Spectral likelihood expansions for Bayesian inference

Journal Article

Author(s):
Nagel, Joseph B.; Sudret, Bruno ⓘD

Spectral likelihood expansions for Bayesian inference

Joseph B. Nagel[a,*], Bruno Sudret[a,**]

a ETH Zürich, Institute of Structural Engineering
Chair of Risk, Safety & Uncertainty Quantification
Stefano-Franscini-Platz 5, CH-8093 Zürich, Switzerland

Abstract

A spectral approach to Bayesian inference is presented. It pursues the emulation of the posterior probability density. The starting point is a series expansion of the likelihood function in terms of orthogonal polynomials. From this spectral likelihood expansion all statistical quantities of interest can be calculated semi-analytically. The posterior is formally represented as the product of a reference density and a linear combination of polynomial basis functions. Both the model evidence and the posterior moments are related to the expansion coefficients. This formulation avoids Markov chain Monte Carlo simulation and allows one to make use of linear least squares instead. The pros and cons of spectral Bayesian inference are discussed and demonstrated on the basis of simple applications from classical statistics and inverse modeling.

Keywords: Uncertainty quantification, Bayesian inference, Inverse problem, Surrogate modeling, Polynomial chaos, Posterior expansion

1. Introduction

In view of inverse modeling [1, 2] and uncertainty quantification [3, 4], Bayesian inference establishes a convenient framework for the data analysis of engineering systems [5–7]. It adopts probability theory in order to represent, propagate and update epistemic parameter uncertainties. The prior distribution captures the uncertainty of unknown model parameters before the data are analyzed. A posterior is then constructed as an updated distribution that captures the remaining uncertainty after the data have been processed. The computation of this posterior is the primary task in Bayesian inference.

Some simplified statistical models admit closed-form expressions of the posterior density. Beyond these so-called conjugate cases, computational approaches either aim at evaluating expectation values under the posterior or drawing samples from it [8]. This is usually accomplished with stochastic methods such as Markov chain Monte Carlo [9, 10]. Nowadays this class of techniques constitutes the mainstay of Bayesian computations. The posterior is explored by realizing an appropriate Markov chain over the prior support that exhibits the posterior as its long-run distribution. In turn, the obtained sample is used to empirically approximate the statistical quantities of interest. These include the characteristics of the posterior and the predictive distribution. This way of proceeding suffers from some inherent deficiencies. The presence of sample autocorrelation and the absence of a convergence criterion cause severe practical problems. Moreover, Markov chain Monte Carlo typically requires a large number of serial forward model runs. Since in engineering applications even a single model run can be computationally taxing, this may be prohibitive. In the recent past, numerous enhancements have been proposed in order to accelerate Markov chain Monte Carlo for Bayesian inverse problems. This includes the implementation of more efficient sampling algorithms, e.g. transitional Markov chain Monte Carlo [11, 12] or Hamiltonian Monte Carlo [13–15], and the substitution

---

*Corresponding author
**Principal corresponding author
Email addresses: `nagel@ibk.baug.ethz.ch` (Joseph B. Nagel), `sudret@ibk.baug.ethz.ch` (Bruno Sudret)

of the forward model with an inexpensive metamodel, e.g. based on Gaussian process models [16, 17] or polynomial chaos expansions [18–20]. Although these approaches promise significant speedups, they still inherit all principle shortcomings of sample-based posterior representations.

Unfortunately there are only few fundamental alternatives to stochastic sampling. Variational Bayesian inference establishes such an alternative where the posterior is sought through deterministic optimization [21–23]. In particular, a member from a simple parametric family of probability densities is selected such that some distance to the posterior is minimized. In this regard, the Kullback-Leibler divergence is often chosen as a relative measure of the dissimilarity of two probability densities. The procedure commonly rests upon some simplifying independence assumptions. Variational methods are regarded as less computing intensive than Markov chain Monte Carlo, yet they are only approximate. They are prominently used in machine learning and computer science [24, 25], and since recently such methods are applied to inverse problems [26, 27], too. A particularly interesting implementation of variational Bayesian inference has been proposed in [28]. The posterior is parametrized as a transformation of the prior density and can be computed based on the corresponding back-transformation. More specifically, a random variable transformation is sought in polynomial form such that the Kullback-Leibler divergence between the prior density and the back-transformed posterior density is minimized. This formulation is supported by arguments from optimal transport theory which also allows for a practical regularization of the problem. Finally, samples from the posterior distribution are obtained by independently sampling from the prior and applying the polynomial map. Another approach to certain Bayesian inverse problems has been recently devised in [29, 30]. Based on monomial Taylor expansions of the forward model and of the posterior density, the computation of expectation values under the posterior is tackled by sparse numerical quadrature.

In this paper we propose a novel approach to surrogate the posterior probability density in itself. The main idea is to decompose the likelihood function into a series of polynomials that are orthogonal with respect to the prior density. It is shown that all statistical quantities of interest can then be easily extracted from this spectral likelihood expansion. Emulators of the joint posterior density and its marginals are derived as the product of the prior that functions as the reference density and a linear combination of polynomials that acts as an adjustment. In doing so, the model evidence simply emerges as the coefficient of the constant term in the expansion. Moreover, closed-form expressions for the first posterior moments in terms of the low-order expansion coefficients are given. The propagation of the posterior uncertainty through physical models can be easily accomplished based on a further postprocessing of the expansion coefficients. In this sense, spectral Bayesian inference is semi-analytic. While the corrections required for an expansion of the posterior with respect to the prior as the reference density may be large, they can be small for an expansion around a properly chosen auxiliary density. A change of the reference density is therefore suggested in order to increase the efficiency of computing a posterior surrogate. The devised formulation entirely avoids Markov chain Monte Carlo. Instead it draws on the machinery of spectral methods [31–33] and approximation theory [34–36]. It is proposed to compute the expansion coefficients via linear least squares [37, 38]. This allows one to make use of the wealth of statistical learning methods [39, 40] that are designed for this type of problems. The approach features a natural convergence criterion and it is amenable to parallel computing.

The scope of applicability of the proposed approach covers problems from Bayesian inference for which the likelihood function can be evaluated and for which polynomials can be constructed that are orthogonal with respect to the prior, possibly after a carefully chosen variable transformation. This excludes statistical models that involve intractable likelihoods [41, 42], i.e. the likelihood cannot be exactly evaluated. It also excludes improper prior distributions [43, 44], i.e. the prior does not integrate to one or any finite value, and models with pathologic priors such as the Cauchy distribution for which the moments are not defined [45, 46]. Many hierarchical Bayesian models [47, 48] are not covered by the devised problem formulation. They are either based on conditional priors which does not allow for orthogonal polynomials or on integrated likelihoods which can only be evaluated subject to noise.

Spectral likelihood expansions complement the existing array of Bayesian methods with a way of surrogating the posterior density directly. They have the potential to remedy at least some of the shortcomings of Markov chain Monte Carlo. Yet, their practical implementation poses challenges. Hence, the goal of this paper is to discuss and investigate the possibilities and limitations of the approach. The method of spectral likelihood expansions is therefore applied to well-known calibration problems from classical statistics and

inverse heat conduction problem. We restrict the analysis to low-dimensional problems. The final results are compared with corresponding results from Markov chain Monte Carlo simulations.

The manuscript is structured as follows. The principles of Bayesian inference are summarized in Section 2. Surrogate forward modeling with polynomial chaos expansions is reviewed in Section 3. After that, spectral likelihood expansions are introduced as an alternative approach to Bayesian inference in Section 4. Two well-known Gaussian fitting examples and the identification of thermal properties of a composite material serve as numerical demonstrations in Section 5. Finally, it is summarized and concluded in Section 6.

2. Bayesian inference

Let $\boldsymbol{x} = (x_1, \ldots, x_M)^\top \in \mathcal{D}_{\boldsymbol{x}}$ with $\mathcal{D}_{\boldsymbol{x}} = \mathcal{D}_{x_1} \times \ldots \times \mathcal{D}_{x_M} \subset \mathbb{R}^M$ be a vector of unknown parameters. The goal of statistical inference is to deduce these unknowns from the observed data $\boldsymbol{y} = (y_1, \ldots, y_N)^\top \in \mathbb{R}^N$. In Bayesian statistics one adapts probabilistic models for representing uncertainty. Hence, let $(\Omega, \mathcal{F}, \mathcal{P})$ be a suitable probability space with a sample space Ω, a σ-field $\mathcal{F}$ and a probability measure $\mathcal{P}$. On this space one can define a prior model of the unknowns and an observational model of the data that represent the encountered parameter uncertainties and the experimental situation, respectively.

The epistemic uncertainty of the parameter values is cast as a $\mathcal{D}_{\boldsymbol{x}}$-valued random vector $\boldsymbol{X} \colon \Omega \to \mathcal{D}_{\boldsymbol{x}} \subset \mathbb{R}^M$. Here, the components of $\boldsymbol{X} = (X_1, \ldots, X_M)^\top$ are $\mathcal{D}_{x_i}$-valued random variables $X_i \colon \Omega \to \mathcal{D}_{x_i} \subset \mathbb{R}$ for $i = 1, \ldots, M$. Since the data have not been processed at this stage, the joint density of $\boldsymbol{X} \sim \pi(\boldsymbol{x})$ is called the *prior density*. Similarly, a $\mathbb{R}^N$-valued random vector $\boldsymbol{Y} \colon \Omega \to \mathbb{R}^N$ represents the observables. The components of $\boldsymbol{Y} = (Y_1, \ldots, Y_N)^\top$ are real-valued random variables $Y_i \colon \Omega \to \mathbb{R}$ for $i = 1, \ldots, N$. In order to draw inferences from the data about the unknowns, one has to formulate an observational model that establishes a relationship between those quantities. Commonly this is a probabilistic representation of the observables $\boldsymbol{Y} | \boldsymbol{x} \sim f(\boldsymbol{y} | \boldsymbol{x})$ that is conditional on the unknown parameters. For the actually acquired data $\boldsymbol{Y} = \boldsymbol{y}$, the *likelihood function* $\mathcal{L}(\boldsymbol{x}) = f(\boldsymbol{y} | \boldsymbol{x})$ is defined by interpreting the conditional density $f(\boldsymbol{y} | \boldsymbol{x})$ as a function of the unknowns $\boldsymbol{x}$.

Given this setup, one can formulate an updated probability density $\pi(\boldsymbol{x} | \boldsymbol{y})$ of the unknowns that is conditioned on the realized data. This so-called *posterior density* results from Bayes' law

$$\pi(\boldsymbol{x} | \boldsymbol{y}) = \frac{\mathcal{L}(\boldsymbol{x}) \pi(\boldsymbol{x})}{Z}. \tag{1}$$

It completely summarizes the available information about the unknowns after the data have been analyzed. The *model evidence* Z properly normalizes the posterior density. It can be written as

$$Z = \int_{\mathcal{D}_{\boldsymbol{x}}} \mathcal{L}(\boldsymbol{x}) \, \pi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{2}$$

One is often interested in the marginals and moments of the posterior. The posterior marginal $\pi(x_j | \boldsymbol{y})$ of a single unknown x_j with $j \in \{1, \ldots, M\}$ is defined as

$$\pi(x_j | \boldsymbol{y}) = \int_{\mathcal{D}_{\boldsymbol{x}_{\sim j}}} \pi(\boldsymbol{x} | \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x}_{\sim j}. \tag{3}$$

Here, the simplifying notation $\boldsymbol{x}_{\sim j} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_M)^\top$ is introduced. For the mean $\mathbb{E}[\boldsymbol{X} | \boldsymbol{y}]$ and the covariance $\mathrm{Cov}[\boldsymbol{X} | \boldsymbol{y}]$ of the posterior one has

$$\mathbb{E}[\boldsymbol{X} | \boldsymbol{y}] = \int_{\mathcal{D}_{\boldsymbol{x}}} \boldsymbol{x} \, \pi(\boldsymbol{x} | \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x}, \tag{4}$$

$$\mathrm{Cov}[\boldsymbol{X} | \boldsymbol{y}] = \int_{\mathcal{D}_{\boldsymbol{x}}} (\boldsymbol{x} - \mathbb{E}[\boldsymbol{X} | \boldsymbol{y}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{X} | \boldsymbol{y}])^\top \pi(\boldsymbol{x} | \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x}. \tag{5}$$

3

More generally, Bayesian inference focuses the computation of posterior expectation values of the *quantities of interest* (QoI) $h \colon \mathcal{D}_{\boldsymbol{x}} \to \mathbb{R}$. These expectations may be formally expressed as

$$\mathbb{E}[h(\boldsymbol{X})|\boldsymbol{y}] = \int\limits_{\mathcal{D}_{\boldsymbol{x}}} h(\boldsymbol{x})\,\pi(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}. \tag{6}$$

For later considerations, it is remarked that this integration over the posterior density can be interpreted as a reweighted integration over the prior density

$$\mathbb{E}[h(\boldsymbol{X})|\boldsymbol{y}] = \int\limits_{\mathcal{D}_{\boldsymbol{x}}} h(\boldsymbol{x})\frac{\mathcal{L}(\boldsymbol{x})}{Z}\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \frac{1}{Z}\mathbb{E}[h(\boldsymbol{X})\mathcal{L}(\boldsymbol{X})]. \tag{7}$$

2.1. Bayesian inverse problems

The Bayesian framework described above can be applied to a vast range of scenarios from classical statistics [49, 50] and inverse modeling [51, 52]. In inverse problems, a so-called *forward model* $\mathcal{M}$ establishes a mathematical representation of the physical system under consideration. It is the function

$$\begin{aligned} \mathcal{M} \colon \mathcal{D}_{\boldsymbol{x}} &\to \mathbb{R}^N \\ \boldsymbol{x} &\mapsto \mathcal{M}(\boldsymbol{x}) \end{aligned} \tag{8}$$

that maps model inputs $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}} \subset \mathbb{R}^M$ to outputs $\tilde{\boldsymbol{y}} = \mathcal{M}(\boldsymbol{x}) \in \mathbb{R}^N$. *Inversion* is the process of inferring the unknown forward model parameters $\boldsymbol{x}$ with the measured data $\boldsymbol{y}$ of its response.

A probabilistic model of the observables is commonly constructed supposing that they can be represented as the sum $\boldsymbol{Y} = \tilde{\boldsymbol{Y}} + \boldsymbol{E}$ of the model response vector $\tilde{\boldsymbol{Y}} = \mathcal{M}(\boldsymbol{X}) \colon \Omega \to \mathbb{R}^N$ and a random vector $\boldsymbol{E} \colon \Omega \to \mathbb{R}^N$. The latter accounts for measurement noise and forward model inadequacy. It is assumed that the *residual vector* $\boldsymbol{E}$ is statistically independent from $\boldsymbol{X}$. An unknown realization $\boldsymbol{E} = \boldsymbol{\varepsilon}$ measures the discrepancy between the actually measured data $\boldsymbol{y} = \tilde{\boldsymbol{y}} + \boldsymbol{\varepsilon}$ and the model response $\tilde{\boldsymbol{y}} = \mathcal{M}(\boldsymbol{x})$ at the true value $\boldsymbol{x}$. Typically one starts from the premise that the residual $\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{\varepsilon}|\boldsymbol{0}, \boldsymbol{\Sigma})$ follows a Gaussian distribution. Here, $\boldsymbol{\Sigma}$ is a positive-definite covariance matrix. The observational model is then simply given as $\boldsymbol{Y}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{y}|\mathcal{M}(\boldsymbol{x}), \boldsymbol{\Sigma})$. For the likelihood this implies

$$\mathcal{L}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}\big(\boldsymbol{y} - \mathcal{M}(\boldsymbol{x})\big)^{\top} \boldsymbol{\Sigma}^{-1} \big(\boldsymbol{y} - \mathcal{M}(\boldsymbol{x})\big)\right). \tag{9}$$

For the actually acquired data $\boldsymbol{y}$ this is understood as a function of the unknowns $\boldsymbol{x}$. The Bayesian solution to the inverse problem posed is then the posterior in Eq. (1) where the likelihood is given as in Eq. (9). It summarizes the collected information about the unknown forward model inputs.

2.2. Markov chain Monte Carlo

Apart from some exceptional cases, the posterior density in Eq. (1) does not exhibit a closed-form expression. Thus one settles either for computing expectation values under the posterior or for sampling from the posterior. The former can be accomplished through stochastic integration techniques such as *Monte Carlo* (MC) [53] or *importance sampling* [54]. For the latter one usually has to resort to *Markov chain Monte Carlo* (MCMC) sampling [9, 10]. An ergodic Markov chain $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots$ over the support $\mathcal{D}_{\boldsymbol{x}}$ is constructed in such a way that the posterior arises as the invariant distribution

$$\pi(\boldsymbol{x}^{(t+1)}|\boldsymbol{y}) = \int\limits_{\mathcal{D}_{\boldsymbol{x}}} \pi(\boldsymbol{x}^{(t)}|\boldsymbol{y})\,\mathcal{K}(\boldsymbol{x}^{(t)}, \boldsymbol{x}^{(t+1)})\,\mathrm{d}\boldsymbol{x}^{(t)}. \tag{10}$$

Here, $\mathcal{K}(\boldsymbol{x}^{(t)}, \boldsymbol{x}^{(t+1)})$ denotes the density of the transition probability from the state $\boldsymbol{x}^{(t)}$ of the Markov chain at a time t to its state $\boldsymbol{x}^{(t+1)}$ at time $t+1$. The *Metropolis-Hastings* (MH) algorithm [55, 56] suggests an easy

principle for the construction of the Markov kernel $\mathcal{K}$ that satisfies Eq. (10). It is based on sampling candidates from a proposal distribution and a subsequent accept/reject decision. The transition kernel defined this ways satisfies detailed balance, i.e. time reversibility $\pi(\boldsymbol{x}^{(t)}|\boldsymbol{y})\,\mathcal{K}(\boldsymbol{x}^{(t)},\boldsymbol{x}^{(t+1)}) = \pi(\boldsymbol{x}^{(t+1)}|\boldsymbol{y})\,\mathcal{K}(\boldsymbol{x}^{(t+1)},\boldsymbol{x}^{(t)})$. This is a sufficient condition for Eq. (10) to apply. In practice, one initializes the Markov chain at some $\boldsymbol{x}^{(1)} \in \mathcal{D}_{\boldsymbol{x}}$ and then iteratively applies the MH updates from $\boldsymbol{x}^{(t)}$ to $\boldsymbol{x}^{(t+1)}$ for a finite number of times T. The *ergodic theorem* then ensures that one can approximate the population average in Eq. (6) in an asymptotically consistent way as the time average

$$\mathbb{E}[h(\boldsymbol{X})|\boldsymbol{y}] \approx \frac{1}{T}\sum_{t=1}^{T} h(\boldsymbol{x}^{(t)}). \tag{11}$$

A whole string a unpleasant consequences is entailed by the fact that MCMC updates are typically local and serially dependent. The quality of the posterior approximation is governed by the MCMC sample autocorrelation. In order to ensure an efficient posterior exploration one has to carefully design and tune the proposal distribution. This is an extremely tedious and problem-dependent task. Yet, even for comparably efficient MCMC updates, a large number of MCMC iterations may be required in order to achieve an acceptable degree of fidelity of the final results. In inverse modeling this requires an even larger number of serial forward solves which can be prohibitively expensive for demanding models. Another intrinsic MCMC weakness is that it lacks a clear convergence and stopping criterion, i.e. for diagnosing when the chain has forgotten its initialization and has converged to the target distribution in Eq. (10), and for the assessment of when the MC error in Eq. (11) has become sufficiently small. Even though there are more or less sophisticated convergence diagnostics [57, 58], those heuristic checks may very well fail, e.g. when separated posterior modes have not yet been detected.

The model evidence in Eq. (2) is important in the context of model comparison and selection [59]. In engineering applications it often happens that one wants to judge the performance of various competing models against measured data [60, 61]. While in variational Bayesian inference at least a lower bound of the model evidence is implicitly computed as a side product, in the MH algorithm it is not computed at all. Avoiding the explicit computation of the model evidence is beneficial for parameter estimation, but it does not allow for model selection. To this effect one has to rely on dedicated methods [62, 63].

