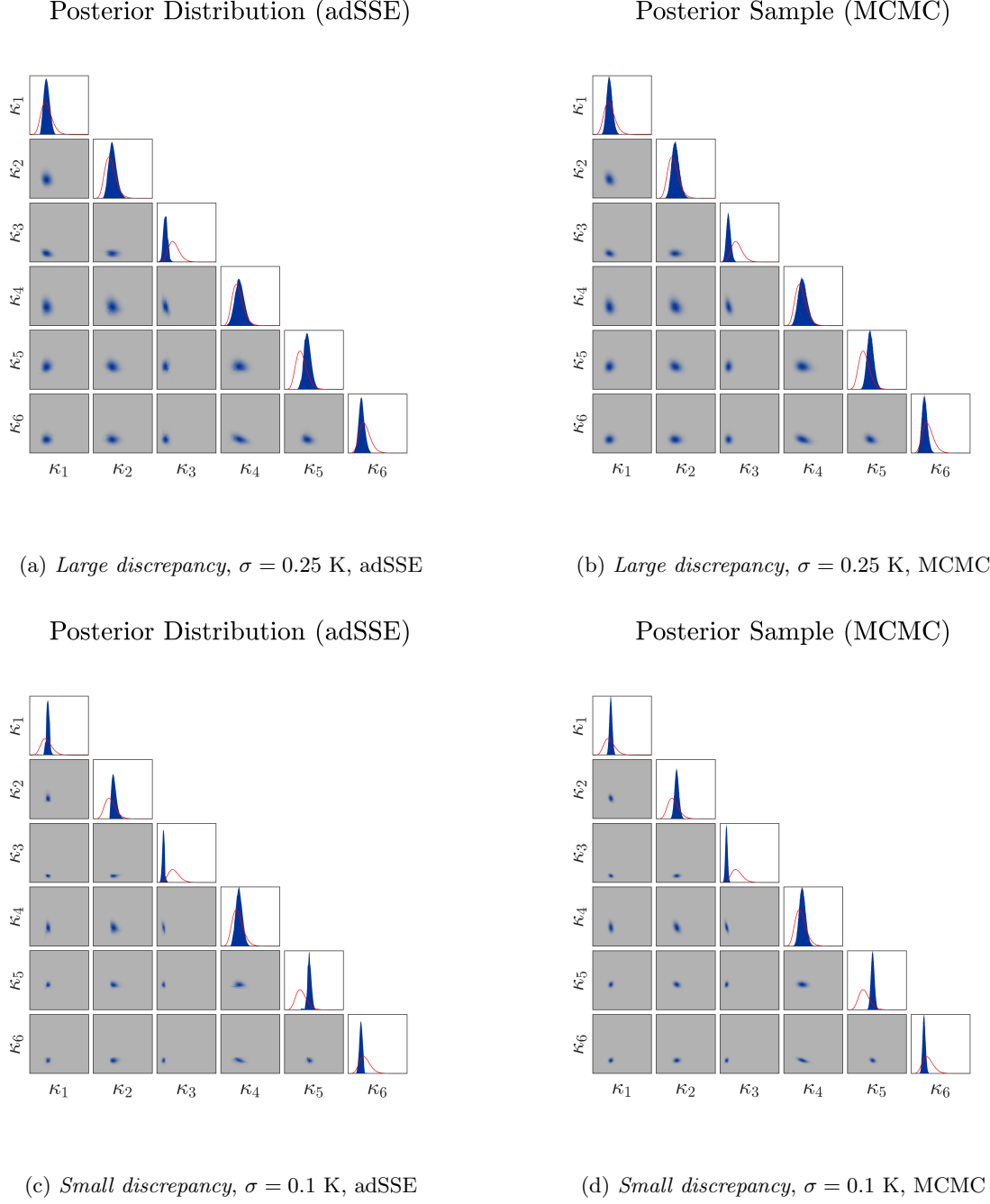


Figure 8: *Moderate-dimensional heat transfer problem*: Comparison plots of the posterior distribution marginals computed from adSSE ( $N_{\text{ED}} = 30,000$ ) and MCMC ( $N_{\text{ED}} = 1.5 \cdot 10^6$ ). The prior marginals are shown in red.

