Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Computational Physics 217 (2006) 63–81
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

On the construction and analysis of stochastic models: Characterization and propagation of the errors associated with limited data

Roger G. Ghanem [a,*], Alireza Doostan [b]

[a] *Aerospace and Mechanical Engineering, University of Southern California, 3620 S. Vermont, 210 KAP Hall USC, Los Angeles, CA 90089, USA*
[b] *Department of Civil Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA*

Received 28 July 2005; received in revised form 4 January 2006; accepted 19 January 2006
Available online 7 March 2006

Abstract

This paper investigates the predictive accuracy of stochastic models. In particular, a formulation is presented for the impact of data limitations associated with the calibration of parameters for these models, on their overall predictive accuracy. In the course of this development, a new method for the characterization of stochastic processes from corresponding experimental observations is obtained. Specifically, *polynomial chaos* representations of these processes are estimated that are consistent, in some useful sense, with the data. The estimated polynomial chaos coefficients are themselves characterized as random variables with known probability density function, thus permitting the analysis of the dependence of their values on further experimental evidence. Moreover, the error in these coefficients, associated with limited data, is propagated through a physical system characterized by a stochastic partial differential equation (SPDE). This formalism permits the rational allocation of resources in view of studying the possibility of validating a particular predictive model. A Bayesian inference scheme is relied upon as the logic for parameter estimation, with its computational engine provided by a Metropolis-Hastings Markov chain Monte Carlo procedure.

Keywords: Uncertainty quantification; Computational Bayesian analysis; Markov chain Monte Carlo; Stochastic finite elements; Parameter estimation; Polynomial chaos; Karhunen–Loeve

1. Introduction

Predictive science can be viewed as the extrapolation of observed behavior to new, yet to be observed, contexts. Observed behavior typically reflects controlled laboratory experiments or accumulated field observations while the unobserved context refers to a target design or behavior at some unobserved instant in time or space. The extrapolation from observed reality to predicted behavior is usually carried out by relying on

* Corresponding author. Tel.: +1 213 740 9528; fax: +1 213 740 2037.
E-mail addresses: ghanem@usc.edu (R.G. Ghanem), doostan@jhu.edu (A. Doostan).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2006.01.037

expert opinion which is typically available either as a list of actions predicated on operational conditions, or as a list of governing physical phenomena together with their associated mathematical forms. Either way, a procedure is typically developed for the operational implementation of this extrapolation, which could take one of various forms, ranging from a code of practice to a computer code. A number of factors contribute to a discrepancy between predicted and actual behavior. These clearly include limitations on the weight of experimentally observed evidence on which the predictive process is predicated. Such limitations stem from, in addition to observed variability in this evidence, its finiteness, both in terms of its temporal and spatial coverage as well as in terms of its limited quantity. In addition to these information-related errors, discrepancies between predictions and reality can also be attributed to shortcomings in the extrapolation process, be it driven by empirical or physical expert opinion, as well as by limitations on the implementation of this process, as manifested, for instance, through a computational model.

The characterization of an error budget that describes in quantitative terms the contribution of the above sources of error will permit the judicious allocation of resources between experimental, theoretical, and computational efforts in order to achieve a target confidence in the associated predictions.

The present paper adopts a probabilistic framework for characterizing uncertainty associated with the task of prediction. In doing so, it is emphasized that this probabilistic framework is merely adopted for convenience and does not describe an intrinsic property of these errors. This convenience is manifested in that a rigorous mathematical construction ensues that permits the structured statement of the problem.

In a series of previous publications, the authors have presented a mathematical framework for the characterization and propagation of uncertainty in physical systems [7,15–20,26]. That work is based on the adaptation of multiple Wiener integral representations [40,5] to finite-dimensional spaces and their implementation into a weighted residual scheme for the stochastic characterization of the solution of stochastic partial differential equations. The restriction of the representations to finite-dimensional uncertainty (i.e. stochastic processes characterized by a finite-dimensional random vector) permits the generalization of the Wiener constructions which had used polynomials orthogonal with respect to Gaussian measure. Extensions of that work to the Askey scheme were recently developed [41,42] together with extensions using non-orthogonal representations [1,3] and representations in terms of wavelets [23,24]. While the above extensions are limited to representations in terms of independent random variables, extensions using finite-dimensional dependent random vectors (such as appearing, for instance, in the finite-dimensional Karhunen–Loeve representation of an arbitrary stochastic process) have also been completed [36]. These so-called polynomial chaos expansions, coupled with stochastic projection mechanisms, provide a general method for characterizing the solution of problems of mathematical physics whose parameters have been described as stochastic processes. An analysis of the error associated with this method has already been developed [4,3]. The mathematical foundation of these developments, which permits their rigorous mathematical analysis, lies in functional analysis and in the observation that second-order random vectors (i.e. loosely speaking, those with finite second order statistics) form a Hilbert space.