3. Surrogate forward modeling

In the analysis of engineering systems it has become a widespread practice to substitute expensive computer models with inexpensive *metamodels* or *surrogate models*. Those approximations mimic the functional relationship between the inputs and the outputs of the original model in Eq. (8). Metamodeling promises significant gains in situations that require a large number of forward model runs, e.g. for optimization problems, uncertainty analysis and inverse modeling. Important classes of metamodels are based on Gaussian process models or Kriging [64, 65] and polynomial chaos expansions [66]. More recent introductions to these subjects can be found in [67, 68] and [69, 70], respectively. Nowadays, the application of Kriging [16, 17] and polynomial chaos surrogates [18–20] is commonplace in Bayesian inverse problems.

Next we focus on polynomial chaos metamodels. The idea is to decompose the forward model response into polynomial terms that are orthogonal with respect to a weight function. In stochastic analysis this weight is often identified with a probability distribution in order to facilitate uncertainty propagation. In inverse analysis it is commonly equated with the prior in order to enhance MCMC posterior sampling. The formalism of polynomial chaos expansions is rooted in spectral methods and functional approximations with orthogonal polynomials. Hence, the function space point of view is emphasized in this section. We also concentrate on linear least squares for the practical computation of the expansions coefficients.

3.1. L^2_π function space

From here on it is assumed that the components of the uncertain parameter vector $\boldsymbol{X} = (X_1, \ldots, X_M)^\top$ are independent random variables X_i. Thus their joint density can be written as

$$\pi(\boldsymbol{x}) = \prod_{i=1}^{M} \pi_i(x_i). \tag{12}$$

Let $L^2_\pi(\mathcal{D}_{\boldsymbol{x}}) = \{u \colon \mathcal{D}_{\boldsymbol{x}} \to \mathbb{R} \,|\, \int_{\mathcal{D}_{\boldsymbol{x}}} u^2(\boldsymbol{x})\,\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} < \infty\}$ be the Hilbert space of functions that are square integrable with respect to the prior density in Eq. (12). For $u, v \in L^2_\pi(\mathcal{D}_{\boldsymbol{x}})$ a weighted inner product $\langle \cdot, \cdot \rangle_{L^2_\pi}$ and its associated norm $\|\cdot\|_{L^2_\pi}$ are defined as

$$\langle u, v \rangle_{L^2_\pi} = \int_{\mathcal{D}_{\boldsymbol{x}}} u(\boldsymbol{x})v(\boldsymbol{x})\,\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{13}$$

$$\|u\|_{L^2_\pi} = \langle u, u \rangle_{L^2_\pi}^{1/2}. \tag{14}$$

Given that $u, v \in L^2_\pi(\mathcal{D}_{\boldsymbol{x}})$, the real-valued random variables $u(\boldsymbol{X}), v(\boldsymbol{X}) \colon \Omega \to \mathbb{R}$ on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ have a finite variance. One can then write the inner product in Eq. (13) as the expectation

$$\langle u, v \rangle_{L^2_\pi} = \mathbb{E}[u(\boldsymbol{X})v(\boldsymbol{X})]. \tag{15}$$

In the further course of the presentation, the identity in Eq. (15) is frequently used in order to switch back and forth between expectation values under the prior distribution and weighted inner products.

3.2. Orthonormal polynomials

Now a basis of the space $L^2_\pi(\mathcal{D}_{\boldsymbol{x}})$ is constructed with orthogonal polynomials [71–73]. Let $\{\Psi^{(i)}_{\alpha_i}\}_{\alpha_i \in \mathbb{N}}$ be a family of univariate polynomials in the input variable $x_i \in \mathcal{D}_{x_i}$. Each member is characterized by its polynomial degree $\alpha_i \in \mathbb{N}$. The polynomials are required to be orthonormal in the sense that

$$\left\langle \Psi^{(i)}_{\alpha_i}, \Psi^{(i)}_{\beta_i} \right\rangle_{L^2_{\pi_i}} = \delta_{\alpha_i \beta_i} = \begin{cases} 1 & \text{if } \alpha_i = \beta_i, \\ 0 & \text{if } \alpha_i \neq \beta_i. \end{cases} \tag{16}$$

These polynomials form a complete orthogonal system in $L^2_{\pi_i}(\mathcal{D}_{x_i})$. Next, a set of multivariate polynomials $\{\Psi_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{N}^M}$ in the input variables $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}}$ is constructed as the tensor product

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \prod_{i=1}^{M} \Psi^{(i)}_{\alpha_i}(x_i). \tag{17}$$

Here, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M) \in \mathbb{N}^M$ is a multi-index that characterizes the polynomials. By construction, namely due to Eqs. (12), (16) and (17), they are orthonormal in the sense that

$$\langle \Psi_{\boldsymbol{\alpha}}, \Psi_{\boldsymbol{\beta}} \rangle_{L^2_\pi} = \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}} = \begin{cases} 1 & \text{if } \boldsymbol{\alpha} = \boldsymbol{\beta}, \\ 0 & \text{if } \boldsymbol{\alpha} \neq \boldsymbol{\beta}. \end{cases} \tag{18}$$

These polynomials establish a complete orthogonal basis in $L^2_\pi(\mathcal{D}_{\boldsymbol{x}})$. Note that the constant term is always given as $\Psi^{(i)}_0 = 1$ in the univariate case. This ensures the proper normalization $\|\Psi^{(i)}_0\|_{L^2_{\pi_i}} = 1$. In the multivariate case one similarly has $\Psi_{\boldsymbol{0}} = 1$ with $\|\Psi_{\boldsymbol{0}}\|_{L^2_\pi} = 1$.

3.3. Hermite and Legendre polynomials

Two classical univariate families are the *Hermite* and the *Legendre polynomials* $\{H_{\alpha_i}(x_i)\}_{\alpha_i \in \mathbb{N}}$ for $x_i \in \mathbb{R}$ and $\{P_{\alpha_i}(x_i)\}_{\alpha_i \in \mathbb{N}}$ for $x_i \in [-1, 1]$, respectively. The former are orthogonal with respect to the weight function $\mathcal{N}(x_i | 0, 1) = (2\pi)^{-1/2} \exp(-x_i^2/2)$, the latter with respect to $\mathcal{U}(x_i | -1, 1) = \mathrm{I}_{[-1,1]}(x_i)/2$. Here, $\mathrm{I}_{[-1,1]}$ denotes the indicator function of the interval $[-1, 1]$. A short summary of these two univariate families is given in Table 1. Over the respective domains, their first members are defined as given in the Appendix A. Classical orthogonal polynomials $\{\psi_{\alpha_i}^{(i)}(x_i)\}_{\alpha_i \in \mathbb{N}}$ are typically not normalized, e.g. the aforementioned Hermite or Legendre families. An orthonormal family $\{\Psi_{\alpha_i}^{(i)}\}_{\alpha_i \in \mathbb{N}}$ is then obtained through an appropriate normalization with $\Psi_{\alpha_i}^{(i)} = \psi_{\alpha_i}^{(i)}/\|\psi_{\alpha_i}^{(i)}\|_{L^2_{\pi_i}}$.

Table 1: Two families of orthogonal polynomials.

Input type	Polynomials	$\mathcal{D}_{x_i}$	$\pi_i(x_i)$	$\psi_{\alpha_i}^{(i)}(x_i)$	$\|\psi_{\alpha_i}^{(i)}\|_{L^2_{\pi_i}}$
Gaussian	Hermite	$\mathbb{R}$	$\mathcal{N}(x_i | 0, 1)$	$H_{\alpha_i}(x_i)$	$\sqrt{\alpha_i!}$
Uniform	Legendre	$[-1, 1]$	$\mathcal{U}(x_i | -1, 1)$	$P_{\alpha_i}(x_i)$	$\sqrt{1/(2\alpha_i + 1)}$

In practice, the parameter space $\mathcal{D}_{\boldsymbol{x}}$ and the input distribution $\pi(\boldsymbol{x})$ are often not directly suitable for an expansion based on the two standardized families in Table 1. One possibility is then to employ suitably chosen or constructed polynomials [74, 75]. Another possibility is to use an invertible function $\mathcal{T} \colon \mathbb{R}^M \to \mathbb{R}^M$, sufficiently well-behaved and as linear as possible, in order to transform the physical variables $\boldsymbol{x}$ into *standardized variables* $\boldsymbol{\xi} = \mathcal{T}(\boldsymbol{x})$, i.e. the image $\mathcal{D}_{\boldsymbol{\xi}} = \mathcal{T}(\mathcal{D}_{\boldsymbol{x}})$ and the transformed weight function $\pi_{\mathcal{T}}(\boldsymbol{\xi}) = \pi(\mathcal{T}^{-1}(\boldsymbol{\xi})) |\det J_{\mathcal{T}^{-1}}(\boldsymbol{\xi})|$ are of a standard form. Here, $J_{\mathcal{T}^{-1}} = \mathrm{d}\mathcal{T}^{-1}/\mathrm{d}\boldsymbol{\xi}$ is the Jacobian matrix. If such a change of variables is needed, the considerations that are given below for $\mathcal{M}$ and h in the variables $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}}$ can be straightforwardly repeated for $\mathcal{M} \circ \mathcal{T}^{-1}$ and $h \circ \mathcal{T}^{-1}$ in the variables $\boldsymbol{\xi} \in \mathcal{D}_{\boldsymbol{\xi}}$. In this case, the expectation in Eq. (6) follows the integration by substitution

$$\mathbb{E}[h(\boldsymbol{X})|\boldsymbol{y}] = \frac{1}{Z} \int_{\mathcal{D}_{\boldsymbol{\xi}}} h\left(\mathcal{T}^{-1}(\boldsymbol{\xi})\right) \mathcal{L}\left(\mathcal{T}^{-1}(\boldsymbol{\xi})\right) \pi_{\mathcal{T}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}. \tag{19}$$

3.4. Polynomial chaos expansions

For simplicity, herein the presentation is restricted to scalar-valued models $\mathcal{M} \colon \mathcal{D}_{\boldsymbol{x}} \to \mathbb{R}$. The extension to the multivariate case is straightforward. It is supposed that the forward model $\mathcal{M} \in L^2_\pi(\mathcal{D}_{\boldsymbol{x}})$ is mean-square integrable. This is a reasonable assumption as seen from a physical perspective. In $L^2_\pi(\mathcal{D}_{\boldsymbol{x}})$ the *generalized Fourier expansion* of $\mathcal{M}$ in terms of the orthogonal polynomials $\{\Psi_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{N}^M}$ is then given as

$$\mathcal{M} = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}, \quad \text{with} \tag{20}$$

$$a_{\boldsymbol{\alpha}} = \langle \mathcal{M}, \Psi_{\boldsymbol{\alpha}} \rangle_{L^2_\pi} = \int_{\mathcal{D}_{\boldsymbol{x}}} \mathcal{M}(\boldsymbol{x}) \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \, \pi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{21}$$

The *generalized Fourier coefficients* $\{a_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{N}^M}$ of the series expansion in Eq. (20) are defined as the orthogonal projection of $\mathcal{M}$ onto the basis elements in Eq. (21). The corresponding Fourier series of the second-order random variable $\tilde{Y} = \mathcal{M}(\boldsymbol{X})$ on $(\Omega, \mathcal{F}, \mathcal{P})$ is a so-called *polynomial chaos expansion* (PCE) $\mathcal{M}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$.

PCEs have been popularized in the context of uncertainty propagation where the goal is the quantification of the distribution of $\tilde{Y} = \mathcal{M}(\boldsymbol{X})$. For this purpose it comes in handy that the mean and the variance of this random variable can be easily determined from its PCE coefficients. Indeed, with Eq. (18) it is easy to

verify that they are simply given as

$$\mathbb{E}[\mathcal{M}(\boldsymbol{X})] = \left\langle \Psi_{\boldsymbol{0}}, \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}} \right\rangle_{L_\pi^2} = a_{\boldsymbol{0}}, \tag{22}$$

$$\mathrm{Var}[\mathcal{M}(\boldsymbol{X})] = \left\| \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}} - a_{\boldsymbol{0}} \Psi_{\boldsymbol{0}} \right\|_{L_\pi^2}^2 = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M \setminus \{\boldsymbol{0}\}} a_{\boldsymbol{\alpha}}^2. \tag{23}$$

The simple identities in Eqs. (22) and (23) follow from the definitions of the inner product and the associated norm in Eqs. (13) and (14), respectively.

3.5. Truncated series

For a practical computation one has to truncate the infinite series in Eq. (20). Let the total degree of a multivariate polynomial $\Psi_{\boldsymbol{\alpha}}$ be defined as $\|\boldsymbol{\alpha}\|_1 = \sum_{i=1}^M |\alpha_i|$. A *standard truncation scheme* is then adopted by limiting the terms in Eq. (20) to the finite set of multi-indices

$$\mathcal{A}_p = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_1 \leq p \right\}. \tag{24}$$

This specifies a set of polynomials $\{\Psi_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{A}_p}$ such that their total degree $\|\boldsymbol{\alpha}\|_1$ is smaller than or equal to a chosen p. The total number of terms retained in the set $\mathcal{A}_p$ is given as

$$P = \binom{M+p}{p} = \frac{(M+p)!}{M!\,p!}. \tag{25}$$

The dramatic increase of the total number of terms P with the input dimensionality M and the maximal polynomial degree p, that is described by Eq. (25), is commonly referred to as the *curse of dimensionality*. A simple idea to limit the number of regressors relies on hyperbolic truncation sets. For $0 < q < 1$ a quasinorm is defined as $\|\boldsymbol{\alpha}\|_q = (\sum_{i=1}^M |\alpha_i^q|)^{1/q}$. The corresponding *hyperbolic truncation scheme* is then given as $\mathcal{A}_p^q = \{\boldsymbol{\alpha} \in \mathbb{N}^M | \|\boldsymbol{\alpha}\|_q \leq p\}$. Adopting the standard scheme in Eq. (24), a finite version of Eq. (20) can be written as

$$\hat{\mathcal{M}}_p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_p} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}). \tag{26}$$

In engineering problems, one uses $\hat{\mathcal{M}}_p(\boldsymbol{x})$ as a functional approximation of $\mathcal{M}(\boldsymbol{x})$, i.e. as a *polynomial response surface* [76]. This is justified because the approximation converges in the mean-square sense

$$\left\| \mathcal{M} - \hat{\mathcal{M}}_p \right\|_{L_\pi^2}^2 = \mathbb{E}\left[\left(\mathcal{M}(\boldsymbol{X}) - \hat{\mathcal{M}}_p(\boldsymbol{X}) \right)^2 \right] = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M \setminus \mathcal{A}_p} a_{\boldsymbol{\alpha}}^2 \to 0, \quad \text{for } p \to \infty. \tag{27}$$

The rate of the convergence in Eq. (27) depends on the regularity of $\mathcal{M}$. On top of that, the response surface in Eq. (26) is also optimal in the mean-square sense

$$\left\| \mathcal{M} - \hat{\mathcal{M}}_p \right\|_{L_\pi^2}^2 = \inf_{\mathcal{M}^\star \in \mathbb{P}_p} \|\mathcal{M} - \mathcal{M}^\star\|_{L_\pi^2}^2, \quad \text{where } \mathbb{P}_p = \mathrm{span}\left(\{\Psi_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{A}_p} \right). \tag{28}$$

According to Eq. (28), the response surface in Eq. (26) minimizes the mean-square error over the space of polynomials $\mathbb{P}_p = \mathrm{span}(\{\Psi_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{A}_p})$ having a total degree of at most p.

3.6. Least squares

In order to find a metamodel of the form Eq. (26), one computes approximations of the exact expansion coefficients in Eq. (21). Broadly speaking, one distinguishes between *intrusive* and *non-intrusive* computations. While the former class of techniques is based on manipulations of the governing equations, the

8

latter is exclusively build upon calls to the forward model at chosen input values. Stochastic Galerkin methods belong to the class of intrusive techniques [77, 78], whereas stochastic collocation [79, 80] and projection through numerical quadrature [81, 82] are non-intrusive approaches. Herein we focus on another non-intrusive formulation that is based on least squares regression analysis [83, 84]. This formulation is based on linear least squares [37, 38] and related ideas from statistical learning theory [39, 40]. Since this includes sparsity-promoting fitting techniques from high-dimensional statistics [85, 86], recently least squares projection methods receive considerable attention. This includes frequentist [87–90] and Bayesian implementations [91–94] of shrinkage estimators. Current results regarding the convergence behavior of such regression methods can be found in [95–97].

We introduce a simplifying vector notation such that $\boldsymbol{a} = (a_1, \ldots, a_P)^\top$ and $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_P)^\top$ gather and order the coefficients and the polynomials for all $\boldsymbol{\alpha} \in \mathcal{A}_p$. For the truncated expression in Eq. (26) one thus has $\hat{\mathcal{M}}_p = \boldsymbol{a}^\top \boldsymbol{\Psi}$. The problem of finding $\hat{\mathcal{M}}_p \in \mathbb{P}_p$ that minimizes the mean-square error in Eq. (28) may then be equivalently rephrased as

$$\boldsymbol{a} = \underset{\boldsymbol{a}^\star \in \mathbb{R}^P}{\arg\min} \, \mathbb{E}\left[\left(\mathcal{M}(\boldsymbol{X}) - \boldsymbol{a}^{\star\top} \boldsymbol{\Psi}(\boldsymbol{X})\right)^2\right]. \tag{29}$$

The stochastic optimization objective in Eq. (29) establishes an alternative to the orthogonal projection in Eq. (21). This formulation may be more amenable to a numerical computation. At the very least it allows one the draw on the machinery of linear least squares in order to compute an approximation $\hat{\boldsymbol{a}}$ of the exact coefficients $\boldsymbol{a}$. To that end one discretizes Eq. (29) as

$$\hat{\boldsymbol{a}} = \underset{\boldsymbol{a}^\star \in \mathbb{R}^P}{\arg\min} \, \frac{1}{K} \sum_{k=1}^{K} \left(\mathcal{M}(\boldsymbol{x}^{(k)}) - \boldsymbol{a}^{\star\top} \boldsymbol{\Psi}(\boldsymbol{x}^{(k)})\right)^2. \tag{30}$$

Here, the *experimental design* $\mathcal{X} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(K)})$ is a representative sample of K forward model inputs, e.g. randomly drawn from the input distribution in Eq. (12). It is then required to compute the corresponding model responses $\mathcal{Y} = (\mathcal{M}(\boldsymbol{x}^{(1)}), \ldots, \mathcal{M}(\boldsymbol{x}^{(K)}))^\top$ in K training runs.

Now let $\boldsymbol{A} \in \mathbb{R}^{K \times P}$ be the matrix with entries $A_{k,l} = \Psi_l(\boldsymbol{x}^{(k)})$ for $k = 1, \ldots, K$ and $l = 1, \ldots, P$. Moreover, let the system $\mathcal{Y} = \boldsymbol{A}\hat{\boldsymbol{a}}$ be overdetermined with $K \geq P$. The linear least squares problem in Eq. (30) may then be written as

$$\hat{\boldsymbol{a}} = \underset{\boldsymbol{a}^\star \in \mathbb{R}^P}{\arg\min} \|\mathcal{Y} - \boldsymbol{A}\boldsymbol{a}^\star\|^2 \tag{31}$$

The *normal equations* $(\boldsymbol{A}^\top \boldsymbol{A})\hat{\boldsymbol{a}} = \boldsymbol{A}^\top \mathcal{Y}$ establish the first-order condition for Eq. (31) to apply. Given that the matrix $\boldsymbol{A}^\top \boldsymbol{A}$ is non-singular, this linear system is solved by

$$\hat{\boldsymbol{a}} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \mathcal{Y}. \tag{32}$$

The positive definiteness of $\boldsymbol{A}^\top \boldsymbol{A}$, i.e. the columns of $\boldsymbol{A}$ are linearly independent, is the second-order condition for Eq. (32) to be the minimum of Eq. (31). The *ordinary least squares* (OLS) solution Eq. (32) is commonly computed by means of linear algebra methods. An alternative to OLS is *least angle regression* (LAR) [98, 99]. LAR is well-suited for high-dimensional PCE regression problems [87] and can be even applied in the underdetermined case. It is based on selecting only the most dominant regressors from a possibly large candidate set. The resulting predictor is thus sparse as compared to the OLS solution.