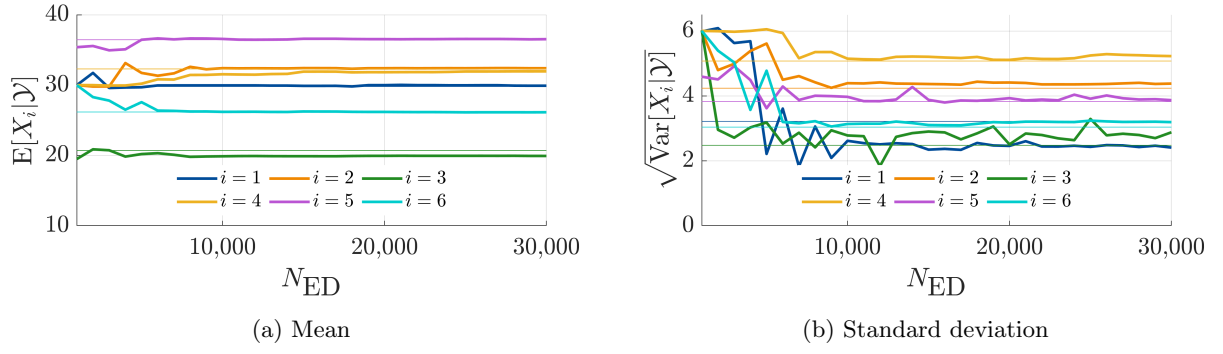
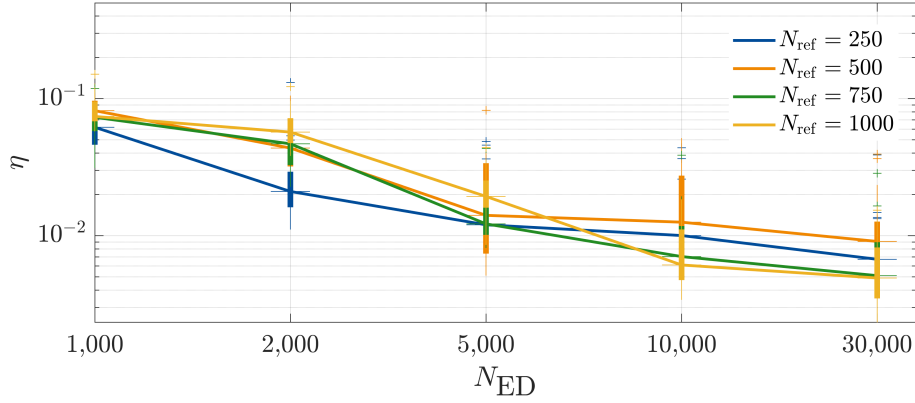
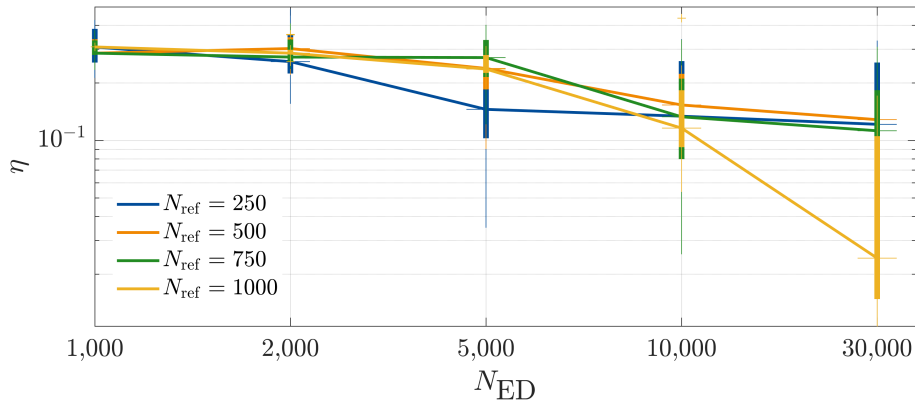