3.7. Prediction errors

After the computation of a metamodel, one typically wants to assess its prediction accuracy. Moreover, when a number of candidate surrogates is computed, one wants to compare their performances in order to eventually select the best. Hence, one needs to define an appropriate criterion that allows for an accurate and efficient quantification of the approximation errors. The natural measure of the mismatch between the forward model $\mathcal{M}$ and an approximation $\hat{\mathcal{M}}_p$ is the *generalization error* $E_{\mathrm{Gen}} = \mathbb{E}[(\mathcal{M}(\boldsymbol{X}) - \hat{\mathcal{M}}_p(\boldsymbol{X}))^2]$. This is exactly the error the minimization of which is posed by Eq. (29). Since it remains unknown, it cannot

be used as a performance measure. One could estimate E_{Gen} based on MC simulation, though. However, this is not very efficient since it requires the execution of additional forward model runs. In contrast, the *empirical error* $E_{\text{Emp}} = K^{-1} \sum_{k=1}^{K} (\mathcal{M}(\boldsymbol{x}^{(k)}) - \hat{\mathcal{M}}_p(\boldsymbol{x}^{(k)}))^2$ is the quantity that is practically minimized according to Eq. (30). This error indicator is obtained for free, however, it does not account for overfitting and thus tends to severely underestimate the real generalization error E_{Gen}.

In order to construct an estimate of E_{Gen} that is more efficient than the MC estimate and more accurate than E_{Emp}, one sometimes resorts to leave-one-out (LOO) cross validation [100]. Let $\hat{\mathcal{M}}_{\sim k}$ be the surrogate model that is obtained from the reduced experimental design $\mathcal{X}_{\sim k} = (\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k-1)}, \boldsymbol{x}^{(k+1)}, \dots, \boldsymbol{x}^{(K)})$, i.e. a single input $\boldsymbol{x}^{(k)}$ has been dropped. The *LOO error* is then defined as

$$E_{\text{LOO}} = \frac{1}{K} \sum_{k=1}^{K} \left(\mathcal{M}(\boldsymbol{x}^{(k)}) - \hat{\mathcal{M}}_{\sim k}(\boldsymbol{x}^{(k)}) \right)^2 . \tag{33}$$

Without the need for re-running the forward model, this error allows for a fair assessment of how well the performance of a metamodel $\hat{\mathcal{M}}_p$ generalizes beyond the used experimental design. Yet, Eq. (33) calls for conducting K separate regressions for finding $\hat{\mathcal{M}}_{\sim k}$ with an experimental design of the size $K - 1$. A remarkably simple result from linear regression analysis states that E_{LOO} can be also computed as

$$E_{\text{LOO}} = \frac{1}{K} \sum_{k=1}^{K} \left(\frac{\mathcal{M}(\boldsymbol{x}^{(k)}) - \hat{\mathcal{M}}_p(\boldsymbol{x}^{(k)})}{1 - h_k} \right)^2 . \tag{34}$$

Here, $\hat{\mathcal{M}}_p$ is computed from the full experimental design $\mathcal{X}$ and h_k denotes the k-th diagonal entry of the matrix $\boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top$. This is more efficient since it does not require repeated fits. A derivation of the formula in Eq. (34) can be found in [101].

One may define $\epsilon_{\text{Emp}} = E_{\text{Emp}}/\text{Var}[\mathcal{Y}]$ and $\epsilon_{\text{LOO}} = E_{\text{LOO}}/\text{Var}[\mathcal{Y}]$ as normalized versions of the empirical and the LOO error, respectively. Here, $\text{Var}[\mathcal{Y}]$ is the empirical variance of the response sample $\mathcal{Y}$. These normalized errors can be used in order to judge and compare the performance of metamodels. In turn, this enables a practical convergence analysis, e.g. by monitoring the errors over repeated metamodel computations for an increasing experimental design size K and expansion order p.

4. Spectral Bayesian inference

Different types of probability density approximations are encountered in statistical inference. This includes series expansions for the population density of a random data sample in nonparametric distribution fitting. Here, the unknown density of the data is either directly represented as a linear combination of polynomial basis functions [102] or as the product of a base density times a superposition of polynomials [103]. The latter type of expansion is also encountered in parametric density estimation of random data where one-dimensional posterior densities of the unknown parameter are expanded about a Gaussian baseline. This can be based on a Taylor sum at a maximum likelihood estimate [104, 105] or on Stein's lemma and numerical integration [106, 107]. Moreover, different types of likelihood approximations are encountered in inverse modeling. This includes direct approaches where the likelihood is approximated itself [108, 109] and indirect methods where the likelihood is approximated based on a surrogate of the forward model [18–20]. These techniques facilitate Bayesian inference within the limits of MCMC sampling.

By linking likelihood approximations to density expansions, now we present a spectral formulation of Bayesian inference which targets the emulation of the posterior density. Based on the theoretical and computational machinery of PCEs, the likelihood function itself is decomposed into polynomials that are orthogonal with respect to the prior distribution. This spectral likelihood expansion enables semi-analytic Bayesian inference. Simple formulas are derived for the joint posterior density and its marginals. They are regarded as expansions of the posterior about the prior as the reference density. The model evidence is shown to be the coefficient of the constant expansion term. General QoI-expectations under the posterior and the first posterior moments are obtained through a mere postprocessing of the spectral coefficients. After a discussion of the advantages and shortcomings of the spectral method, a change of the reference density is proposed in order to improve the efficacy.

10

4.1. Spectral likelihood expansions

The authors C. Soize and R. Ghanem start their paper [110] with the following sentence: "Characterizing the membership of a mathematical function in the most suitable functional space is a critical step toward analyzing it and identifying sequences of efficient approximants to it." As discussed in Section 2, given the data $\boldsymbol{y}$ and the statistical model $f(\boldsymbol{y}|\boldsymbol{x})$, the likelihood is the function

$$
\begin{aligned}
\mathcal{L} \colon \mathcal{D}_{\boldsymbol{x}} &\to \mathbb{R}^+ \\
\boldsymbol{x} &\mapsto f(\boldsymbol{y}|\boldsymbol{x}).
\end{aligned}
\tag{35}
$$

It maps the parameter space $\mathcal{D}_{\boldsymbol{x}}$ into the set of non-negative real numbers $\mathbb{R}^+$. At this place we assume that the likelihood $\mathcal{L} \in L_\pi^2(\mathcal{D}_{\boldsymbol{x}})$ is square integrable with respect to the prior. From a statistical point of view, this is a reasonable supposition which is necessary in order to invoke the theory of Section 3. On condition that the likelihood is bounded from above, the mean-square integrability follows immediately from the axioms of probability. Note that maximum likelihood estimates (MLE) implicitly rest upon this presumption. If $\boldsymbol{x}_{\mathrm{MLE}} \in \arg\max_{\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}}} \mathcal{L}(\boldsymbol{x})$ is a MLE, i.e. for all $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}}$ it applies that $\mathcal{L}(\boldsymbol{x}) \leq \mathcal{L}(\boldsymbol{x}_{\mathrm{MLE}}) < \infty$, then one trivially has $\int_{\mathcal{D}_{\boldsymbol{x}}} \mathcal{L}^2(\boldsymbol{x})\,\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \leq \mathcal{L}^2(\boldsymbol{x}_{\mathrm{MLE}}) \int_{\mathcal{D}_{\boldsymbol{x}}} \pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \mathcal{L}^2(\boldsymbol{x}_{\mathrm{MLE}}) < \infty$.

Having identified $L_\pi^2(\mathcal{D}_{\boldsymbol{x}})$ as a suitable function space for characterizing the likelihood, allows one to represent the likelihood with respect to the orthonormal basis $\{\Psi_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{N}^M}$. This representation is

$$
\mathcal{L} = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}, \quad \text{with}
\tag{36}
$$

$$
b_{\boldsymbol{\alpha}} = \langle \mathcal{L}, \Psi_{\boldsymbol{\alpha}} \rangle_{L_\pi^2} = \int_{\mathcal{D}_{\boldsymbol{x}}} \mathcal{L}(\boldsymbol{x}) \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x})\, \pi(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}.
\tag{37}
$$

We refer to Eqs. (36) and (37) as a *spectral likelihood expansion* (SLE). Notice that the SLE coefficients $\{b_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{N}^M}$ are data-dependent. This reflects the fact that the likelihood in Eq. (35) depends on the data. With the truncation scheme in Eq. (24), one can limit the infinite series in Eq. (36) to the finite number of terms for which $\boldsymbol{\alpha} \in \mathcal{A}_p$. A mean-square convergent response surface of the likelihood is then given as

$$
\hat{\mathcal{L}}_p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_p} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}).
\tag{38}
$$

For the time being we assume that the coefficients of the SLE in Eq. (36) or its response surface in Eq. (38) are already known. One can then accomplish Bayesian inference by extracting the joint posterior density or a *posterior density surrogate*, its marginals and the corresponding QoI-expectations directly from the SLE.

4.2. Joint posterior density

We begin with the joint posterior density function and the model evidence. By plugging Eq. (36) in Eq. (1) one simply obtains the "nonparametric" expression

$$
\pi(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z} \left(\sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \right) \pi(\boldsymbol{x}).
\tag{39}
$$

Due to the orthonormality of the basis, the model evidence is simply found as the coefficient of the constant SLE term. This is easily verified by writing Eq. (2) as

$$
Z = \langle 1, \mathcal{L} \rangle_{L_\pi^2} = \left\langle \Psi_{\mathbf{0}}, \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}} \right\rangle_{L_\pi^2} = b_{\mathbf{0}}.
\tag{40}
$$

11

The remarkably simple result in Eq. (40) completes the expression of the posterior density in Eq. (39). It is interesting to note that the posterior density is of the form

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \pi(\boldsymbol{x}) \left(1 + \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M \setminus \{\boldsymbol{0}\}} b_{\boldsymbol{0}}^{-1} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \right). \tag{41}$$

In essence, the posterior density is here represented as a "perturbation" around the prior. The latter establishes the leading term of the expansion and acts as the *reference density*. The expression in Eq. (41) is reminiscent of an *Edgeworth expansion* for a density function in asymptotic statistics [111–113].

4.3. Quantities of interest

Based on the joint posterior density one can calculate the corresponding QoI-expectations in Eq. (6). At this point, the identities in Eqs. (7) and (15) finally play a key role. They allow one to express and treat the posterior expectation of a QoI with $h \in L^2_\pi(\mathcal{D}_{\boldsymbol{x}})$ as the weighted projection onto the likelihood

$$\mathbb{E}[h(\boldsymbol{X})|\boldsymbol{y}] = \frac{1}{Z}\mathbb{E}[h(\boldsymbol{X})\mathcal{L}(\boldsymbol{X})] = \frac{1}{Z}\langle h, \mathcal{L}\rangle_{L^2_\pi}. \tag{42}$$

Let $h = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}$ with $c_{\boldsymbol{\alpha}} = \langle h, \Psi_{\boldsymbol{\alpha}}\rangle_{L^2_\pi}$ be the QoI representation in the polynomial basis used. The general posterior expectation in Eq. (6) follows then from Eq. (42) by *Parseval's theorem*

$$\mathbb{E}[h(\boldsymbol{X})|\boldsymbol{y}] = \frac{1}{b_{\boldsymbol{0}}} \left\langle \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}, \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}} \right\rangle_{L^2_\pi} = \frac{1}{b_{\boldsymbol{0}}} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} c_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}. \tag{43}$$

If the QoI $h \in \mathbb{P}_p$ is known by a multivariate monomial representation of finite order, then its representation in the orthogonal basis can always be recovered by a change of basis. Examples that relate to the first posterior moments are given shortly hereafter. In a more complex case the QoI is a computational model itself and one would have to numerically compute a PCE surrogate. While Eq. (22) facilitates the propagation of the prior uncertainty, Eq. (43) promotes the propagation of the posterior uncertainty.

4.4. Posterior marginals

Now the posterior marginals in Eq. (3) are derived. For some $j \in \{1, \ldots, M\}$ let us introduce the new set of multi-indices $\mathcal{A}^{(j)} = \{(\alpha_1, \ldots, \alpha_M) | \alpha_i = 0 \text{ for } i \neq j\}$. With a slight abuse of the notation, the *sub-expansion* of the SLE that only contains terms with $\boldsymbol{\alpha} \in \mathcal{A}^{(j)}$ is denoted as

$$\mathcal{L}_j(x_j) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}^{(j)}} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \sum_{\mu \in \mathbb{N}} b_\mu^{(j)} \Psi_\mu^{(j)}(x_j), \quad \text{where } b_\mu^{(j)} = b_{(0,\ldots,0,\mu,0,\ldots,0)}. \tag{44}$$

It collects all the polynomials that are constant in all variables $\boldsymbol{x}_{\sim j}$, i.e. they are non-constant in the single variable x_j only. In this sense the sub-expansion in Eq. (44) is virtually a function of x_j only. For the posterior marginal of the single unknown x_j in Eq. (3) one can derive

$$\pi(x_j|\boldsymbol{y}) = \frac{\pi_j(x_j)}{Z} \int_{\mathcal{D}_{\boldsymbol{x}_{\sim j}}} \mathcal{L}(\boldsymbol{x})\, \pi(\boldsymbol{x}_{\sim j})\, \mathrm{d}\boldsymbol{x}_{\sim j} = \frac{1}{b_{\boldsymbol{0}}} \mathcal{L}_j(x_j) \pi_j(x_j). \tag{45}$$

These equalities apply due to the independent prior $\pi(\boldsymbol{x}) = \pi(\boldsymbol{x}_{\sim j})\pi_j(x_j)$ in Eq. (12), the orthonormality of the univariate polynomials in Eq. (16) and the tensor structure of the multivariate ones in Eq. (17). For a pair $j, k \in \{1, \ldots, M\}$ with $j \neq k$ let us introduce yet another set of multi-indices $\mathcal{A}^{(j,k)} = \{(\alpha_1, \ldots, \alpha_M) | \alpha_i = 0 \text{ for } i \neq j, k\}$. The sub-expansion of the full SLE that only contains terms with $\boldsymbol{\alpha} \in \mathcal{A}^{(j,k)}$ is denoted as

$$\mathcal{L}_{j,k}(x_j, x_k) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}^{(j,k)}} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \sum_{\mu, \nu \in \mathbb{N}} b_{\mu,\nu}^{(j,k)} \Psi_\mu^{(j)}(x_j) \Psi_\nu^{(k)}(x_k),$$

$$\text{where } b_{\mu,\nu}^{(j,k)} = b_{(0,\ldots,0,\mu,0,\ldots,0,\nu,0,\ldots,0)}. \tag{46}$$

12

Since it only contains terms that are constant in $\boldsymbol{x}_{\sim j,k}$, the sub-expansion in Eq. (46) can be seen as a function of x_j and x_k. The posterior density can then be marginalized as follows

$$\pi(x_j, x_k | \boldsymbol{y}) = \int\limits_{\mathcal{D}_{\boldsymbol{x}_{\sim j,k}}} \pi(\boldsymbol{x} | \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x}_{\sim j,k} = \frac{1}{b_{\boldsymbol{0}}} \mathcal{L}_{j,k}(x_j, x_k) \pi_j(x_j) \pi_k(x_k). \tag{47}$$

Note that the dependency structure of $\pi(x_j, x_k | \boldsymbol{y})$ in Eq. (47) is induced by those terms of $\mathcal{L}_{j,k}(x_j, x_k)$ that are not present in $\mathcal{L}_j(x_j)$ and $\mathcal{L}_k(x_k)$, i.e. the terms $b_{\mu,\nu}^{(j,k)} \Psi_\mu^{(j)}(x_j) \Psi_\nu^{(k)}(x_k)$ with $\mu, \nu \neq 0$.

4.5. First posterior moments

With the two marginalizations of the posterior density in Eqs. (45) and (47) one can calculate the entries of the posterior mean in Eq. (4) and the covariance matrix in Eq. (5). Let $\{d_0^{(j)}, d_1^{(j)}\}$ be defined such that $x_j = d_0^{(j)} \Psi_0^{(j)}(x_j) + d_1^{(j)} \Psi_1^{(j)}(x_j)$. With this univariate representation and Eq. (45) one easily obtains

$$\mathbb{E}[X_j | \boldsymbol{y}] = \frac{1}{b_{\boldsymbol{0}}} \langle x_j, \mathcal{L}_j \rangle_{L_{\pi_j}^2} = \frac{1}{b_{\boldsymbol{0}}} \left(d_0^{(j)} b_0^{(j)} + d_1^{(j)} b_1^{(j)} \right). \tag{48}$$

Note that one actually has $b_0^{(j)} = b_{\boldsymbol{0}}$ in this notation. Diagonal entries of the covariance matrix in Eq. (5) can be similarly deduced. Let $\{e_0^{(j)}, e_1^{(j)}, e_2^{(j)}\}$ the coefficients of the univariate representation $(x_j - \mathbb{E}[X_j | \boldsymbol{y}])^2 = \sum_{\mu=0}^2 e_\mu^{(j)} \Psi_\mu^{(j)}(x_j)$. Then one simply has

$$\mathrm{Var}[X_j | \boldsymbol{y}] = \frac{1}{b_{\boldsymbol{0}}} \left\langle (x_j - \mathbb{E}[X_j | \boldsymbol{y}])^2, \mathcal{L}_j \right\rangle_{L_{\pi_j}^2} = \frac{1}{b_{\boldsymbol{0}}} \sum_{\mu=0}^2 e_\mu^{(j)} b_\mu^{(j)}. \tag{49}$$

Finally, let $\{e_{0,0}^{(j,k)}, e_{0,1}^{(j,k)}, e_{1,0}^{(j,k)}, e_{1,1}^{(j,k)}\}$ be the coefficients of the bivariate PCE with $(x_j - \mathbb{E}[X_j | \boldsymbol{y}])(x_k - \mathbb{E}[X_k | \boldsymbol{y}]) = \sum_{\mu,\nu=0}^1 e_{\mu,\nu}^{(j,k)} \Psi_\mu^{(j)}(x_j) \Psi_\nu^{(k)}(x_k)$. For an off-diagonal entry of Eq. (5) one then finds

$$\mathrm{Cov}[X_j, X_k | \boldsymbol{y}] = \frac{1}{b_{\boldsymbol{0}}} \sum_{\mu,\nu=0}^1 e_{\mu,\nu}^{(j,k)} b_{\mu,\nu}^{(j,k)}. \tag{50}$$

Notation-wise, Eqs. (48) to (50) may seem to be somewhat cumbersome. Nevertheless, they establish simple recipes of how to obtain the first posterior moments by a postprocessing of the low-degree SLE terms in closed-form. Higher-order moments could be obtained similarly. Some examples of how the corresponding QoIs can be represented in terms of orthogonal polynomials can be found in the Appendix B.

4.6. Discussion of the advantages

In spectral Bayesian inference the posterior is genuinely characterized through the SLE and its coefficients. The essential advantage of this approach is that all quantities of inferential relevance can be computed semi-analytically. Simple formulas for the joint posterior density and the model evidence emerge in Eqs. (39) and (40). They allow to establish Eq. (41) as the posterior density surrogate. General QoI-expectations under the posterior are then calculated via Parseval's formula in Eq. (43). The posterior marginals are obtained based on sub-expansions of the full SLE in Eq. (45) and the first posterior moments have closed-form expressions in Eqs. (48) to (50). These striking characteristics clearly distinguish spectral inference from integration and sampling approaches where the posterior is summarized by expected values or random draws only. As for the latter, one has to rely on kernel estimates of the posterior density and on empirical sample approximations of the QoI-expectations. Also, the model evidence is not computed explicitly.

The practical computation of the SLE in Eq. (36) can be accomplished analogously to finding the PCE approximation of the forward model in Eq. (20), e.g. by solving a linear least squares problem as in Eq. (30). This allows one to draw on the vast number of tools that were developed for carrying out this well-known

type of regression analysis. An attractive feature of this procedure is that the prediction error of the obtained SLE acts as a natural convergence indicator. We recall that the LOO error in Eq. (33) can be efficiently evaluated as per Eq. (34), i.e. without the need for additional forward model runs or regression analyses. The existence of an intrinsic convergence criterion is an advantage over traditional MCMC techniques. Another advantage of the formulation its amenability to parallel computations. While the workload posed by MCMC is inherently serial, running the forward model for each input in the experimental design is embarrassingly parallel. Parallelization is also possible on the level of the linear algebra operations that are necessary in order to solve the normal equations.

4.7. Discussion of the shortcomings

The approximate nature of SLE computations is twofold, i.e. only a finite number of terms are kept in the expansion and the coefficients are inexact. Unfortunately, a number of inconveniences may arise from these inevitable approximations. The SLE and the correspondingly computed posterior density could spuriously take on negative values. Also the estimated model evidence in Eq. (40) could take on negative values $Z < 0$. Still note that the approximate posterior in Eq. (39) always integrates to one. For reasonably adequate SLEs, we expect that negative values only occur in the distributional tails. Even so, the presence of negative values hampers the interpretation of the obtained posterior surrogate as a proper probability density, e.g. it leads to finite negative probabilities that are somehow irritating. From a more practical rather than a technical or philosophical perspective, densities are ultimately instrumental to the evaluation of more concrete quantities such as the expectations in Eq. (43). The severity of negative densities has thus to be judged with respect to the distortions of these relevant values. As long as their accurate approximation is guaranteed, the possibility of negative density values is an unavoidable artifact that can be regarded as a minor blemish. And the obtained surrogate density still proves to be expedient to effectively characterize the posterior distribution. In this light, it is more unpleasant that the a posteriori estimates of the model parameters may violate the restrictions that were imposed a priori. In Eq. (48) it could indeed happen that $\mathbb{E}[X_j|\boldsymbol{y}] \notin \mathcal{D}_{x_j}$. Estimations of the second order moments in Eq. (49) could result in unnatural values $\mathrm{Var}[X_j|\boldsymbol{y}] < 0$, too. Although these problems cannot be remedied unless one solves an appropriately constrained version of Eq. (30), they unlikely occur if the SLE is sufficiently accurate. Anticipating outcomes from later numerical demonstrations, we remark that the occurrence of negative density values is observed, while unphysical or unnatural estimates of the first posterior moments are not found.

The SLE decomposition into a globally smooth basis of tensorized polynomials suffers from some other intrinsic problems. Generally, there is the curse of dimensionality, i.e. the increase of the number of regressors in Eq. (25). Furthermore, the SLE convergence rate in Eq. (27) depends on the regularity of the underlying likelihood function. For discontinuous forward or error models the SLE approximation with smooth polynomials converges only slowly. Likelihood functions often show a peaked structure around the posterior modes and a vanishing behavior elsewhere. Hence, any adequate superposition of polynomials has to capture those two different behavioral patterns through some kind of "constructive" and "destructive" interaction between its terms, respectively. Due to their global nature, the employed polynomial basis may not admit sparse likelihood representations. In turn, a high number of terms might be necessary in order to accurately represent even simple likelihood functions. Especially in the case of high-dimensional and unbounded parameter spaces this may cause severe practical problems. Of course, in Eqs. (36) and (37) one could expand the likelihood in a different basis. Yet, note that the QoIs in Eq. (43) would also have to be expanded in that basis.

The role of the prior for spectral inference is manifold. Initially the posterior expectations in Eq. (6) have been rewritten as the weighted prior expectations in Eq. (7). This formulation is accompanied by difficulties in computing posteriors that strongly deviate from the prior. The same situation arises for crude MC integration and eventually motivates importance or MCMC sampling that allow to focus on localized regions of high posterior mass. Those difficulties become manifest if the prior acts as the sampling distribution for the experimental design. In this case, the SLE is only accurate over the regions of the parameter space that accumulate considerable shares of the total prior probability mass. Approximation errors of the SLE in regions that are less supported by the prior then induce errors in the computed posterior surrogate. Note that this difficulty is also encountered in MCMC posterior exploration with prior-based PCE

metamodels. Also, the error estimate in Eq. (34) then only measures the SLE accuracy with respect to the prior which may be misleading for the assessment of the posterior accuracy. It is not clear how the errors of the likelihood expansion relate to the induced errors of the posterior surrogate and the posterior moments. Moreover, since the prior acts as the reference density of the posterior expansion in Eq. (41), the spectral SLE representation of significantly differing posteriors requires higher order corrections. Otherwise put, SLEs are expected to perform better for posteriors that only slightly update or perturb the prior.

4.8. Change of the reference density

As just discussed, a major drawback of SLEs is their dependency on the prior π as the reference density function. The errors are minimized and measured with respect to the prior and the posterior is represented as correction of the standard reference. In case that high-order corrections are required, SLEs also suffer from the curse of dimensionality. While fully maintaining the advantages of the spectral problem formulation, these shortcomings can be remedied through the introduction of an *auxiliary density* g over the prior support $\mathcal{D}_{\boldsymbol{x}}$ that would optimally mimic the posterior in some sense. This *reference density change* allows for the construction of auxiliary expansions that are more accurate with respect to the posterior and for a more convenient series expansions of the joint posterior density. It is analogous to the adjustment of the integration weight in importance sampling, where the average over a distribution is replaced by a weighted average over another ancillary distribution. An iterative use of this reference change naturally allows for adaptive SLE approaches.

Given that $g(\boldsymbol{x}) \neq 0$ for all $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}}$, one may define the auxiliary quantity $\mathcal{G} = \mathcal{L}\pi/g$. Under the additional assumption that $\mathcal{G} \in L_g^2(\mathcal{D}_{\boldsymbol{x}})$, one can expand this quantity in terms of polynomials $\{\Psi_{\boldsymbol{\alpha}}^g\}_{\boldsymbol{\alpha} \in \mathbb{N}^M}$ that are orthogonal with respect to the auxiliary reference. Analogous to the expansion of $\mathcal{L} \in L_\pi^2(\mathcal{D}_{\boldsymbol{x}})$ in Eqs. (36) and (37), this is

$$\mathcal{G} = \frac{\mathcal{L}\pi}{g} = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m} b_{\boldsymbol{\alpha}}^g \Psi_{\boldsymbol{\alpha}}^g, \quad \text{with} \tag{51}$$

$$b_{\boldsymbol{\alpha}}^g = \langle \mathcal{G}, \Psi_{\boldsymbol{\alpha}}^g \rangle_{L_g^2} = \int_{\mathcal{D}_{\boldsymbol{x}}} \mathcal{G}(\boldsymbol{x})\Psi_{\boldsymbol{\alpha}}^g(\boldsymbol{x})\, g(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \int_{\mathcal{D}_{\boldsymbol{x}}} \mathcal{L}(\boldsymbol{x})\Psi_{\boldsymbol{\alpha}}^g(\boldsymbol{x})\, \pi(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}. \tag{52}$$

The coefficients of this *auxiliary SLE* (aSLE) are denoted as $\{b_{\boldsymbol{\alpha}}^g\}_{\boldsymbol{\alpha} \in \mathbb{N}^M}$. They equal the projections $b_{\boldsymbol{\alpha}}^g = \langle \mathcal{L}, \Psi_{\boldsymbol{\alpha}}^g \rangle_{L_\pi^2}$ of the likelihood onto the polynomials $\Psi_{\boldsymbol{\alpha}}^g$. Note that if the new reference $g = \pi$ equals the prior, then $\mathcal{G} = \mathcal{L}$ is simply the likelihood and the formulation remains unchanged. If the density $g = \pi(\cdot|\boldsymbol{y})$ equals the posterior, then the quantity $\mathcal{G} = Z$ equals the model evidence. In this case the aSLE $\mathcal{G} = b_{\boldsymbol{0}}^g \Psi_{\boldsymbol{0}}^g = b_{\boldsymbol{0}}^g$ is a constant with a single nonzero term. If $g \approx \pi(\cdot|\boldsymbol{y})$ only applies in an approximate sense, then one may still speculate that the aSLE is sparser than the corresponding SLE.

As in importance sampling, one can then rewrite the expectation values under π in Eqs. (2) and (7) as expectations under g. Similar to Eq. (40), the model evidence then emerges again as the zeroth expansion term

$$Z = \int_{\mathcal{D}_{\boldsymbol{x}}} \mathcal{G}(\boldsymbol{x})\, g(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = b_{\boldsymbol{0}}^g. \tag{53}$$

Let $h = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} c_{\boldsymbol{\alpha}}^g \Psi_{\boldsymbol{\alpha}}^g$ with $c_{\boldsymbol{\alpha}}^g = \langle h, \Psi_{\boldsymbol{\alpha}}^g \rangle_{L_g^2}$ be the auxiliary expansion of a QoI. Similar to Eq. (43), for general QoI posterior expectations one may then write

$$\mathbb{E}[h(\boldsymbol{X})|\boldsymbol{y}] = \frac{1}{Z} \int_{\mathcal{D}_{\boldsymbol{x}}} h(\boldsymbol{x})\mathcal{G}(\boldsymbol{x})\, g(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \frac{1}{b_{\boldsymbol{0}}^g} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} c_{\boldsymbol{\alpha}}^g b_{\boldsymbol{\alpha}}^g. \tag{54}$$

In accordance with the aSLE in Eqs. (51) and (52) the joint density of the posterior distribution is obtained as the asymptotic series

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \frac{\mathcal{G}(\boldsymbol{x})g(\boldsymbol{x})}{Z} = \frac{1}{b_{\boldsymbol{0}}^g} \left(\sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} b_{\boldsymbol{\alpha}}^g \Psi_{\boldsymbol{\alpha}}^g(\boldsymbol{x}) \right) g(\boldsymbol{x}) = g(\boldsymbol{x}) \left(1 + \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M \setminus \{\boldsymbol{0}\}} \frac{b_{\boldsymbol{\alpha}}^g}{b_{\boldsymbol{0}}^g} \Psi_{\boldsymbol{\alpha}}^g(\boldsymbol{x}) \right). \tag{55}$$

As opposed to Eq. (41) where the posterior density is represented around the prior π, in Eq. (55) the posterior is expanded about the new reference g. If the latter resembles the posterior adequately well, the formulation only calls for small corrections.

5. Numerical examples

Next the potential and the difficulties of the theory presented in the preceding section are investigated. The goal is to give a proof of concept for the basic feasibility of spectral Bayesian inference. It is verified that the theory can be successfully applied in practice and further insight into its functioning is obtained. Moreover, it is learned about its current shortcomings. Four instructive calibration problems from classical statistics and inverse modeling are solved for these purposes. The analysis is confined to problems with low-dimensional parameter spaces. First, the mean value of a normal distribution is inferred with random data under a conjugate normal prior. Second, the mean and standard deviation of a normal distribution are fitted for a joint prior with independent and uniform marginals. Third, an inverse heat conduction problem in two spatial dimensions with two unknowns is solved. Finally, a similar thermal problem with six unknowns is considered. Synthetically created pseudo-data are used in all these example applications.

As it turns out, one can gain valuable insights into the characteristics of likelihood expansions and posterior emulators by way of comparison. Therefore, the analyses for the first three examples proceed analogously. For rich experimental designs, the convergence behavior of high-degree SLEs is studied by reference to the LOO error. More importantly, the capability of lower-degree SLEs to accurately capture the posterior QoI-expectations is explored for scarcer experimental designs. Eventually, aSLE-based posterior surrogates are investigated in order to mitigate the curse of dimensionality. All results are compared to reference solutions. Where possible, the exact solutions from a conjugate Bayesian analysis are used to this effect. Otherwise, corresponding approximations are computed via classical MCMC sampling.

The uncertainty quantification platform UQLab [114, 115] is used throughout the numerical demonstrations. It provides a flexible environment for the uncertainty analysis of engineering systems, e.g. for uncertainty propagation. In this context it ships with a range of regression tools that allow one to easily compute PCEs. These tools can be directly applied to the likelihood function itself in order to compute SLEs. OLS is employed as the standard solving routine in the following examples.

5.1. 1D normal fitting

First of all, we consider the problem of fitting a Gaussian distribution $\mathcal{N}(y_i|\mu, \sigma^2)$ to random realizations y_i with $i = 1, \ldots, N$. The goal is to estimate the unknown mean μ whereas the standard deviation σ is assumed to be already known. Given a Gaussian prior, this one-dimensional normal model with known variance exhibits a Gaussian posterior density. Moreover, a closed-form expression for the model evidence can be derived. Since this offers the possibility of comparing the SLE results with analytical solutions, this simple statistical model is used as a first SLE testbed. Let the data $\boldsymbol{y} = (y_1, \ldots, y_N)^\top$ be comprised of N independent samples from the normal distribution. For the observational model one may then write

$$\boldsymbol{Y}|\mu \sim \prod_{i=1}^{N} \mathcal{N}(y_i|\mu, \sigma^2), \quad \text{with known } \sigma^2. \tag{56}$$

Consequently, the likelihood function can be simply written as $\mathcal{L}(\mu) = \prod_{i=1}^{N} \mathcal{N}(y_i|\mu, \sigma^2)$. A Bayesian prior distribution $\pi(\mu)$ captures the epistemic uncertainty of the true value of μ before the data analysis. For the posterior distribution, that aggregates the information about the unknown after the data have been analyzed, one then has $\pi(\mu|\boldsymbol{y}) = Z^{-1}\mathcal{L}(\mu)\pi(\mu)$.

The conjugate prior for the data model in Eq. (56) is a Gaussian $\pi(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2)$. Its mean μ_0 and standard deviation σ_0 have to be conveniently specified by the experimenter and data analyst. This prior choice ensures that the posterior is a Gaussian $\pi(\mu|\boldsymbol{y}) = \mathcal{N}(\mu|\mu_N, \sigma_N^2)$ whose parameters $\mu_N = \mathbb{E}[\mu|\boldsymbol{y}]$ and $\sigma_N^2 = \text{Var}[\mu|\boldsymbol{y}]$ are easily found as

$$\mu_N = \left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}\right)^{-1}\left(\frac{\mu_0}{\sigma_0^2} + \frac{N\overline{y}}{\sigma^2}\right), \quad \text{and } \sigma_N^2 = \left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}\right)^{-1}. \tag{57}$$

Here, $\overline{y} = N^{-1} \sum_{i=1}^{N} y_i$ is the empirical sample mean of the data. Likewise, an explicit expression for the model evidence $Z = \int_{\mathbb{R}} (\prod_{i=1}^{N} \mathcal{N}(y_i|\mu, \sigma^2)) \mathcal{N}(\mu|\mu_0, \sigma_0^2) \, d\mu$ can be derived. Let $\overline{y^2} = N^{-1} \sum_{i=1}^{N} y_i^2$ denote the sample mean of the squared observations. A straightforward calculation based on simple algebra and a Gaussian integral then yields

$$Z = \sigma_0^{-1} \left(\sigma \sqrt{2\pi}\right)^{-N} \left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}\right)^{-1/2} \exp\left(-\frac{1}{2}\left(\frac{\mu_0^2}{\sigma_0^2} + \frac{N\overline{y^2}}{\sigma^2} - \left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}\right)^{-1}\left(\frac{\mu_0}{\sigma_0^2} + \frac{N\overline{y}}{\sigma^2}\right)^2\right)\right). \quad (58)$$

For the following computer experiment, the parameters of the data distribution in Eq. (56) are specified as $\mu = 10$ and $\sigma = 5$, respectively. In the course of the procedure only the mean is treated as an unknown, whereas the standard deviation is assumed to be known. We consider a situation where $N = 10$ samples are randomly drawn from the data distribution. For the numerical experiment, the pseudo-random numbers $\boldsymbol{y} = (8.78, 4.05, 12.58, 3.60, 11.05, 8.70, 20.80, 1.23, 19.36, 12.07)^{\top}$ are used as synthetic data. The prior distribution is set to be a Gaussian $\pi(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2)$ with $\mu_0 = 11.5$ and $\sigma_0 = 1.5$.

5.1.1. Posterior density

In order to better understand the principles of spectral Bayesian inference we now proceed as follows. Spectral expansions $\hat{\mathcal{L}}_p$ of the likelihood function $\mathcal{L}$ defined above are computed and compared for experimental designs of varying size K and polynomial terms of varying degree p. Hermite polynomials are used in combination with an appropriate linear transformation to standardized variables $\xi_\mu \in \mathbb{R}$ with a Gaussian weight function $\mathcal{N}(\xi_\mu|0,1)$. Accordingly, the unknown can be represented as $\mu = \mu_0 + \sigma_0 \xi_\mu$. The experimental designs are one-dimensional Sobol sequences that are appropriately transformed.

First the convergence behavior and the accuracy of the likelihood approximation are analyzed. For a rich experimental design with $K = 5 \times 10^4$, SLEs are computed for an increasing order up to $p = 20$. The normalized empirical error ϵ_{Emp} and the normalized LOO error ϵ_{LOO} are monitored over these computations. While the former can be directly computed according to its definition, the computation of the latter relies on the reformulation in Eq. (34). This serves the purpose of assessing the prediction accuracy of the computed SLE as a function of the degree p. The results are plotted in Fig. 1. It can be seen how the error estimates approach zero, i.e. the SLE converges to the likelihood function. For $p = 20$ the empirical error amounts to $\epsilon_{\text{Emp}} = 1.05 \times 10^{-12}$ and the LOO amounts to $\epsilon_{\text{LOO}} = 1.82 \times 10^{-10}$. These small error magnitudes show that the likelihood function $\mathcal{L}$ can be indeed spectrally expanded in a Hermite basis.
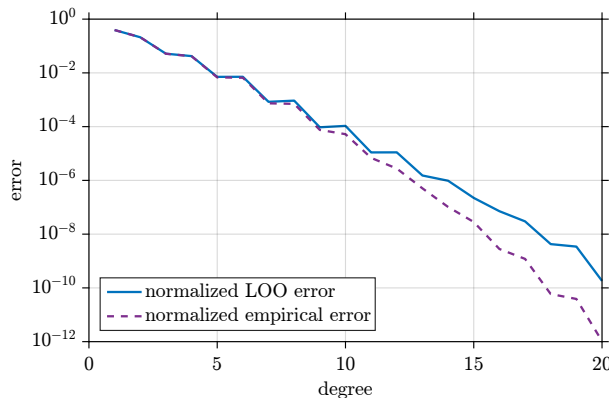

Figure 1: 1D normal fitting: Convergence of the SLE.

The functional likelihood approximation $\hat{\mathcal{L}}_p$ provided by the most accurate SLE with $p = 20$ is visualized in Fig. 2. Moreover, the plot shows a low-order SLE with $p = 5$ and $K = 1 \times 10^2$ for which the error estimates $\epsilon_{\text{Emp}} = 2.61 \times 10^{-4}$ and $\epsilon_{\text{LOO}} = 8.41 \times 10^{-4}$ are obtained. For the sake of comparison the exact likelihood function $\mathcal{L}$ is shown as well. It can be seen that the SLEs are able to accurately represent the likelihood around its peak, i.e. roughly speaking in the interval $\mu \in [8, 15]$ for $p = 5$ and in $\mu \in [5, 18]$ for $p = 20$.

17

Note that these regions accumulate the largest proportions of the total prior probability mass. Outside of these ranges, however, the SLEs $\hat{\mathcal{L}}_p$ start strongly deviating from $\mathcal{L}$ and taking on negative values. These phenomena can be attributed to an imperfect polynomial cancellation of the finite series approximation of the likelihood in the regions of the parameter space that are only sparsely covered by the experimental design. Indeed, for unbounded parameter spaces it is clearly hopeless to achieve a global net cancellation of a finite polynomial expansion that is necessary in order to emulate the vanishing behavior of the likelihood far from its peaks. The extent to which this impacts on the approximation of the posterior density and its first moments is investigated next.