Figure 9: *Moderate-dimensional heat transfer problem*: Evolution of the posterior moment estimation for a typical run of the adSSE algorithm on the *small discrepancy problem*. The thin lines show the MCMC reference solution.



(a) *Large discrepancy*,  $\sigma = 0.25$  K



(b) *Small discrepancy*,  $\sigma = 0.1$  K

Figure 10: *Moderate-dimensional heat transfer problem*: Convergence of the  $\eta$  error measure (Eq. (28)) as a function of the experimental design size  $N_{ED}$  in 20 replications for the proposed adSSE approach with different  $N_{ref}$  parameters. We display the two discrepancy standard deviation cases  $\sigma = \{0.25, 0.1\}$  K.

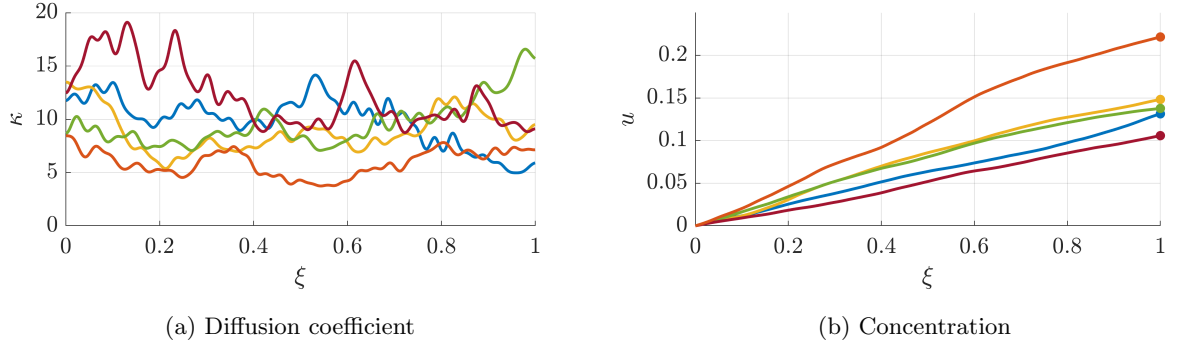


Figure 11: *High-dimensional diffusion problem*: 5 independent realizations of  $\mathbf{X}$  and the resulting  $\kappa$  and  $u$  with the model output  $u(1)$  highlighted with a circle.

true high error subdomains and choose the split directions that maximize Eq. (25).

### 5.3 High-dimensional diffusion problem

The last cast study shows that adSSE for Bayesian model inversion remains feasible in high dimensional problems. The considered forward model is often used as a standard benchmark in UQ computations (Shin and Xiu, 2016; Fajraoui et al., 2017). It represents the 1-D diffusion along a domain with coordinate  $\xi \in [0, 1]$  given by the following boundary value problem:

$$-\frac{\partial}{\partial \xi} \left[ \kappa(\xi) \frac{\partial u}{\partial \xi}(\xi) \right] = 1, \quad \text{with} \quad \begin{cases} u(0) = 0, \\ \frac{\partial u}{\partial \xi}(1) = 1. \end{cases} \quad (33)$$