Expanding the likelihood function is only a means to the end of surrogating the posterior density. Approximations of the posterior density $\pi(\mu|\boldsymbol{y}) \approx b_0^{-1}\hat{\mathcal{L}}_p(\mu)\pi(\mu)$ are computed from the SLEs with $p = 5$ and $p = 20$ through Eqs. (39) and (40). The results are plotted in Fig. 3. In addition to the SLE approximations, the prior density $\pi(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2)$ and the exact solution $\pi(\mu|\boldsymbol{y}) = \mathcal{N}(\mu|\mu_N, \sigma_N^2)$ from a conjugate analysis based on Eq. (57) are shown. The posterior surrogate for $p = 5$ shows minor deviations from the the analytical result, while the approximation for $p = 20$ perfectly matches the true density. It is noted that the discrepancies between $\hat{\mathcal{L}}_p$ and $\mathcal{L}$ shown in Fig. 2 are attenuated. The underlying reason is that for large enough $|\mu| \to \infty$ the exponential decay of the Gaussian prior $\pi(\mu) \propto \exp(-(\mu - \mu_0)^2)$ dominates the polynomial increase of $\hat{\mathcal{L}}_p(\mu) = \sum_{\alpha=0}^{p} b_\alpha \Psi_\alpha(\mu)$ in the sense that $\hat{\mathcal{L}}_p(\mu)\pi(\mu) \to 0$. This absorbs the effects of the SLE approximation that is increasingly inadequate for large values of $|\mu|$. In this sense, the prior reference density guards the posterior surrogate against the inadequacies of the SLE. Therefore, the posterior emulation may very well be more accurate than the SLE approximation of the likelihood.
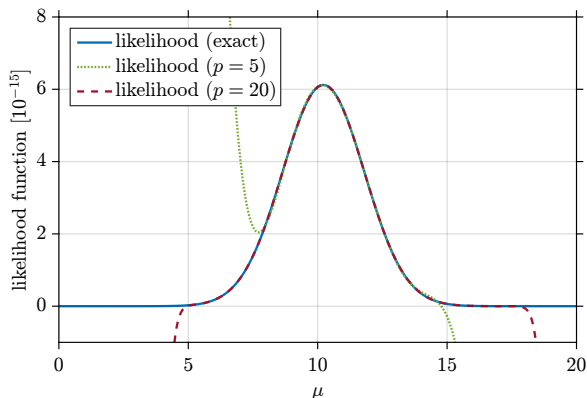
Figure 2: 1D normal fitting: Likelihood function.
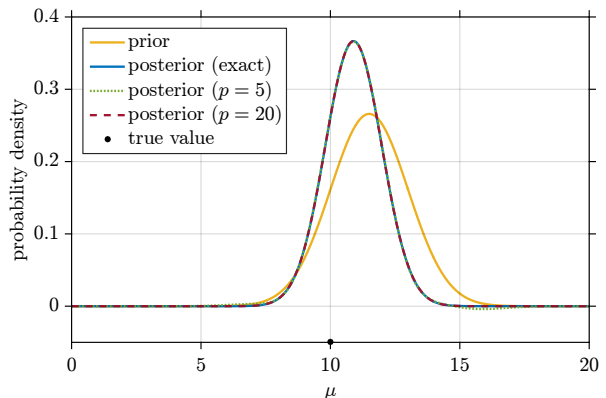
Figure 3: 1D normal fitting: Posterior density.

5.1.2. Quantities of interest

Commonly one employs posterior means as parameter estimates and posterior standard deviations as measures of the estimation uncertainty. In order to investigate how well one can approximate the model evidence together with these meaningful quantities in spectral Bayesian inference, SLEs are computed for experimental designs of varying size K and for a selection of expansion orders p. The corresponding SLE-based approximations of the model evidence Z, the posterior mean μ_N and the standard deviation σ_N are then computed from Eqs. (40), (48) and (49). Note that the effects of the transformation to standard variables have to be appropriately taken care of at this place. This happens via Eq. (19). The SLE approximations can then be compared to the analytical solutions that are obtained from the conjugate analysis in Eqs. (57) and (58). In Table 2 the results of this procedure are summarized. Note that all the SLE estimates attain admissible values, e.g. the model evidence is non-negative. Furthermore, it is noticed that Z, μ_N and σ_N can be recovered with high accuracy even for very scarce experimental designs and low-order SLEs, say for $K = 1 \times 10^3$ and $p = 10$. It is concluded that, in some sense, the accurate estimation of the model evidence and the first posterior moments require significantly less computational effort than the accurate estimation of the posterior density.

18

Table 2: 1D normal fitting: Statistical quantities.

	K	p	ϵ_{LOO}	$Z\ [10^{-15}]$	μ_N	σ_N
SLE	1×10^2	5	8.41×10^{-4}	3.71	10.85	0.92
	5×10^2	8	2.49×10^{-4}	3.75	10.91	1.14
	1×10^3	10	2.58×10^{-5}	3.74	10.90	1.07
	5×10^3	12	8.21×10^{-6}	3.74	10.89	1.09
	1×10^4	15	3.84×10^{-7}	3.74	10.89	1.09
	5×10^4	20	1.82×10^{-10}	3.74	10.89	1.09
exact results				3.74	10.89	1.09

5.2. 2D normal fitting

Next, we consider the problem of fitting both the unknown mean μ and the standard deviation σ of a Gaussian distribution $\mathcal{N}(y_i|\mu,\sigma^2)$. A number of independent samples y_i with $i = 1,\ldots,N$ from the normal distribution constitute the available data $\boldsymbol{y} = (y_1,\ldots,y_N)^\top$. The data model for this situation is written as

$$\boldsymbol{Y}|\mu,\sigma \sim \prod_{i=1}^{N} \mathcal{N}(y_i|\mu,\sigma^2). \tag{59}$$

For the likelihood function one then has $\mathcal{L}(\mu,\sigma) = \prod_{i=1}^{N} \mathcal{N}(y_i|\mu,\sigma^2)$. Given a Bayesian prior $\pi(\mu,\sigma)$, the posterior distribution is $\pi(\mu,\sigma|\boldsymbol{y}) = Z^{-1}\mathcal{L}(\mu,\sigma)\pi(\mu,\sigma)$. This distribution aggregates the information about the two unknowns after the data have been analyzed.

The true values of the mean and standard deviation are set as $\mu = 30$ and $\sigma = 5$, respectively. These values are treated as unknowns in the further course of the computer experiment. We consider a situation where $N = 10$ samples are randomly drawn from the distribution in Eq. (59). The pseudo-random numbers $\boldsymbol{y} = (31.23, 27.50, 24.91, 25.99, 32.88, 36.41, 27.81, 25.19, 37.96, 34.84)^\top$ are used as synthetic data. We consider an independent prior $\pi(\mu,\sigma) = \pi(\mu)\pi(\sigma)$ with uniform marginals $\pi(\mu) = \mathcal{U}(\mu|\underline{\mu},\overline{\mu})$ and $\pi(\sigma) = \mathcal{U}(\sigma|\underline{\sigma},\overline{\sigma})$ over bounded supports $\mathcal{D}_\mu = [\underline{\mu},\overline{\mu}] = [20,40]$ and $\mathcal{D}_\sigma = [\underline{\sigma},\overline{\sigma}] = [2,10]$. As opposed to the conjugate example above, this two-dimensional model does not permit a closed-form expression of the posterior density and the model evidence.

5.2.1. Posterior density

Now we proceed analogously to the investigation of the normal model with known variance. Expansions $\hat{\mathcal{L}}_p$ of the likelihood $\mathcal{L}$ are computed and contrasted for different experimental designs of size K and different polynomial orders p. An appropriate linear transformation to uniform standardized variables is applied such that the unknowns are represented as $\mu = (\overline{\mu} - \underline{\mu})/2 \cdot \xi_\mu + (\underline{\mu} + \overline{\mu})/2$ and $\sigma = (\overline{\sigma} - \underline{\sigma})/2 \cdot \xi_\sigma + (\underline{\sigma} + \overline{\sigma})/2$, respectively. Here, $\xi_\mu, \xi_\sigma \in [-1,1]$ are the corresponding standardized variables with a uniform weight function. Accordingly, tensorized Legendre polynomials form the trial basis. Two-dimensional Sobol sequences are utilized as uniformly space-filling experimental designs.

As before, the speed of convergence and the prediction accuracy of the SLE are analyzed first. The normalized empirical error ϵ_{Emp} and the normalized LOO error ϵ_{LOO} are therefore monitored throughout a series of runs that are conducted for an experimental design of the fixed size $K = 1 \times 10^5$ and for an increasing expansion order up to $p = 50$. In Fig. 4 a corresponding plot is shown, where the convergence of the SLE $\hat{\mathcal{L}}_p$ to the likelihood function $\mathcal{L}$ is diagnosed. The reason that ϵ_{Emp} and ϵ_{LOO} do not significantly differ is that the large size of the experimental design prevents overfitting. For $p = 50$ the normalized empirical error and the normalized LOO error are found as $\epsilon_{\text{Emp}} = 5.56 \times 10^{-11}$ and $\epsilon_{\text{LOO}} = 6.05 \times 10^{-11}$, respectively. This shows that the likelihood function $\mathcal{L}$ can be indeed expanded in the Legendre basis. For the uniform prior distribution that is used here, the normalized SLE errors effectively measure the errors of the posterior density.
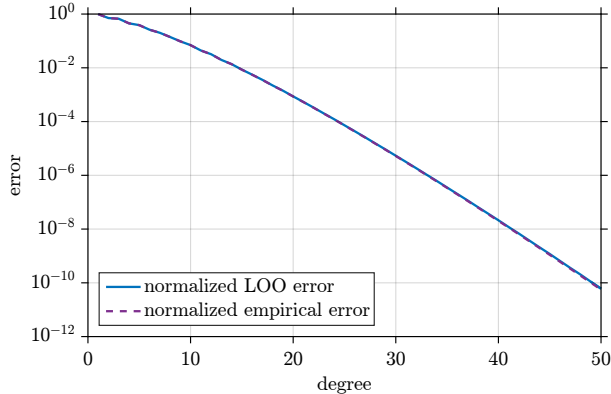
Figure 4: 2D normal fitting: Convergence of the SLE.

Now the joint posterior density $\pi(\mu, \sigma | \boldsymbol{y})$ is computed and plotted in Fig. 5. For comparison purposes the posterior is sampled by means of MCMC simulation first. A simple random walk Metropolis (RWM) sampler with a Gaussian instrumental distribution is utilized. With this algorithm an unusually large number of 10^7 MCMC samples is drawn from the posterior. This serves the purpose of providing very accurate results that act as references for the SLE-based estimates. In Fig. 5a a normalized histogram of the obtained RWM sample is shown. Next, the joint posterior density $\pi(\mu, \sigma | \boldsymbol{y}) \approx b_{\boldsymbol{0}}^{-1} \hat{\mathcal{L}}_p(\mu, \sigma) \pi(\mu, \sigma)$ is computed via Eqs. (39) and (40). The SLE $\hat{\mathcal{L}}_p(\mu, \sigma)$ with $p = 50$ that features the lowest LOO error is used. In Fig. 5b the posterior surrogate that arises from the SLE is plotted. For a later comparison with the heat conduction example, in Fig. 5c the SLE posterior surrogate from Fig. 5b is plotted again from a different angle. By visual inspection the obvious similarity between the density $\pi(\mu, \sigma | \boldsymbol{y})$ sampled by MCMC and emulated by the SLE is noticed.

Now the posterior marginals $\pi(\mu | \boldsymbol{y})$ and $\pi(\sigma | \boldsymbol{y})$ are computed from the joint posterior. On the one hand, samples from the posterior marginals are obtained by restricting the analysis to the corresponding components of the joint MCMC sample. On the other hand, functional approximations of the posterior marginals are extracted based on sub-expansions $\hat{\mathcal{L}}_{\mu,p}(\mu)$ and $\hat{\mathcal{L}}_{\sigma,p}(\sigma)$ of a joint SLE $\hat{\mathcal{L}}_p(\mu, \sigma)$ as in Eqs. (44) and (45). For the SLEs with $p = 9$ and $p = 50$ the results are visualized in Fig. 6. Histogram-based MCMC sample representations and functional SLE approximations of the marginal densities are shown, too. As it can be seen, the marginal posteriors as obtained by MCMC and the SLE with $p = 50$ exactly match each other. For $p = 9$ the posteriors marginals display some wavelike fluctuations in their tails.

5.2.2. Quantities of interest

Since the posterior density itself is of little inferential use, the model evidence and the first posterior moments are computed for a selection of SLEs with varying size of the experimental design K and degree p. According to Eqs. (40) and (48) to (50), the SLE estimates of these quantities are obtained from the expansions coefficients. In Table 3 a summary of the results is given. Compliant with Eq. (40) the SLE estimates of the model evidence Z are obtained as the coefficient of the constant expansion term. According to Eqs. (48) and (49), the SLE estimates of the posterior mean $\mathbb{E}[\mu | \boldsymbol{y}]$ and the standard deviation $\text{Std}[\mu | \boldsymbol{y}] = \text{Var}[\mu | \boldsymbol{y}]^{1/2}$ of the location parameter μ are computed. Likewise, the corresponding estimates for the spread parameter σ follow through a simple postprocessing of the low-order expansion coefficients. The SLE estimates of the linear coefficient of correlation $\rho[\mu, \sigma | \boldsymbol{y}] = \text{Cov}[\mu, \sigma | \boldsymbol{y}] / \text{Std}[\mu | \boldsymbol{y}] / \text{Std}[\sigma | \boldsymbol{y}]$ are computed based on Eq. (50). Additionally, the LOO error ϵ_{LOO} is listed to indicate the SLE prediction accuracy. Note that all those estimates comply with the natural bounds and restrictions of the estimated quantities, e.g. the posterior means comply with the prior bounds.
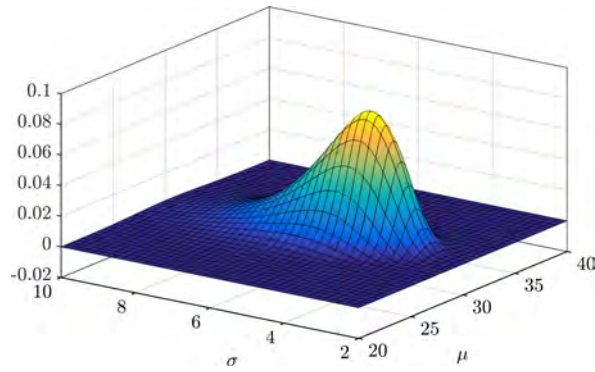
For the sake of comparison, associated results are listed for the simulated MCMC sample, too. These MCMC results are simply obtained as the corresponding sample approximations. The reference estimate of the model evidence is obtained by crude MC simulation instead, i.e. the arithmetic mean of the likelihood
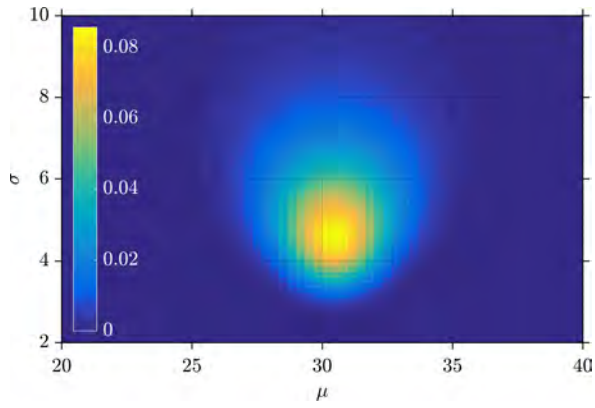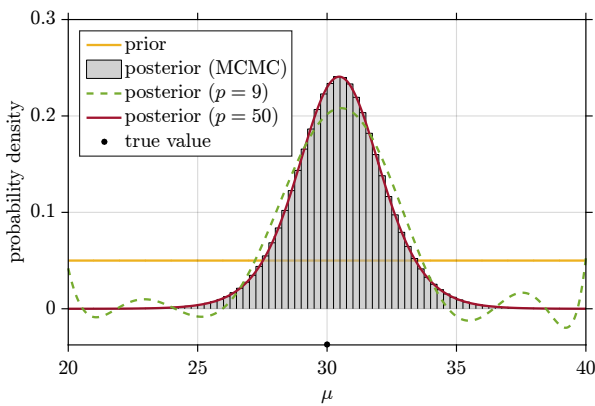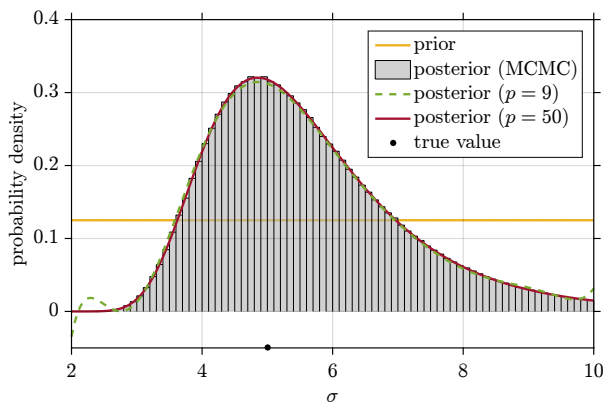
(a) MCMC reference sample.
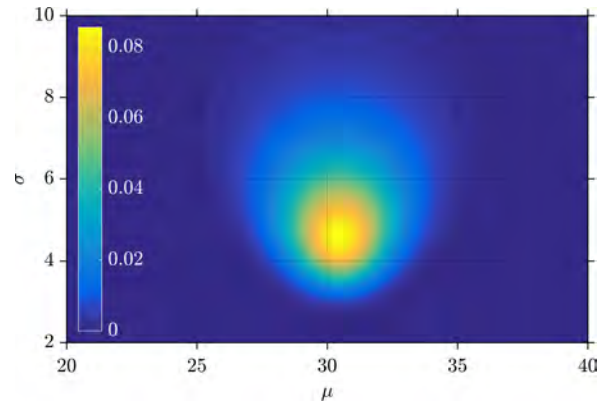
(b) SLE with $p = 50$.

(c) SLE with $p = 50$.

Figure 5: 2D normal fitting: Joint posterior.

(a) Mean value μ.

(b) Standard deviation σ.

Figure 6: 2D normal fitting: Posterior marginals.

is computed for a number of 10^8 independent draws from the prior. It is interesting to note that the SLEs can reproduce the MCMC results for moderate experimental designs and degrees, say for $K = 5 \times 10^3$ and $p = 21$. Even though a large number of input samples and a large polynomial degree is necessary to reproduce the shape of the joint posterior density, significantly smaller experimental designs and polynomial orders suffice to reproduce the first posterior moments.

Table 3: 2D normal fitting: Statistical quantities.

	K	p	ϵ_{LOO}	$Z\,[10^{-14}]$	$\mathbb{E}[\mu|\boldsymbol{y}]$	$\mathbb{E}[\sigma|\boldsymbol{y}]$	$\text{Std}[\mu|\boldsymbol{y}]$	$\text{Std}[\sigma|\boldsymbol{y}]$	$\rho[\mu,\sigma|\boldsymbol{y}]$
	5×10^2	5	4.24×10^{-1}	1.19	30.34	5.57	2.03	1.39	0.18
	1×10^3	9	1.19×10^{-1}	1.20	30.39	5.54	2.01	1.41	0.08
SLE	5×10^3	21	9.64×10^{-4}	1.18	30.48	5.56	1.79	1.38	-0.01
	1×10^4	32	5.86×10^{-6}	1.18	30.47	5.56	1.81	1.38	0.00
	5×10^4	45	1.30×10^{-9}	1.18	30.47	5.56	1.81	1.38	-0.00
	1×10^5	50	6.05×10^{-11}	1.18	30.47	5.56	1.81	1.38	-0.00
(MC)MC				1.18	30.47	5.56	1.81	1.38	-0.00

5.3. 2D inverse heat conduction

Finally, an inverse heat conduction problem (IHCP) is considered. The heat equation is a partial differential equation (PDE) that describes the distribution and evolution of heat in a system where conduction is the dominant mode of heat transfer. We consider a stationary heat equation of the form

$$\nabla \cdot (\kappa \nabla \tilde{T}) = 0. \tag{60}$$

The temperature is denoted as $\tilde{T}$ and the thermal conductivity is denoted as κ. Commonly one is interested in the solution of the boundary value problem that is posed when Eq. (60) is satisfied over a physical domain subject to appropriate boundary conditions. We consider the steady state situation in two spatial dimensions. The Euclidean coordinate vector is denoted as $\boldsymbol{r} = (r_1, r_2)^\top$ in the following.

It is dealt with the identification of thermal conductivities of inclusions in a composite material with close-to-surface measurements of the temperature. The setup of the simplified thermal problem is visualized in Fig. 7. The thermal conductivity of the background matrix is denoted as κ_0, while the conductivities of the material inclusions are termed as κ_1 and κ_2, respectively. It is assumed that the material properties are not subject to a further spatial variability. At the "top" of the domain a Dirichlet boundary condition $\tilde{T}_1$ is imposed, while at the "bottom" the Neumann boundary condition $q_2 = -\kappa_0\, \partial\tilde{T}/\partial r_2$ is imposed. Zero heat flux conditions $\partial\tilde{T}/\partial r_1 = 0$ are imposed at the "left" and "right" hand side.

We consider the IHCP that is posed when the thermal conductivities $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)^\top$ are unknown and their inference is intended. With this in mind, a number of N measurements $\boldsymbol{T} = (T(\boldsymbol{r}_1), \ldots, T(\boldsymbol{r}_N))^\top$ of the temperature field at the measurement locations $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ is available. The forward model $\mathcal{M} \colon \boldsymbol{\kappa} \mapsto \tilde{\boldsymbol{T}}$ establishes the connection between the data and the unknowns. It formalizes the operation of solving Eq. (60) for $\tilde{\boldsymbol{T}}$ as a function of $\boldsymbol{\kappa}$. Measured temperatures $\boldsymbol{T} = \tilde{\boldsymbol{T}} + \boldsymbol{\varepsilon}$ consist of the corresponding model response $\tilde{\boldsymbol{T}} = \mathcal{M}(\boldsymbol{\kappa})$ and a residual term $\boldsymbol{\varepsilon}$. The latter accounts for measurement uncertainty and forward model inadequacy. We consider residuals that are distributed according to a Gaussian $\mathcal{N}(\boldsymbol{\varepsilon}|\boldsymbol{0}, \boldsymbol{\Sigma})$ with covariance matrix $\boldsymbol{\Sigma}$. In compliance with Eq. (9) the likelihood function is given as $\mathcal{L}(\boldsymbol{\kappa}) = \mathcal{N}(\boldsymbol{T}|\mathcal{M}(\boldsymbol{\kappa}), \boldsymbol{\Sigma})$. Provided that a prior distribution $\pi(\boldsymbol{\kappa})$ can be elicited, the posterior is given as $\pi(\boldsymbol{\kappa}|\boldsymbol{T}) = Z^{-1}\mathcal{L}(\boldsymbol{\kappa})\pi(\boldsymbol{\kappa})$.

The thermal conductivity of the background matrix is set to $\kappa_0 = 15\,\text{W/m/K}$, while the thermal conductivities of the inclusions are specified as $\kappa_1 = 32\,\text{W/m/K}$ and $\kappa_1 = 28\,\text{W/m/K}$. The material properties of the inclusions are treated as unknowns subsequently. Moreover, the boundary conditions $\tilde{T}_1 = 200\,\text{K}$ and $q_2 = 2000\,\text{W/m}^2$ are imposed. A finite element (FE) model is used to solve a weak form of the governing PDE. The FE solution for the experimental setup described above is shown in Fig. 8. We consider a uniform prior distribution $\pi(\boldsymbol{\kappa}) = \pi(\kappa_1)\pi(\kappa_2)$ with independent marginals $\pi(\kappa_1) = \mathcal{U}(\kappa_1|\underline{\kappa}_1, \overline{\kappa}_1)$ and

$\pi(\kappa_2) = \mathcal{U}(\kappa_2 | \underline{\kappa}_2, \overline{\kappa}_2)$. The prior bounds are chosen as $\underline{\kappa}_1 = \underline{\kappa}_2 = 20\,\text{W/m/K}$ and $\overline{\kappa}_1 = \overline{\kappa}_2 = 40\,\text{W/m/K}$, respectively. A number of $N = 12$ close-to-surface observations is analyzed. Their measurement locations are indicated by the black dots in Fig. 7. Independent Gaussian measurement noise with $\boldsymbol{\Sigma} = \sigma_T^2 \boldsymbol{1}$ and $\sigma_T = 0.25\,\text{K}$ is considered. Based on this setup, synthetic data are simulated for conducting the computer experiment. This means that the forward model responses $\tilde{\boldsymbol{T}}$ for the true parameter setup are computed and pseudo-random noise is added in order to obtain $\boldsymbol{T}$.
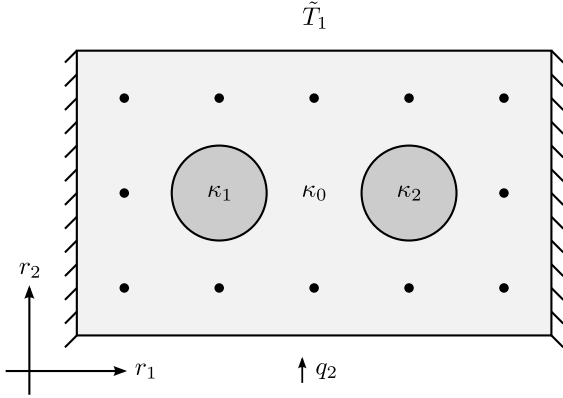
Figure 7: 2D IHCP: Heat conduction setup.
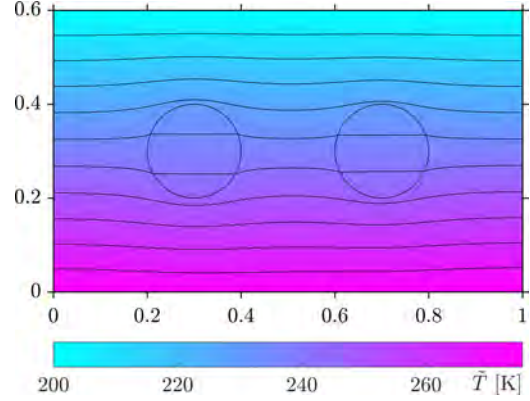
Figure 8: 2D IHCP: Steady state solution.

5.3.1. Posterior density

The analyses proceed analogously to the preceding section. By comparing the present IHCP and the non-conjugate Gaussian example, that have a two-dimensional parameter space and uniform priors in common, one can gain interesting insight into spectral Bayesian inference. First, the convergence behavior of the SLE is investigated. Spectral expansions $\hat{\mathcal{L}}_p$ of the likelihood $\mathcal{L}$ are therefore computed for an experimental design of size $K = 1 \times 10^5$ and candidate bases with polynomials up to degree $p = 50$. All practical issues are handled analogously to the procedure in the non-conjugate Gaussian example. In Fig. 9 the normalized versions of the empirical error ϵ_{Emp} and the LOO error ϵ_{LOO} are shown as a function of p. Comparing these results to Fig. 4 reveals that the convergence rate of the SLE $\hat{\mathcal{L}}_p$ is considerably slower than the corresponding one for the Gaussian example. For the SLE with $p = 50$ the error estimates amount to $\epsilon_{\text{Emp}} = 6.26 \times 10^{-4}$ and $\epsilon_{\text{LOO}} = 7.56 \times 10^{-4}$. These errors are around seven orders of magnitude higher then the errors observed for the Gaussian example. The difference in the SLE convergence rate presumably originates from a difference in the underlying likelihood functions and posterior densities. This is now investigated in more detail.
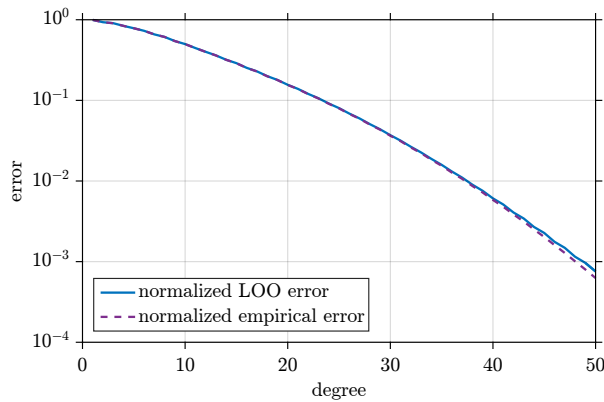

Figure 9: 2D IHCP: Convergence of the SLE.

A RWM approximation with 10^7 samples and the SLE-based emulation with $p = 50$ of the posterior density $\pi(\kappa_1, \kappa_2|\boldsymbol{T}) \approx b_{\boldsymbol{0}}^{-1}\hat{\mathcal{L}}_p(\kappa_1, \kappa_2)\pi(\kappa_1, \kappa_2)$ are depicted in Fig. 10. In order to reduce the numerical cost of MCMC sampling, the FE model $\mathcal{M}$ is replaced by a PCE surrogate $\hat{\mathcal{M}}_p$. For $i = 1, \ldots, N$, separate PCEs $\hat{\mathcal{M}}_{i,p}$ of the temperature $\tilde{T}_i = \mathcal{M}_{i,p}(\kappa_1, \kappa_2)$ at the location $\boldsymbol{r}_i$ are fitted as a function of the unknown conductivities. After an appropriate transformation to standardized variables, tensorized Legendre polynomials up to degree $p = 10$ act as the trial basis. Based on an experimental design of the size $K = 10^3$, the LOO errors of the regressions amount to about $\epsilon_{\mathrm{LOO}} \approx 10^{-10}$. Accordingly, the PCE is considered an adequate replacement of the full FE model. Note that it would be also possible to use $\hat{\mathcal{M}}_p$ as a forward model surrogate during the likelihood training runs.

The posteriors in Fig. 10 can be compared to the posteriors of the Gaussian example in Fig. 5 of the previous section. Relative to the respective prior, the posterior of the thermal problem $\pi(\kappa_1, \kappa_2|\boldsymbol{T})$ contains more information than the posterior of the normal problem $\pi(\mu, \sigma|\boldsymbol{y})$, i.e. the likelihood $\mathcal{L}(\kappa_1, \kappa_2)$ has a slightly more peaked and localized structure than $\mathcal{L}(\mu, \sigma)$. In order to capture these different behaviors nearby and far from the posterior mode, the SLEs $\hat{\mathcal{L}}_p(\kappa_1, \kappa_2)$ and $\hat{\mathcal{L}}_p(\mu, \sigma)$ require a different number of expansions terms. The more localized the posterior modes are with respect to the prior, the more terms are required in order to achieve the cancellation in the tails. Moreover, as opposed to $\pi(\mu, \sigma|\boldsymbol{y})$ the posterior $\pi(\kappa_1, \kappa_2|\boldsymbol{T})$ exhibits a pronounced correlation structure. In turn, this requires non-vanishing interaction terms. As a consequence, the SLE $\hat{\mathcal{L}}_p(\kappa_1, \kappa_2)$ of the IHCP example is less accurate than the SLE $\hat{\mathcal{L}}_p(\mu, \sigma)$ of the Gaussian example. This is also reflected in the fact that the posterior surrogate fluctuates and takes on negative values around the points $[\underline{\kappa}_1, \underline{\kappa}_2]$ and $[\overline{\kappa}_1, \overline{\kappa}_2]$. In order to see this more clearly, the SLE posterior surrogate from Fig. 10b is plotted again from a different angle in Fig. 10c. A small wavelike posterior structure spans the parameter space between these corners. These artifacts stem from an imperfect polynomial cancellation of the finite series approximation. This stands in contrast to the posterior of the Gaussian example in Fig. 5c where these phenomena were not observed.

Via Eq. (45) the posterior marginals $\pi(\kappa_1|\boldsymbol{T})$ and $\pi(\kappa_2|\boldsymbol{T})$ can be extracted from the joint SLEs. The resulting densities are shown in Fig. 11 together with a histogram-based MCMC sample representation. As it can be seen, for $p = 50$ the marginals are captured fairly well, while the moderate-order surrogate for $p = 21$ still exhibits discrepancies at the bounds of the parameter space. The approximation of the posterior marginals by sub-SLEs seems to be more accurate, at least in the sense of the maximum deviation, than the approximation of the joint posterior $\pi(\kappa_1, \kappa_2|\boldsymbol{T})$ by the full SLE in Fig. 10c. This phenomenon can be explained through the absence of all non-constant polynomial terms in the variables that are marginalized out.

5.3.2. Quantities of interest

Now we investigate how well one can extract the statistically interesting quantities. Results from SLEs with varying K and p are compared with the results from MCMC sampling. A summary of the findings is provided in Table 4. The LOO error ϵ_{LOO} of various SLEs is shown together with some basic posterior characteristics obtained by a postprocessing of the SLE coefficients. For $j = 1, 2$ the posterior mean $\mathbb{E}[\kappa_j|\boldsymbol{T}]$ and the standard deviation $\mathrm{Std}[\kappa_j|\boldsymbol{T}] = \mathrm{Var}[\kappa_j|\boldsymbol{T}]^{1/2}$ of the posterior distribution are given in physical units of W/m/K. In addition, the model evidence Z and the linear coefficient of correlation $\rho[\kappa_1, \kappa_2|\boldsymbol{T}] = \mathrm{Cov}[\kappa_1, \kappa_2|\boldsymbol{T}]/\mathrm{Std}[\kappa_1|\boldsymbol{T}]/\mathrm{Std}[\kappa_2|\boldsymbol{T}]$ are specified. In comparison to Table 3, where the results for the non-conjugate normal example are listed, the SLE results for the IHCP match their MCMC counterparts less accurately. Nevertheless, it can be observed that the lowest-degree quantities of inferential interest can be extracted with a comparably small experimental design and relatively low number of regressors, say with $K = 1 \times 10^4$ and $p = 29$. Note that all the estimates attain admissible values.

5.4. 6D inverse heat conduction

In the previous sections it was demonstrated that likelihood functions can be indeed spectrally expanded and that the posterior density with its moments can be computed accordingly. For low-dimensional problems the SLE convergence behavior up to a high degree was studied by monitoring the LOO error. It was shown that the expansion error can be arbitrarily reduced by increasing the order of the expansion and adding
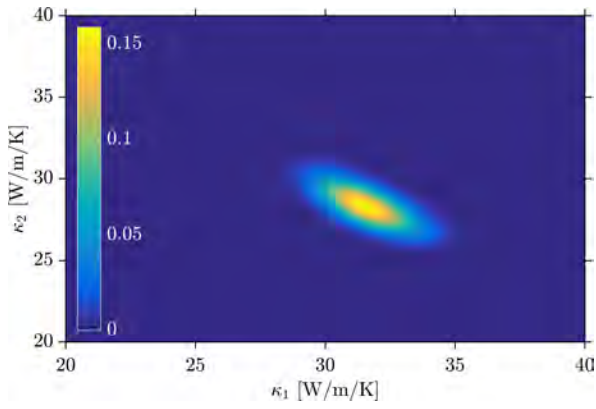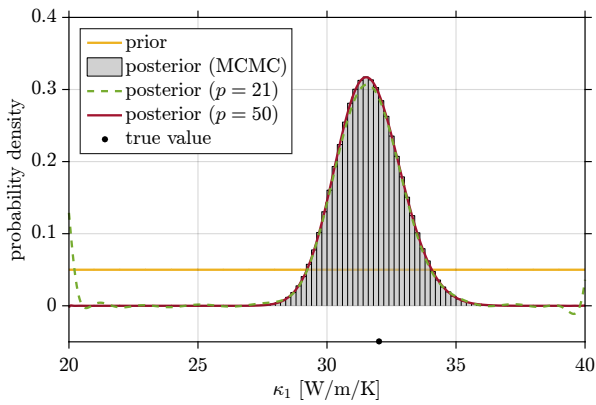
(a) MCMC reference sample.
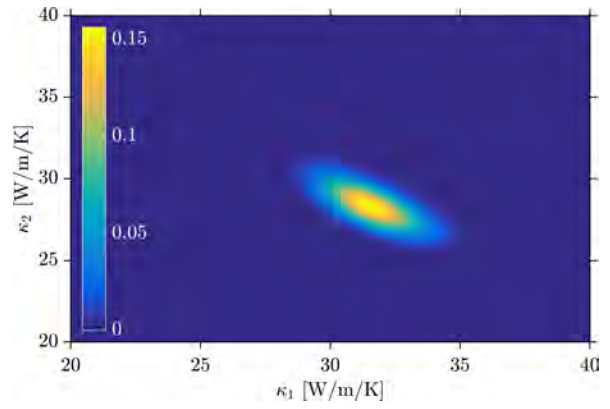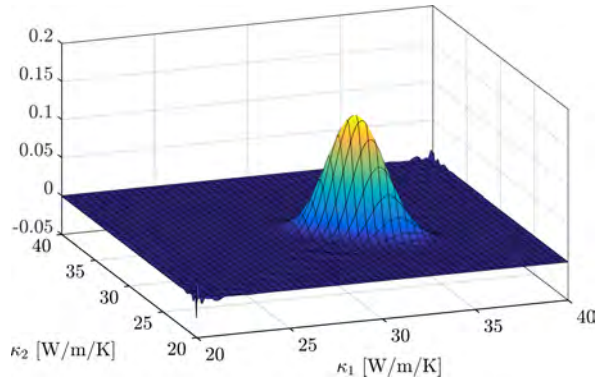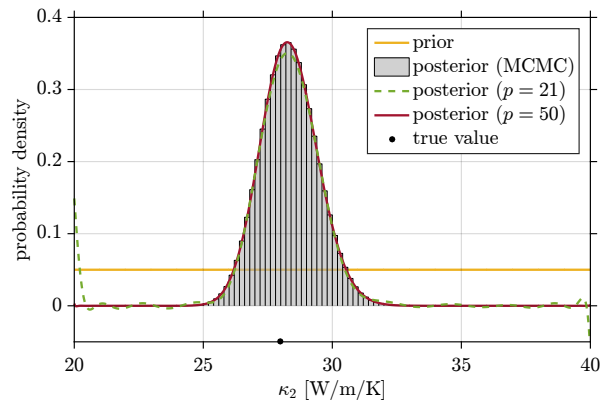
(b) SLE with $p = 50$.

(c) SLE with $p = 50$.

Figure 10: 2D IHCP: Joint posterior.

(a) Thermal conductivity κ_1.

(b) Thermal conductivity κ_2.

Figure 11: 2D IHCP: Posterior marginals.

25

Table 4: 2D IHCP: Statistical quantities.

	K	p	ϵ_{LOO}	$Z\ [10^{-1}]$	$\mathbb{E}[\kappa_1|\boldsymbol{T}]$	$\mathbb{E}[\kappa_2|\boldsymbol{T}]$	$\text{Std}[\kappa_1|\boldsymbol{T}]$	$\text{Std}[\kappa_2|\boldsymbol{T}]$	$\rho[\kappa_1,\kappa_2|\boldsymbol{T}]$
	5×10^2	5	8.24×10^{-1}	8.45	31.33	28.36	1.74	1.33	0.28
	1×10^3	9	6.08×10^{-1}	7.81	31.40	28.22	2.02	1.53	0.15
SLE	5×10^3	21	1.50×10^{-1}	7.47	31.32	28.13	2.16	1.61	0.34
	1×10^4	29	5.79×10^{-2}	7.21	31.56	28.30	1.61	1.39	-0.05
	5×10^4	35	1.63×10^{-2}	7.18	31.62	28.34	1.24	1.08	-0.75
	1×10^5	50	7.56×10^{-4}	7.18	31.62	28.33	1.26	1.10	-0.68
(MC)MC				7.17	31.62	28.33	1.26	1.09	-0.68

samples to the experimental design. While this is reassuring to know, it does not help in solving higher-dimensional problems for which the computation of high-order expansions is exacerbated by the curse of dimensionality. Hence, now we want to investigate the applicability of SLEs and aSLEs in an inverse problem of moderate dimension.