The concentration field  $u$  can be used to describe any steady-state diffusion driven process (*e.g.*, heat diffusion, concentration diffusion, etc.). Assume that the diffusion coefficient  $\kappa$  is a log-normal random field given by  $\kappa(\xi, \omega) = \exp(10 + 3g(\xi))$  where  $g$  is a standard normal stationary Gaussian random field with exponential autocorrelation function  $\rho(\xi, \xi') = \exp(-3|\xi' - \xi|)$ . Let  $g$  be approximated through a *truncated Karhunen-Loève expansion*

$$g(\xi) \approx \sum_{k=1}^M X_k e_k(\xi), \quad (34)$$

with the pairwise uncorrelated random variables  $X_k$  denoting the *field coefficients* and the real valued function  $e_k$  obtained from the solution of the Fredholm equation for  $\rho$  (Ghanem and Spanos, 1991). The truncation variable is set to  $M = 62$  to explain 99% of the variance. Some realizations of the random field and resulting concentrations are shown in Figure 11.

In this example, the random vector of coefficients  $\mathbf{X} = (X_1, \dots, X_{62})$  shall be inferred using a single measurement of the diffusion field at  $u(\xi = 1)$  given by  $\mathcal{Y} = 0.16$ . The considered model therefore takes as an input a realization of that random vector, and returns the diffusion field at  $\xi = 1$ , *i.e.*,  $\mathcal{M} : \mathbf{x} \mapsto u(1)$ .

We impose a standard normal prior on the field coefficients such that  $\pi(\mathbf{x}) = \prod_{i=1}^{62} \mathcal{N}(x_i|0, 1)$  and assume the standard additive discrepancy model with known discrepancy variance  $\sigma^2 = 10^{-6}$ . This yields the likelihood function

$$\mathcal{L}(\mathbf{x}; \mathcal{Y}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} (0.16 - \mathcal{M}(\mathbf{x}))^2\right). \quad (35)$$

We proceed to compare the performance of standard SLE, non-adaptive SSE and the proposed adSSE approach on this example. We solve the problem with a set of maximum likelihood evaluations  $N_{\text{ED}} = \{2,000; 4,000; 6,000; 8,000; 10,000\}$ .

In the present high-dimensional case, it is necessary to set  $N_{\text{ref}}$  to a *relatively* large number ( $N_{\text{ref}} = 2,000$ ). At smaller  $N_{\text{ref}}$  numbers the variance of the estimator of  $E_{\text{split}}$  in Eq. (24) makes it difficult for the algorithm to correctly identify the splitting direction that maximizes Eq. (25).

To compare the results of the algorithms, they are compared to a reference MCMC solution obtained with the affine-invariant ensemble sampler (Goodman and Weare, 2010) algorithm with 100,000 steps and 100 parallel chains at a total cost of  $10^7$  likelihood evaluations.

To allow a quantitative comparison, we again use the error measure from Eq. (28) with  $M = 62$ . It is plotted for a set of maximum likelihood evaluations in Figure 12. It is clear that both SSE algorithms outperform SLE, while the adSSE approach manages to improve the performance of SLE by an order of magnitude.

The overall small magnitude of the error  $\eta$  in Figure 12 can be attributed to the low *active dimensionality* of this problem. Despite its high nominal dimensionality ( $M = 62$ ), this problem in fact has only very few active dimensions, as the first few variables are significantly more important than the rest. In physical terms, very local fluctuations of the conductivity do not influence the output  $u(1)$  which results from an integration of these fluctuations. Therefore, the biggest change between prior and posterior distribution happens in the first few parameters (see also Table 3), while the other parameters remain unchanged by the updating procedure. This results in a small value of the Jensen-Shannon divergence for the inactive dimensions that lower the average value  $\eta$  as defined in Eq. (28).

To highlight the results of one adSSE algorithm's instance with  $N_{\text{ED}} = 10,000$ , we display plots of the marginal posteriors in Figure 13. Due to the low active dimensionality of the problem, we focus on the first 3 parameters  $\{X_1, X_2, X_3\}$ . The remaining posterior parameters are not significantly influenced by the considered data. The comparative plots show a good agreement between the adSSE and the reference solution, especially *w.r.t.* the interaction between  $X_1$  and  $\{X_2, X_3\}$ . It can also be seen that the data  $\mathcal{Y}$  has the biggest influence on the first parameter  $X_1$ .