An IHCP in two spatial dimensions with six unknown conductivities is considered in this section. The setup of the problem is shown in Fig. 12. The $M = 6$ unknown conductivities $\boldsymbol{\kappa} = (\kappa_1,\ldots,\kappa_6)^\top$ are inferred with $N = 20$ noisy measurements $\boldsymbol{T} = (T_1,\ldots,T_{20})^\top$ of the temperature field $\tilde{T}$. We set $\kappa_0 = 30\,\text{W/m/K}$ and $\boldsymbol{\kappa} = (20, 24, \ldots, 40)^\top\,\text{W/m/K}$. The prior is set to a multivariate lognormal distribution $\pi(\boldsymbol{\kappa}) = \prod_{i=1}^6 \pi(\kappa_i)$ with independent marginals $\pi(\kappa_i) = \mathcal{LN}(\kappa_i|\mu_0,\sigma_0^2)$ with $\mu_0 = 30\,\text{W/m/K}$ and $\sigma_0 = 6\,\text{W/m/K}$. These parameters describe the mean $\mu_0 = \mathbb{E}[\kappa_i]$ and standard deviation $\sigma_0 = \text{Std}[\kappa_i]$ of the lognormal prior. They are related to the parameters of the associated normal distribution $\mathcal{N}(\log(\kappa_i)|\lambda_0,\varsigma_0^2)$ via $\mu_0 = \exp(\lambda_0 + \varsigma_0^2/2)$ and $\sigma_0^2 = (\exp(\varsigma_0^2) - 1)\exp(2\lambda_0 + \varsigma_0^2)$. Otherwise than that, the problem setup is exactly as described in the previous section, i.e. the likelihood function is given as $\mathcal{L}(\boldsymbol{\kappa}) = \mathcal{N}(\boldsymbol{T}|\mathcal{M}(\boldsymbol{\kappa}),\boldsymbol{\Sigma})$. In accordance with this setup, in the following synthetic data are simulated and analyzed in order to compute the joint posterior $\pi(\boldsymbol{\kappa}|\boldsymbol{T}) = Z^{-1}\mathcal{L}(\boldsymbol{\kappa})\pi(\boldsymbol{\kappa})$.
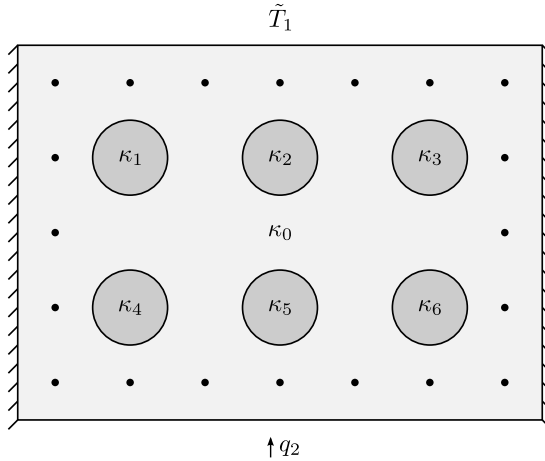

Figure 12: 6D IHCP: Heat conduction setup.

5.4.1. Posterior density

The unknowns are represented as $\kappa_i = \exp(\lambda_0 + \varsigma_0\xi_i)$ in terms of the standardized variables $\xi_i \in \mathbb{R}$ with Gaussian weight functions $\mathcal{N}(\xi_i|0,1)$. A spectral expansion $\hat{\mathcal{L}}_p$ in tensorized Hermite polynomials is then computed for $p = 5$ and $K = 5\times10^4$. The errors of the likelihood approximation are estimated as $\epsilon_{\text{Emp}} = 8.81\times10^{-1}$ and $\epsilon_{\text{LOO}} = 9.14\times10^{-1}$. As compared to the low-dimensional examples that were

studied before, these are large errors. An auxiliary reference density $g(\boldsymbol{\kappa}) = \prod_{i=1}^{6} g(\kappa_i)$ is then constructed as a multivariate lognormal with independent marginals $g(\kappa_i) = \mathcal{LN}(\kappa_i | \mu_i, \sigma_i^2)$. The parameters of the latter are chosen as the means $\mu_i = \mathbb{E}[\kappa_i | \boldsymbol{T}]$ and standard deviations $\sigma_i = \mathrm{Std}[\kappa_i | \boldsymbol{T}]$ of the posterior surrogate corresponding to the coefficients of SLE $\hat{\mathcal{L}}_p$. We remark that this is a simple two-step procedure and that a more refined usage of the reference change would certainly lead to more sophisticated approaches. Subsequently, an aSLE $\hat{\mathcal{G}}_p$ with $p = 5$ and $K = 5 \times 10^4$ is computed. The errors amount to $\epsilon_{\mathrm{Emp}} = 4.81 \times 10^{-1}$ and $\epsilon_{\mathrm{LOO}} = 6.24 \times 10^{-1}$. Notwithstanding that these errors are smaller than the corresponding errors of the SLE, they are still large as compared to the previous examples. Since these errors are now measured with respect to the auxiliary density which is expectedly closer to the true posterior than the prior is, the aSLE presumably leads to a more accurate posterior surrogate.

From the previously computed SLE $\hat{\mathcal{L}}(\boldsymbol{\kappa})$ and the aSLE $\hat{\mathcal{G}}_p(\boldsymbol{\kappa})$ approximations of the joint posterior density are computed via Eqs. (39) and (55). The obtained surrogates $\pi(\boldsymbol{\kappa} | \boldsymbol{T}) \approx \hat{\mathcal{L}}_p(\boldsymbol{\kappa}) \pi(\boldsymbol{\kappa}) / b_{\boldsymbol{0}}$ and $\pi(\boldsymbol{\kappa} | \boldsymbol{T}) \approx \hat{\mathcal{G}}_p(\boldsymbol{\kappa}) g(\boldsymbol{\kappa}) / b_{\boldsymbol{0}}^g$ are now compared to each other. We start with the one-dimensional marginals that can be compiled by collecting terms from the full expansions based on Eq. (45). For $j = 1, \ldots, 6$ the marginals $\pi(\kappa_j | \boldsymbol{T})$ that are extracted that way are shown in Fig. 13. The marginal priors $\pi(\kappa_j)$ and the auxiliary densities $g(\kappa_j)$ are shown, too. While the marginals that are taken from the SLE slightly deviate from their MCMC counterparts, the marginals based on the aSLE match their references perfectly well. The reason is that the posterior can be easier represented as a small adjustment of the auxiliary density than as a large correction to the prior. Thus, with a same expansion order the posterior is more accurately represented through the aSLE than through the SLE. Regarding the size if the error estimates it is surprising that the marginals can be retrieved that well with the aSLE. Even though the SLE-based posterior approximations can hardly be interpreted as proper probability densities, i.e. they conspicuously take on negative values, the moments are recovered sufficiently well for the construction of the auxiliary reference density.

On the basis of Eq. (47) the two-dimensional posterior marginals $\pi(\kappa_j, \kappa_k | \boldsymbol{T})$ can be constructed from the full expansions. For $j = 3$ and $k = 4$ the posterior marginal for the SLE $\hat{\mathcal{L}}_p$ is shown in Fig. 14a. The same two-dimensional distribution is depicted in Fig. 14b for the aSLE $\hat{\mathcal{G}}_p$. A histogram of the MCMC sample is provided in Fig. 14c as a reference. As already found in Figs. 13c and 13d for instance, in Fig. 14 the aSLE-based surrogate appears to be almost exact whereas the SLE-based one is flattened out. Since the aSLE captures the true posterior density more accurately than the SLE, we expect similar findings for the posterior moments.

5.4.2. Quantities of interest

Finally we compute the model evidence and the first posterior moments with the aid of Eqs. (40) and (53) and Eqs. (48) to (50). For the aSLE $\hat{\mathcal{G}}_p$ the analysis proceeds analogously to the SLE $\hat{\mathcal{L}}_p$. In Table 5 a summary of the results is given. As it can be taken from the table, the aSLE consistently gives more accurate estimates of the reference values. This fulfills our earlier expectations. Regarding the inaccuracy of the SLE-based posterior marginals and the concerns about interpreting them as probability densities, the quality of the SLE-based estimates of the moments surpasses our expectations. In particular, the estimated standard deviations are more accurate than the surrogate marginals suggest, e.g. the ones shown in Figs. 13a and 13f. Similar as for the posterior density, we have to conclude that the normalized LOO error does not give conclusive information about the accuracy of the first posterior moments. Nevertheless, it is remarked that the use of resampling methods still ensures a robust fit, i.e. it protects against overfitting.


6. Concluding remarks

A spectral approach to Bayesian inference that focuses on the surrogate modeling of the posterior density was devised. The likelihood was expanded in terms of polynomials that are orthogonal with respect to the prior weight. Ensuing from this spectral likelihood expansion (SLE), the joint posterior density was expressed as the prior that acts the reference density times a polynomial correction term. The normalization factor of the posterior emerged as the zeroth SLE coefficient and the posterior marginals were shown to be easily accessible through sub-expansions of the SLE. Closed-form expressions for the first posterior moments in
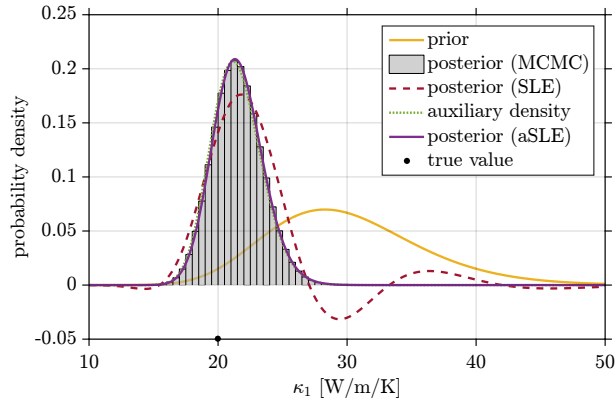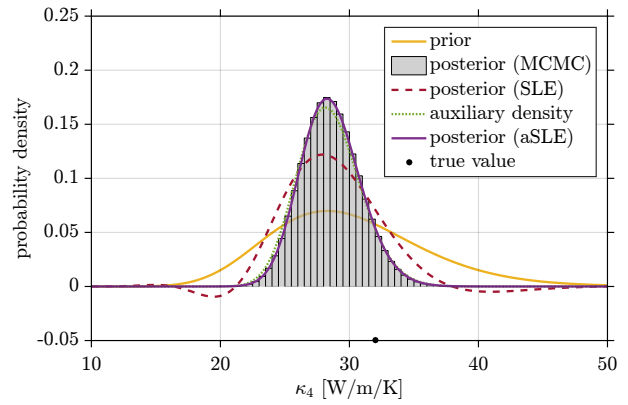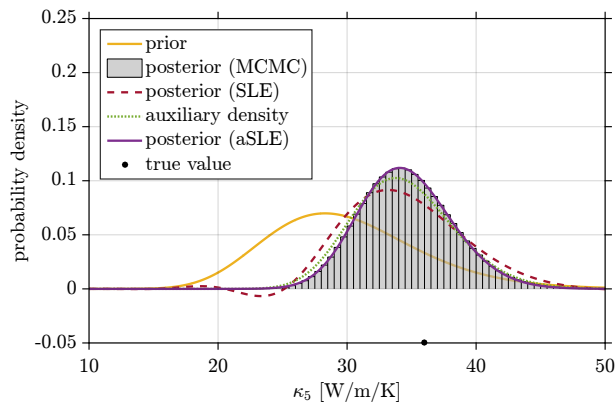
(a) Thermal conductivity κ_1.
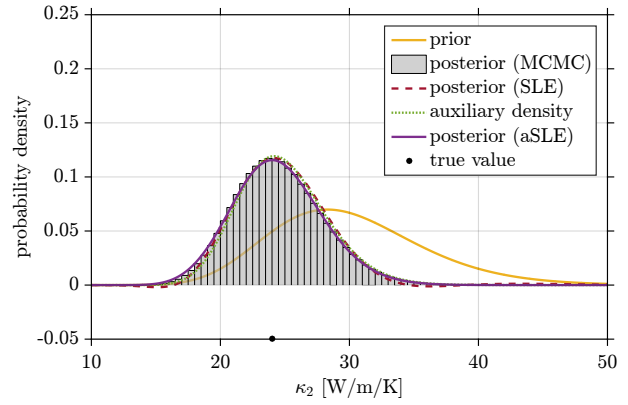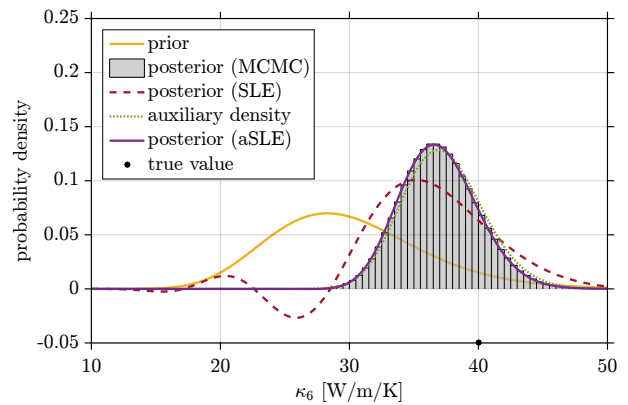
(b) Thermal conductivity κ_2.
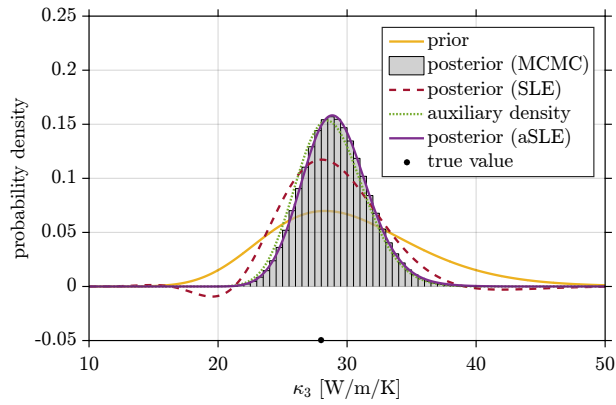
(c) Thermal conductivity κ_3.
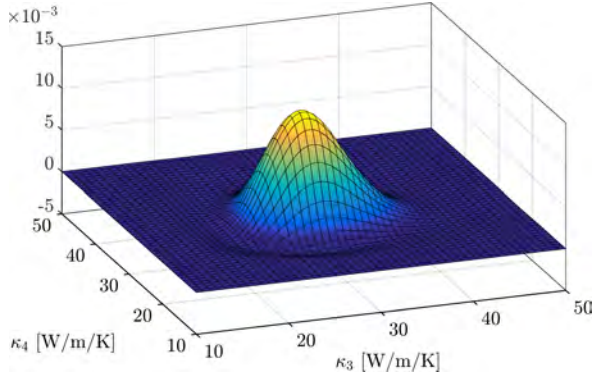
(d) Thermal conductivity κ_4.
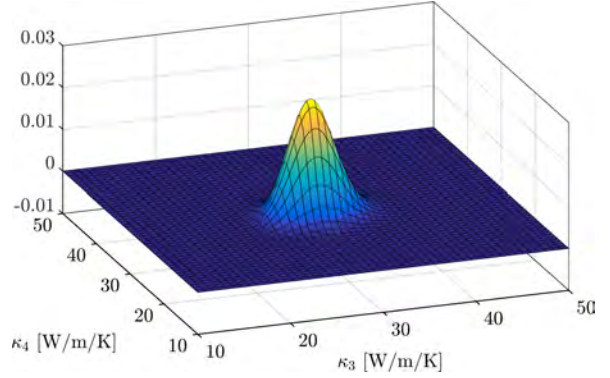
(e) Thermal conductivity κ_5.
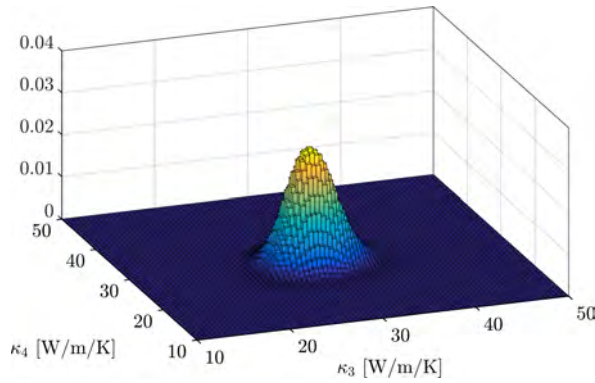
(f) Thermal conductivity κ_6.

Figure 13: 6D IHCP: Posterior marginals.

28

(a) SLE with $p = 5$.

(b) aSLE with $p = 5$.

(c) MCMC reference sample.

Figure 14: 6D IHCP: Posterior marginals.

Table 5: 6D IHCP: Statistical quantities.

	Z $[10^{-3}]$	$\mathbb{E}[\kappa_1|\boldsymbol{T}]$	$\mathbb{E}[\kappa_2|\boldsymbol{T}]$	$\mathbb{E}[\kappa_3|\boldsymbol{T}]$	$\mathbb{E}[\kappa_4|\boldsymbol{T}]$	$\mathbb{E}[\kappa_5|\boldsymbol{T}]$	$\mathbb{E}[\kappa_6|\boldsymbol{T}]$
SLE	4.04	21.42	24.86	28.79	28.45	34.43	37.27
aSLE	3.68	21.53	24.48	29.16	28.57	34.59	36.95
(MC)MC	3.65	21.52	24.57	29.11	28.56	34.64	37.00

	$\text{Std}[\kappa_1|\boldsymbol{T}]$	$\text{Std}[\kappa_2|\boldsymbol{T}]$	$\text{Std}[\kappa_3|\boldsymbol{T}]$	$\text{Std}[\kappa_4|\boldsymbol{T}]$	$\text{Std}[\kappa_5|\boldsymbol{T}]$	$\text{Std}[\kappa_6|\boldsymbol{T}]$	$\rho[\kappa_1,\kappa_2|\boldsymbol{T}]$
SLE	1.95	3.43	2.63	2.43	3.96	3.13	-0.40
aSLE	1.94	3.56	2.61	2.33	3.62	2.99	-0.44
(MC)MC	1.93	3.48	2.56	2.31	3.64	3.00	-0.47

	$\rho[\kappa_1,\kappa_3|\boldsymbol{T}]$	$\rho[\kappa_1,\kappa_4|\boldsymbol{T}]$	$\rho[\kappa_1,\kappa_5|\boldsymbol{T}]$	$\rho[\kappa_1,\kappa_6|\boldsymbol{T}]$	$\rho[\kappa_2,\kappa_3|\boldsymbol{T}]$	$\rho[\kappa_2,\kappa_4|\boldsymbol{T}]$	$\rho[\kappa_2,\kappa_5|\boldsymbol{T}]$
SLE	0.19	-0.39	-0.28	0.05	-0.40	-0.18	-0.30
aSLE	-0.01	-0.29	-0.03	0.10	-0.48	-0.17	-0.28
(MC)MC	-0.02	-0.32	-0.03	0.09	-0.48	-0.17	-0.31

	$\rho[\kappa_2,\kappa_6|\boldsymbol{T}]$	$\rho[\kappa_3,\kappa_4|\boldsymbol{T}]$	$\rho[\kappa_3,\kappa_5|\boldsymbol{T}]$	$\rho[\kappa_3,\kappa_6|\boldsymbol{T}]$	$\rho[\kappa_4,\kappa_5|\boldsymbol{T}]$	$\rho[\kappa_4,\kappa_6|\boldsymbol{T}]$	$\rho[\kappa_5,\kappa_6|\boldsymbol{T}]$
SLE	-0.09	-0.00	0.22	-0.22	-0.20	0.24	-0.11
aSLE	-0.13	0.11	-0.02	-0.32	-0.24	0.13	-0.24
(MC)MC	-0.16	0.10	-0.03	-0.34	-0.26	0.12	-0.24

terms of the low-order spectral coefficients were given. Posterior uncertainty propagation through general quantities of interest was established via a postprocessing of the higher-order coefficients. The semi-analytic reformulation of Bayesian inference was founded on the theory and practice of metamodeling based on polynomial chaos expansions. This allows one to compute the SLE coefficients by solving a linear least squares problem. An analysis of the advantages and disadvantages of the proposed method eventually motivated a change of the reference density. While the expansion of the posterior in terms of the prior may require substantial modifications, its representation with respect to an auxiliary density many only require minor tweaks.

The possibilities and difficulties that arise from the problem formulation were exhaustively discussed and numerically demonstrated. Fitting a parametric distribution to random data and identifying the thermal properties of a composite material served as benchmark problems. These numerical experiments proved that spectral Bayesian inference works in principle and they provided insight into the mechanisms involved. The convergence behavior of the SLE was studied based on the leave-one-out error. It was found that high-degree SLEs are necessary in order to accurately represent the likelihood function and the joint posterior density, whereas lower-order SLEs are sufficient in order to extract the low-level quantities of interest. A change of the reference density allowed for reducing the order of the corrections required in order to represent the posterior with respect to the prior. This helped in alleviating the curse of dimensionality to some extent.

In turn, a number of follow-up questions were given rise to. While the leave-one-out error performs well in quantifying the prediction errors of the SLE, it turned out to be of limited use with regard to the errors of the corresponding posterior surrogate and its marginals. A critical question thus relates to a means to assess the errors of these quantities and to diagnose their convergence. This would assist in choosing experimental designs of a sufficient size. Also, it would be desirable to be able to quantify the estimation errors of single expansion coefficients. This would support the assessment of the efficiency and scalability of the approach and the fair comparison with Monte Carlo, importance and Markov chain Monte Carlo sampling. Another question is whether a constrained optimization problem can be formulated that naturally respects all prior restrictions. This would remedy the potential problem of illegitimate values of the posterior moments. In order to handle a broader spectrum of statistical problems, SLEs would have to be extended to dependent prior distributions and noisy likelihood functions. For increasing the computational efficiency beyond the change of the reference density, it is conceivable to deploy advanced techniques from metamodeling and machine learning. This includes piecewise polynomial models, expansions in a favorable basis and the use of sparsity-promoting regression techniques. Yet another important issue concerns the practical applicability of the presented framework to problems with higher-dimensional parameter spaces. In future research efforts we will try to address the abovementioned issues and to answer this principal question.

Appendix A. Univariate polynomials

The main properties of two classical orthogonal families of polynomials were shortly summarized in Table 1, i.e. the domain of definition, the associated weight function and the norm. Their first six members of these univariate Hermite polynomials $\{H_\alpha\}_{\alpha \in \mathbb{N}}$ and Legendre polynomials $\{P_\alpha\}_{\alpha \in \mathbb{N}}$ are listed in Table A.1. Higher order members can be defined via recursive or differential relations. These polynomials can be used for the construction of the multivariate polynomial basis $\{\Psi_\alpha\}_{\alpha \in \mathbb{N}}$ in Eq. (17). Note that this orthonormal basis is normalized via $\Psi_\alpha = H_\alpha/\sqrt{\alpha!}$ or $\Psi_\alpha = P_\alpha/\sqrt{1/(2\alpha+1)}$.

Appendix B. Low-order QoIs

The representation of six low-order QoIs in terms of the normalized Hermite and Legendre polynomials is given in Table B.2 below. Those expansions can be used in order to compute the first posterior moments, e.g. as shown in Eqs. (48) to (50). Note that the representations in the orthonormal bases directly follow from a change of basis and the substitutions $H_\alpha = \sqrt{\alpha!}\Psi_\alpha$ and $P_\alpha = \sqrt{1/(2\alpha+1)}\Psi_\alpha$.

[1] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, USA, 2005.

Table A.1: Low-order polynomials.

α	$H_\alpha(x)$, $x \in \mathbb{R}$	$P_\alpha(x)$, $x \in [-1, 1]$
0	1	1
1	x	x
2	$x^2 - 1$	$(3x^2 - 1)/2$
3	$x^3 - 3x$	$(5x^3 - 3x)/2$
4	$x^4 - 6x^2 + 3$	$(35x^4 - 30x^2 + 3)/8$
5	$x^5 - 10x^3 + 15x$	$(63x^5 - 70x^3 + 15x)/8$

Table B.2: Low-order QoIs.

QoI	Hermite expansion	Legendre expansion
1	Ψ_0	Ψ_0
x	Ψ_1	$\Psi_1/\sqrt{3}$
x^2	$\sqrt{2}\Psi_2 + \Psi_0$	$(2\Psi_2/\sqrt{5} + \Psi_0)/3$
x^3	$\sqrt{6}\Psi_3 + 3\Psi_1$	$(2\Psi_3/\sqrt{7} + 3\Psi_1/\sqrt{3})/5$
x^4	$2\sqrt{6}\Psi_4 + 6\sqrt{2}\Psi_2 + 3\Psi_0$	$(8\Psi_4/3 + 20\Psi_2/\sqrt{5} + 7\Psi_0)/35$
x^5	$2\sqrt{30}\Psi_5 + 10\sqrt{6}\Psi_3 + 15\Psi_1$	$(8\Psi_5/\sqrt{11} + 28\Psi_3/\sqrt{7} + 27\Psi_1/\sqrt{3})/63$

[2] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer, New York, 2005.

[3] R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Computational Science and Engineering. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, USA, 2014.

[4] T. J. Sullivan. *Introduction to Uncertainty Quantification*, volume 63 of *Texts in Applied Mathematics*. Springer International Publishing, Cham, Switzerland, 2015.

[5] R. Hadidi and N. Gucunski. Probabilistic approach to the solution of inverse problems in civil engineering. *J. Comput. Civil Eng.*, 22(6):338–347, 2008.

[6] J. L. Beck. Bayesian system identification based on probability logic. *Struct. Control. Health Monit.*, 17(7):825–847, 2010.

[7] K.-V. Yuen and S.-C. Kuok. Bayesian methods for updating dynamic models. *Appl. Mech. Rev.*, 64(1):010802:1–18, 2011.

[8] M. Evans and T. Swartz. Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Stat. Sci.*, 10(3):254–272, 1995.

[9] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton, Florida, USA, 1996.

[10] S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, editors. *Handbook of Markov Chain Monte Carlo*. Handbooks of Modern Statistical Methods. Chapman & Hall/CRC, Boca Raton, Florida, USA, 2011.

[11] J. Beck and S. Au. Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. *J. Eng. Mech.*, 128(4):380–391, 2002.

[12] J. Ching and Y. Chen. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *J. Eng. Mech.*, 133(7):816–832, 2007.

[13] S. H. Cheung and J. L. Beck. Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *J. Eng. Mech.*, 135(4):243–255, 2009.

[14] I. Boulkaibet, L. Mthembu, T. Marwala, M. I. Friswell, and S. Adhikari. Finite element model updating using the shadow hybrid Monte Carlo technique. *Mech. Syst. Sig. Process.*, 52–53:115–132, 2015.

[15] J. B. Nagel and B. Sudret. Hamiltonian Monte Carlo and borrowing strength in hierarchical inverse problems. *ASCE-ASME J. Risk Uncertainty Eng. Syst., Part A: Civ. Eng.* (in press), 2016.

[16] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne. Combining field data and computer simulations for calibration and prediction. *SIAM J. Sci. Comput.*, 26(2):448–466, 2004.

[17] D. Higdon, J. D. McDonnell, N. Schunck, J. Sarich, and S. M. Wild. A Bayesian approach for parameter estimation and prediction using a computationally intensive model. *J. Phys. G: Nucl. Part. Phys.*, 42(3):034009:1–18, 2015.

[18] Y. M. Marzouk, H. N. Najm, and L. A. Rahn. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comput. Phys.*, 224(2):560–586, 2007.

[19] Y. M. Marzouk and H. N. Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.*, 228(6):1862–1902, 2009.

[20] Y. Marzouk and D. Xiu. A stochastic collocation approach to Bayesian inference in inverse problems. *Commun. Comput.*

Phys., 6(4):826–847, 2009.

[21] J. T. Ormerod and M. P. Wand. Explaining variational approximations. *Am. Stat.*, 64(2):140–153, 2010.

[22] C. W. Fox and S. J. Roberts. A tutorial on variational Bayesian inference. *Artif. Intell. Rev.*, 38(2):85–95, 2012.

[23] S. Sun. A review of deterministic approximate inference techniques for Bayesian machine learning. *Neural Comput. Appl.*, 23(7–8):2039–2050, 2013.

[24] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Mach. Learn.*, 37(2):183–233, 1999.

[25] T. S. Jaakkola and M. I. Jordan. Bayesian parameter estimation via variational methods. *Stat. Comput.*, 10(1):25–37, 2000.

[26] M. A. Chappell, A. R. Groves, B. Whitcher, and M. W. Woolrich. Variational Bayesian inference for a nonlinear forward model. *IEEE Trans. Signal Process.*, 57(1):223–236, 2009.

[27] B. Jin and J. Zou. Hierarchical Bayesian inference for ill-posed problems via variational method. *J. Comput. Phys.*, 229(19):7317–7343, 2010.

[28] T. A. El Moselhy and Y. M. Marzouk. Bayesian inference with optimal maps. *J. Comput. Phys.*, 231(23):7815–7850, 2012.

[29] C. Schwab and A. M. Stuart. Sparse deterministic approximation of Bayesian inverse problems. *Inverse Prob.*, 28(4):045003:1–32, 2012.

[30] C. Schillings and C. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Prob.*, 29(6):065011:1–28, 2013.

[31] J. P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover Books on Mathematics. Dover Publications, Mineola, New York, USA, 2nd edition, 2001.

[32] D. A. Kopriva. *Implementing Spectral Methods for Partial Differential Equations: Algorithms for Scientists and Engineers*. Scientific Computation. Springer, Dordrecht, Netherlands, 2009.

[33] J. Shen, T. Tang, and L.-L. Wang. *Spectral Methods: Algorithms, Analysis and Applications*, volume 41 of *Springer Series in Computational Mathematics*. Springer-Verlag Berlin Heidelberg, 2011.

[34] O. Christensen and K. L. Christensen. *Approximation Theory: From Taylor Polynomials to Wavelets*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, Massachusetts, USA, 2004.

[35] R. M. Trigub and E. S. Belinsky. *Fourier Analysis and Approximation of Functions*. Springer, Dordrecht, Netherlands, 2004.

[36] L. N. Trefethen. *Approximation Theory and Approximation Practice*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, USA, 2013.

[37] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, USA, 1995.

[38] Å. Björck. *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, USA, 1996.

[39] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Statistics for Engineering and Information Science. Springer, New York, 2nd edition, 2000.

[40] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, 2nd edition, 2009.

[41] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate Bayesian computational methods. *Stat. Comput.*, 22(6):1167–1180, 2012.

[42] M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz. Approximate Bayesian computation. *PLoS Comput. Biol.*, 9(1):e1002803:1–10, 2013.

[43] G. Taraldsen and B. H. Lindqvist. Improper priors are not improper. *Am. Stat.*, 64(2):154–158, 2010.

[44] Kitanidis. P. K. Generalized priors in Bayesian inversion problems. *Adv. Water Resour.*, 36:3–10, 2012.

[45] A. Gelman, A. Jakulin, M. G. Pittau, and Y.-S. Su. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.*, 2(4):1360–1383, 2008.

[46] J. A. Fúquene, J. D. Cook, and L. R. Pericchi. A case for robust Bayesian priors with applications to clinical trials. *Bayesian Anal.*, 4(4):817–846, 2009.

[47] J. B. Nagel and B. Sudret. Bayesian multilevel model calibration for inverse problems under uncertainty with perfect data. *J. Aerosp. Inf. Syst.*, 12(1):97–113, 2015.

[48] J. B. Nagel and B. Sudret. A unified framework for multilevel uncertainty quantification in Bayesian inverse problems. *Probab. Eng. Mech.*, 43:68–84, 2016.

[49] S. Jackman. *Bayesian Analysis for the Social Sciences*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, West Sussex, UK, 2009.

[50] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Texts in Statistical Science. CRC Press, Boca Raton, Florida, USA, 3rd edition, 2014.

[51] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[52] O. G. Ernst, B. Sprungk, and H.-J. Starkloff. Bayesian inverse problems and Kalman filters. In S. Dahlke, W. Dahmen, M. Griebel, W. Hackbusch, K. Ritter, R. Schneider, C. Schwab, and H. Yserentant, editors, *Extraction of Quantifiable Information from Complex Systems*, volume 102 of *Lecture Notes in Computational Science and Engineering*, pages 133–159. Springer International Publishing, Cham, Switzerland, 2014.

[53] R. E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numer.*, 7:1–49, 1998.

[54] S. T. Tokdar and R. E. Kass. Importance sampling: a review. *Wiley Interdiscip. Rev. Comput. Stat.*, 2(1):54–60, 2010.

[55] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast

computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[56] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[57] M. K. Cowles and B. P. Carlin. Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Am. Stat. Assoc.*, 91(434):883–904, 1996.

[58] S. P. Brooks and G. O. Roberts. Convergence assessment techniques for Markov chain Monte Carlo. *Stat. Comput.*, 8(4):319–335, 1998.

[59] A. Vehtari and J. Ojanen. A survey of Bayesian predictive methods for model assessment, selection and comparison. *Stat. Surv.*, 6:142–228, 2012.

[60] J. Beck and K. Yuen. Model selection using response measurements: Bayesian probabilistic approach. *J. Eng. Mech.*, 130(2):192–203, 2004.

[61] K.-V. Yuen. Recent developments of Bayesian model class selection and applications in civil engineering. *Struct. Saf.*, 32(5):338–346, 2010.

[62] C. Han and B. P Carlin. Markov chain Monte Carlo methods for computing Bayes factors. *J. Am. Stat. Assoc.*, 96(455):1122–1132, 2001.

[63] P. Dellaportas, J. J. Forster, and I. Ntzoufras. On Bayesian model and variable selection using MCMC. *Stat. Comput.*, 12(1):27–36, 2002.

[64] A. O'Hagan and J. F. C. Kingman. Curve fitting and optimal design for prediction. *J. R. Stat. Soc. Ser. B*, 40(1):1–42, 1978.

[65] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Stat. Sci.*, 4(4):409–423, 1989.

[66] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, New York, 1991.

[67] T. J. Santner, B. J. Williams, and W. I. Notz. *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer, New York, 2003.

[68] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, Massachusetts, USA, 2006.

[69] O. P. Le Maître and O. M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Scientific Computation. Springer, Dordrecht, Netherlands, 2010.

[70] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, Princeton, New Jersey, USA, 2010.

[71] H. Stahl and V. Totik. *General Orthogonal Polynomials*, volume 43 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, UK, 1992.

[72] W. Gautschi. *Orthogonal Polynomials: Computation and Approximation*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, UK, 2004.

[73] D. Jackson. *Fourier Series and Orthogonal Polynomials*. Dover Books on Mathematics. Dover Publications, Inc., Mineola, New York, USA, 2004.

[74] D. Xiu and G. E. Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644, 2002.

[75] J. A. S. Witteveen, S. Sarkar, and H. Bijl. Modeling physical uncertainties in dynamic stall induced fluid–structure interaction of turbine blades using arbitrary polynomial chaos. *Comput. Struct.*, 85(11–14):866–878, 2007.

[76] G. E. P. Box and N. R. Draper. *Response Surfaces, Mixtures, and Ridge Analyses*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2nd edition, 2007.

[77] I. Babuška, R. Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2):800–825, 2004.

[78] D. Xiu and J. Shen. Efficient stochastic Galerkin methods for random diffusion equations. *J. Comput. Phys.*, 228(2):266–281, 2009.

[79] D. Xiu and J. S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2005.

[80] D. Xiu. Efficient collocational approach for parametric uncertainty analysis. *Commun. Comput. Phys.*, 2(2):293–309, 2007.

[81] O. P. Le Maître, O. M. Knio, H. N. Najm, and R. G. Ghanem. A stochastic projection method for fluid flow: I. Basic formulation. *J. Comput. Phys.*, 173(2):481–511, 2001.

[82] O. P. Le Maître, M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio. A stochastic projection method for fluid flow: II. Random process. *J. Comput. Phys.*, 181(1):9–44, 2002.

[83] M. Berveiller, B. Sudret, and M. Lemaire. Stochastic finite element: a non intrusive approach by regression. *Eur. J. Comput. Mech.*, 15(1–3):81–92, 2006.

[84] G. Blatman and B. Sudret. Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *C. R. Mécanique*, 336(6):518–523, 2008.

[85] C. Giraud. *Introduction to High-Dimensional Statistics*, volume 139 of *Chapman & Hall/CRC Monographs on Statistics & Applied Probability*. CRC Press, Boca Raton, Florida, USA, 2015.

[86] T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*, volume 143 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, Florida, USA, 2015.

[87] G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *J. Comput. Phys.*, 230(6):2345–2367, 2011.

[88] A. Doostan and H. Owhadi. A non-adapted sparse approximation of PDEs with stochastic inputs. *J. Comput. Phys.*,

230(8):3015–3034, 2011.

[89] L. Yan, L. Guo, and D. Xiu. Stochastic collocation algorithms using ℓ_1-minimization. *Int. J. Uncertainty Quantification*, 2(3):279–293, 2012.

[90] L. Mathelin and K. A. Gallivan. A compressed sensing approach for partial differential equations with random input data. *Commun. Comput. Phys.*, 12(4):919–954, 2012.

[91] K. Sargsyan, C. Safta, H. N. Najm, B. J. Debusschere, D. Ricciuto, and P. Thornton. Dimensionality reduction for complex models via Bayesian compressive sensing. *Int. J. Uncertainty Quantification*, 4(1):63–93, 2014.

[92] J. Ray, Z. Hou, M. Huang, K. Sargsyan, and L. Swiler. Bayesian calibration of the community land model using surrogates. *SIAM/ASA J. Uncertainty Quantification*, 3(1):199–233, 2015.

[93] G. Karagiannis and G. Lin. Selection of polynomial chaos bases via Bayesian model uncertainty methods with applications to sparse approximation of PDEs with stochastic inputs. *J. Comput. Phys.*, 259:114–134, 2014.

[94] G. Karagiannis, B. A. Konomi, and G. Lin. A Bayesian mixed shrinkage prior procedure for spatial–stochastic basis selection and evaluation of gPC expansions: Applications to elliptic SPDEs. *J. Comput. Phys.*, 284:528–546, 2015.

[95] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13(5):819–834, 2013.

[96] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Analysis of discrete L^2 projection on polynomial spaces with random evaluations. *Found. Comput. Math.*, 14(3):419–456, 2014.

[97] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone. Discrete least squares polynomial approximation with random evaluations – application to parametric and stochastic elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 49(3):815–837, 2015.

[98] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Ann. Stat.*, 32(2):407–499, 2004.

[99] T. Hesterberg, N. H. Choi, L. Meier, and C. Fraley. Least angle and ℓ_1 penalized regression: A review. *Stat. Surv.*, 2:61–93, 2008.

[100] S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Stat. Surv.*, 4:40–79, 2010.

[101] G. A. F. Seber and A. J. Lee. *Linear Regression Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2nd edition, 2003.

[102] S. Efromovich. Orthogonal series density estimation. *Wiley Interdiscip. Rev. Comput. Stat.*, 2(4):467–476, 2010.

[103] M. Jiang and S. B. Provost. Improved orthogonal polynomial density estimates. *J. Stat. Comput. Simul.*, 81(11):1495–1516, 2011.

[104] R. A. Johnson. An asymptotic expansion for posterior distributions. *Ann. Math. Stat.*, 38(6):1899–1906, 1967.

[105] R. A. Johnson. Asymptotic expansions associated with posterior distributions. *Ann. Math. Stat.*, 41(3):851–864, 1970.

[106] R. C. Weng. A Bayesian Edgeworth expansion by Stein's identity. *Bayesian Anal.*, 5(4):741–763, 2010.

[107] R. C. Weng and C.-H. Hsu. A study of expansions of posterior distributions. *Commun. Stat. Theory Methods*, 42(2):346–364, 2013.

[108] H. R. B. Orlande, M. J. Colaço, and G. S. Dulikravich. Approximation of the likelihood function in the Bayesian technique for the solution of inverse problems. *Inverse Prob. Sci. Eng.*, 16(6):677–692, 2008.

[109] A. Dietzel and P. Reichert. Bayesian inference of a lake water quality model by emulating its posterior density. *Water Resour. Res.*, 50(10):7626–7647, 2014.

[110] C. Soize and R. Ghanem. Physical systems with random uncertainties: Chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput.*, 26(2):395–410, 2004.

[111] D. E. Barndorff-Nielsen and D. R. Cox. *Asymptotic Techniques for Use in Statistics*, volume 31 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, London, UK, 1989.

[112] J. E. Kolassa. *Series Approximation Methods in Statistics*, volume 88 of *Lecture Notes in Statistics*. Springer, New York, 3rd edition, 2006.

[113] C. G. Small. *Expansions and Asymptotics for Statistics*, volume 115 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, Florida, USA, 2010.

[114] S. Marelli and B. Sudret. UQLab: A framework for uncertainty quantification in MATLAB. In M. Beer, S.-K. Au, and J. W. Hall, editors, *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, chapter 257, pages 2554–2563. American Society of Civil Engineers (ASCE), Reston, Virginia, USA, 2014.

[115] S. Marelli and B. Sudret. *UQLab User Manual: Polynomial Chaos Expansions*. Chair of Risk, Safety & Uncertainty Quantification, ETH Zürich, 0.9-104 edition, 2015.