For the same instance, we also compute the first two posterior moments for all posterior marginals and compare them to the MCMC reference solution. The resulting values are presented in Table 3. Keeping in mind that the prior distribution is a multivariate standard

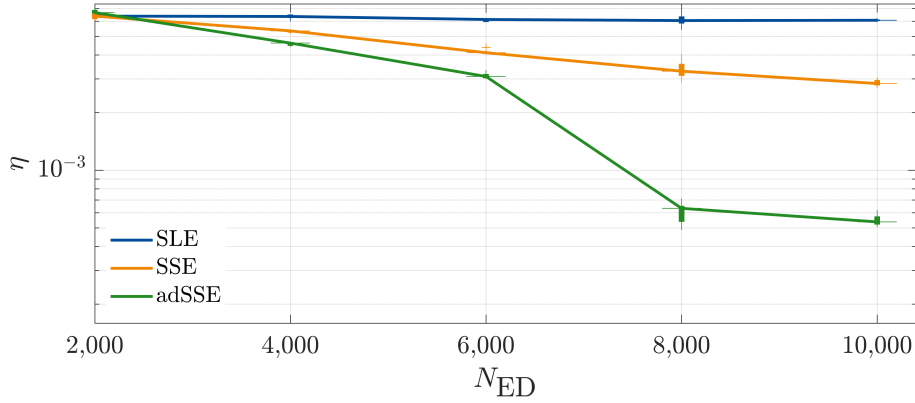


Figure 12: *High-dimensional diffusion problem*: Convergence of the  $\eta$  error measure (Eq. (28)) over five replications for SLE, SSE with a static experimental design and the proposed adSSE approach.

Table 3: *High-dimensional diffusion problem*: Posterior mean and standard deviation for all marginals. The values in brackets are computed from the MCMC reference solution. The prior is a multivariate standard normal distribution ( $\mathbb{E}[X_i] = 0$  and  $\sqrt{\text{Var}[X_i|\mathcal{Y}]} = 1$  for  $i = 1, \dots, 6, 10, 20, \dots, 50, 62$ ).

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$
$\mathbb{E}[X_i \mathcal{Y}]$	-0.88(-0.89)	0.14(0.14)	0.07(0.09)	0.04(-0.03)	-0.04(-0.00)	-0.04(0.03)
$\sqrt{\text{Var}[X_i \mathcal{Y}]}$	0.16(0.14)	1.07(1.04)	0.97(1.03)	1.02(1.00)	1.00(1.00)	0.98(1.03)
	$X_{10}$	$X_{20}$	$X_{30}$	$X_{40}$	$X_{50}$	$X_{62}$
$\mathbb{E}[X_i \mathcal{Y}]$	0.00(0.03)	0.00(0.00)	-0.02(-0.01)	-0.03(-0.01)	-0.00(0.01)	0.00(-0.03)
$\sqrt{\text{Var}[X_i \mathcal{Y}]}$	1.03(1.01)	1.03(0.98)	1.00(1.00)	1.04(1.02)	1.00(0.99)	1.00(0.98)

normal distribution ( $\mathbb{E}[X_i] = 0$  and  $\sqrt{\text{Var}[X_i|\mathcal{Y}]} = 1$  for  $i = 1, \dots, 62$ ) it is obvious from this table that the data most significantly affects the first three parameters.

The adSSE approach manages to accurately recover the first two posterior moments at the relatively low cost of  $N_{ED} = 10,000$  likelihood evaluations. The average absolute error for  $\mathbb{E}[X_i]$  and  $\sqrt{\text{Var}[X_i|\mathcal{Y}]}$  is approximately 0.02.

## 6 Conclusions

Motivated by the recently proposed spectral likelihood expansions (SLE) framework for Bayesian model inversion presented in Nagel and Sudret (2016), we showed that the same analytical post-processing capabilities can be derived when the novel stochastic spectral embedding (SSE) approach from Marelli et al. (2020) is applied to likelihood functions. Because SSE is designed for models with *complex local characteristics*, it was expected to outperform SLE on practically relevant, highly localized likelihood functions. To further improve SSE

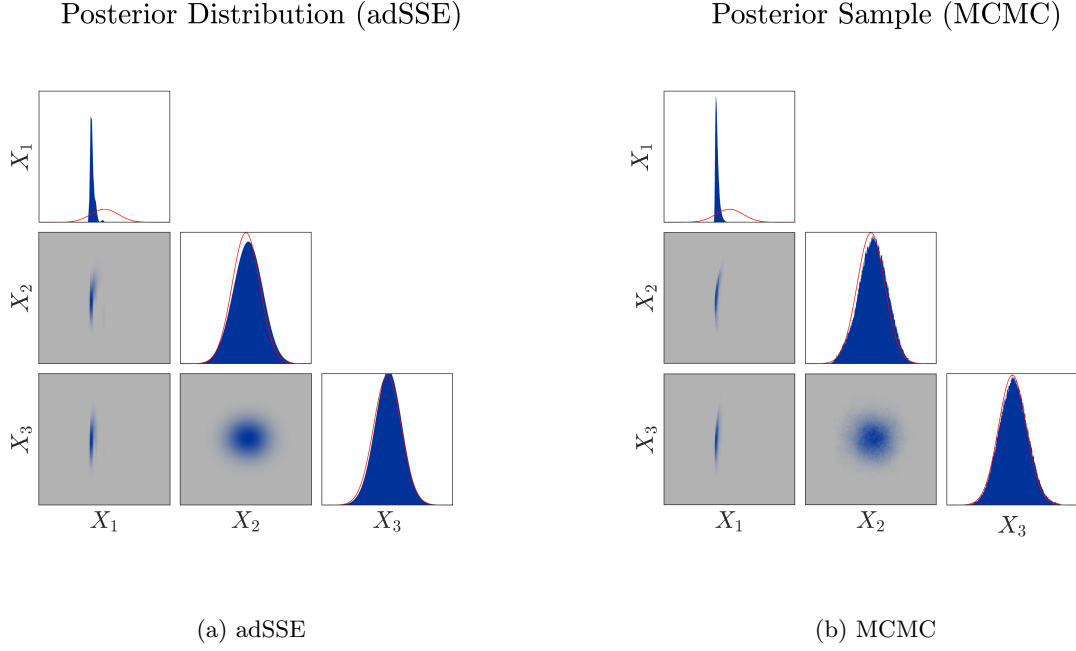


Figure 13: *High-dimensional diffusion problem*: Comparison plots of the posterior distribution marginals for the first 3 parameters  $\{X_1, X_2, X_3\}$  computed from adSSE ( $N_{\text{ED}} = 10,000$ ) and MCMC ( $N_{\text{ED}} = 10^7$ ). The prior marginals are shown in red.

performance in Bayesian model inversion applications, we introduced a novel adaptive sampling procedure and modified partitioning strategy.

There are a few unsolved shortcomings of the SSE that will be addressed in future works. Namely, the discontinuities at the subdomain boundaries may be a source of error that should be addressed. Additionally, for the adaptive SSE algorithm, it is not possible at the moment to specify the optimal  $N_{\text{ref}}$  parameter a priori. In light of the considerable influence of that parameter as shown in Section 5.2.2, it might be necessary to adaptively adjust it, or decouple this parameter from the termination criterion.

Approximating likelihood functions through local PCEs prohibits the enforcement of strict positivity throughout the function domain. For visualization purposes this is not an issue, as negative predictions can simply be set to 0 in a post-processing step. When computing posterior expectations with Eq. (22), however, this can lead to erroneous results such as negative posterior variances. One way to enforce strict positivity is through an initial transformation of the likelihood function (*e.g.*, log-likelihood  $\log \mathcal{L} \approx \sum_{k \in \mathcal{K}} f_k^{\text{PCE}}(\mathbf{X})$ ). This is avoided in the present work because it comes at a loss of the desirable analytical post-processing properties.

The biggest advantage of SSE, however, is that it poses the challenging Bayesian computation in a function approximation setting. This yields an analytical expression of the posterior distribution and preserves the analytical post-processing capabilities of SLE while



delivering highly improved likelihood function approximations. As opposed to many existing algorithms, SSE can even efficiently solve Bayesian problems with multiple posterior modes. As shown in the case studies, the proposed adaptive algorithm further capitalizes on the compact support nature of likelihood functions and leads to significant performance gains, especially at larger experimental designs.

## Acknowledgements

The PhD thesis of the first author is supported by ETH grant #44 17-1.

## References

- Beck, J. and S. Au (2002). Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. *Journal of Engineering Mechanics* 128(4), 380–391.
- Beck, J. L. and L. S. Katafygiotis (1998). Updating models and their uncertainties. I: Bayesian statistical framework. *Journal of Engineering Mechanics* 124(4), 455–461.
- Birolleau, A., G. Poëtte, and D. Lucor (2014). Adaptive Bayesian inference for discontinuous inverse problems, application to hyperbolic conservation laws. *Communications in Computational Physics* 16(1), 1–34.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Information Science and Statistics. Springer.
- Blatman, G. and B. Sudret (2010). An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics* 25, 183–197.
- Blatman, G. and B. Sudret (2011). Adaptive sparse polynomial chaos expansion based on Least Angle Regression. *Journal of Computational Physics* 230, 2345–2367.
- Brooks, S. P. and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7(4), 434–455.
- Chapelle, O., V. Vapnik, and Y. Bengio (2002). Model selection for small sample regression. *Journal of Machine Learning Research* 48(1), 9–23.
- Ching, J. and Y. Chen (2007). Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics* 133(7), 816–832.

- Dick, J., R. N. Gantner, Q. T. Le Gia, and C. Schwab (2017). Multilevel higher-order quasi-Monte Carlo Bayesian estimation. *Mathematical Models and Methods in Applied Sciences* 27(5), 953–995.
- El Moselhy, T. A. and Y. M. Marzouk (2012). Bayesian inference with optimal maps. *Journal of Computational Physics* 231(23), 7815–7850.
- Fajraoui, N., S. Marelli, and B. Sudret (2017). Sequential design of experiment for sparse polynomial chaos expansions. *SIAM/ASA Journal on Uncertainty Quantification* 5(1), 1061–1085.
- Gantner, R. N. and M. D. Peters (2018). Higher-order quasi-Monte Carlo for Bayesian shape inversion. *SIAM/ASA Journal on Uncertainty Quantification* 6(2), 707–736.
- Gautschi, W. (2004). *Orthogonal polynomials: Computation and approximation*. Numerical Mathematics and Scientific Computation. Oxford University Press.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2014). *Bayesian Data Analysis* (3 ed.). Texts in Statistical Science. CRC Press.
- Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4), 457–472.
- Ghanem, R. G. and P. Spanos (1991). *Stochastic finite elements – A spectral approach*. Springer Verlag, New York. (Reedited by Dover Publications, Mineola, 2003).
- Goodman, J. and J. Weare (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science* 5(1), 65–80.
- Haario, H., E. Saksman, and J. Tamminen (2001). An adaptive Metropolis algorithm. *Bernoulli* 7(2), 223–242.
- Harney, H. L. (2016). *Bayesian Inference: Data Evaluation and Decisions* (2 ed.). Springer International Publishing.
- Jakeman, J., M. S. Eldred, and K. Sargsyan (2015). Enhancing  $\ell_1$ -minimization estimates of polynomial chaos expansions using basis selection. *Journal of Computational Physics* 289, 18–34.
- Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 186(1007), 453–461.

- Kennedy, M. C. and A. O’Hagan (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(3), 425–464.
- Li, J. and Y. M. Marzouk (2014). Adaptive construction of surrogates for the Bayesian solution of inverse problems. *SIAM Journal on Scientific Computing* 36(3), A1163–A1186.
- Lin, J. (1991). Divergence measures based on the Shannon entropy. *Transactions on Information Theory* 37(1), 145–151.
- Lüthen, N., S. Marelli, and B. Sudret (2020). Sparse polynomial chaos expansions: Review and benchmark. *SIAM/ASA Journal on Uncertainty Quantification*. submitted, <https://arxiv.org/abs/2002.01290>.
- Marelli, S. and B. Sudret (2014). UQLab: A framework for uncertainty quantification in Matlab. In *Vulnerability, Uncertainty, and Risk (Proceedings of the 2nd International Conference on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom)*, pp. 2554–2563.
- Marelli, S. and B. Sudret (2019). UQLab user manual – Polynomial chaos expansions. Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland. Report # UQLab-V1.3-104.
- Marelli, S., P.-R. Wagner, C. Lataniotis, and B. Sudret (2020). Stochastic spectral embedding. submitted, <https://arxiv.org/abs/2004.04480>.
- Marin, J.-M., P. Pudlo, C. P. Robert, and R. J. Ryder (2012). Approximate Bayesian computational methods. *Statistics and Computing* 22(6), 1167–1180.
- Marzouk, Y., T. Moselhy, M. Parno, and A. Spantini (2016). Sampling via measure transport: An introduction. In R. Ghanem, D. Higdon, and H. Owhadi (Eds.), *Handbook of Uncertainty Quantification*. Springer International Publishing.
- Marzouk, Y. M., H. N. Najm, and L. A. Rahn (2007). Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics* 224, 560–586.
- Marzouk, Y. M. and D. Xiu (2009). A stochastic collocation approach to Bayesian inference in inverse problems. *Communications in Computational Physics* 6(4), 826–847.
- Nagel, J. and B. Sudret (2016). Spectral likelihood expansions for Bayesian inference. *Journal of Computational Physics* 309, 267–294.
- Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*, Handbooks of Modern Statistical Methods, Chapter 5, pp. 113–162. Chapman & Hall/CRC.

- Parno, M. D. (2015). *Transport maps for accelerated Bayesian computation*. PhD thesis, Massachusetts Institute of Technology (MIT).
- Robert, C. P. and G. Casella (2004). *Monte Carlo statistical methods (2<sup>nd</sup> Ed.)*. Springer Series in Statistics. Springer Verlag.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23(3), 470–472.
- Shin, Y. and D. Xiu (2016). Nonadaptive quasi-optimal points selection for least squares linear regression. *SIAM Journal on Scientific Computing* 38(1), A385–A411.
- Sisson, S. A., Y. Fan, and M. A. Beaumont (Eds.) (2018). *Handbook of approximate Bayesian computation*. Chapman and Hall/CRC.
- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics (SIAM).
- Tierney, L. and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* 81(393), 82–86.
- Tierney, L., R. E. Kass, and J. B. Kadane (1989a). Approximate marginal densities of nonlinear functions. *Biometrika* 76(3), 425–433.
- Tierney, L., R. E. Kass, and J. B. Kadane (1989b). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association* 84(407), 710–716.
- Wagner, P.-R., R. Fahrni, M. Klippel, A. Frangi, and B. Sudret (2020). Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels. *Engineering Structures* 205(15), 110063.
- Wand, M. and M. C. Jones (1995). *Kernel smoothing*. Chapman and Hall, Boca Raton.
- Xiu, D. (2010). *Numerical methods for stochastic computations – A spectral method approach*. Princeton University press.
- Xiu, D. and G. E. Karniadakis (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing* 24(2), 619–644.