Implicit in all these developments is the assumption that the parameters in the governing equations have been accurately characterized as stochastic processes. In this case, the formalism described above, based on chaos expansions, can be viewed as an efficient procedure to propagate the probabilistic measures from the system parameters to the solution, and as such appears to be an alternative to approaches based on Monte Carlo sampling or perturbation expansions. As noted previously in Section 1, it is often the case that system parameters are not known with enough resolution to permit their accurate characterization as stochastic processes. In such cases, the approach based on chaos developments provides a unique perspective on the problem as the sensitivity of the system parameters to additional information can be cast as perturbing their coordinates with respect to the polynomial chaos, which in turn is readily described as perturbations in Hilbert spaces. The impact of these perturbations on the chaos coordinates of the solution to the governing equations can then be viewed as quantifying the impact of refining the probabilistic measure of the data on the predictive capability of the mathematical model. Initial efforts in this direction have relied on the maximum-likelihood arguments to compute estimates of the chaos coordinates of parameters from associated statistical samples [8]. The present work addresses this same problem using a Bayesian framework for parameter estimation. The benefit of this approach lies in its ability to characterize the statistics of the estimates, thus enabling the determination of their accuracy and their sensitivity to further data. In addition, a propagation of the error asso-

ciated with these estimates to the predictions of the stochastic governing equations is developed. This clearly provides an essential ingredient for any model validation process. It is worth mentioning that the present approach relaxes some restrictive assumptions of current procedures, [2,21], to characterizing the random parameters of a stochastic system from limited data, in the sense that no particular form for the underlying stochastic parameter is assumed. Moreover, two important features of the proposed technique are in its ability to estimate the error associated with data limitation and also its adaptation to the already well-developed stochastic Galerkin schemes.

The paper is organized as follows. Section 2 summarizes the task to be undertaken together with some of the salient challenges. Following that, a Bayesian parameter estimator is employed in Section 3 to estimate those chaos coefficients together with their asymptotic standard errors. Sections 4 and 5 focus on the representation and propagation of the uncertainty associated with the above Bayesian estimates to the response of a particular class of SPDE's. Finally, in Section 6, a numerical experiment is performed to demonstrate and delineate the performance of the proposed formalism.

2. Representation of random coefficients

Let $a(x,\omega) : \mathscr{D} \times \Omega \to \mathbb{R}$ denote the random field used to describe and provide a mathematical model of the available experimental data. Here $\mathscr{D}$ is an open, bounded polygonal domain in $\mathbb{R}^d$ and is the spatial domain on which a is defined, Ω is the set of elementary events on which a probability space is constructed. It is well known that $a(x,\omega)$ can be efficiently represented using the spectral decomposition of its corresponding two-point correlation function [25]. Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_M$ be M real row vectors in $\mathbb{R}^N$ representing M independent observations from random vector $\boldsymbol{a}(\omega) := (a(x_1,\omega), \ldots, a(x_N,\omega))$, and let $\boldsymbol{C}$ be the underlying covariance matrix of $\boldsymbol{a}$. Moreover, let $\boldsymbol{\Xi}$ denote the set of observation points $x_1, \ldots, x_N$, i.e., $\boldsymbol{\Xi} = \{x_i \in \mathscr{D}, i = 1, \ldots, N\}$. Letting $\bar{\boldsymbol{a}} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{a}_i$, the unbiased estimate of $\boldsymbol{C}$, namely $\hat{\boldsymbol{C}}$, can be obtained as

$$\hat{\boldsymbol{C}} = \frac{1}{M-1} \sum_{i=1}^{M} (\boldsymbol{a}_i - \bar{\boldsymbol{a}})^{\mathrm{T}} (\boldsymbol{a}_i - \bar{\boldsymbol{a}}). \tag{1}$$

The Karhunen–Loeve representation of random vector $\boldsymbol{a}$ is thus given as,

$$\boldsymbol{a}(\omega) = \bar{\boldsymbol{a}} + \sum_{i=1}^{N} \sqrt{\lambda_i}\, \eta^{(i)}(\omega) \boldsymbol{\phi}_i, \tag{2}$$

where $\{\lambda_i\}_{i=1}^{N}$ and $\{\boldsymbol{\phi}_i\}_{i=1}^{N}$ are the eigenvalues and eigenvectors of $\hat{\boldsymbol{C}}$, respectively. Moreover, $\eta^{(i)}$, $i = 1, \ldots, N$, are random variables whose realizations are obtained by

$$\eta_j^{(i)} = \frac{1}{\sqrt{\lambda_i}} \langle \boldsymbol{a}_j - \bar{\boldsymbol{a}}, \boldsymbol{\phi}_i \rangle_{l_2}, \quad j = 1, \ldots, M, \tag{3}$$

where $\langle .,. \rangle_{l_2}$ denotes the scalar product in $\mathbb{R}^N$. A reduced-order representation of random vector $\boldsymbol{a}(\omega)$ can be obtained in the form

$$\boldsymbol{a}(\omega) \approx \bar{\boldsymbol{a}} + \sum_{i=1}^{\mu} \sqrt{\lambda_i}\, \eta^{(i)}(\omega) \boldsymbol{\phi}_i \tag{4}$$

for some $\mu \leqslant N$ such that $\sum_{i=1}^{\mu} \lambda_i / \sum_{i=1}^{N} \lambda_i$ is sufficiently close to one. Based on Eq. (4), vector random variable $\boldsymbol{\eta} := (\eta^{(1)} \ldots \eta^{(\mu)})$ has the following two properties

$$E[\eta^{(n)}] = 0, \quad E[\eta^{(m)} \eta^{(n)}] = \delta_{mn}, \quad n, m = 1, \ldots, \mu. \tag{5}$$

Accordingly, the covariance matrix of $\boldsymbol{\eta}$ is the $\mu \times \mu$ identity matrix, $\boldsymbol{I}_{\mu\times\mu}$. Since, the underlying random field $a(x,\omega)$ is in general non-Gaussian, the random variables $\eta^{(i)}$ are generally non-Gaussian. In the sequel, the goal is to approximate $\boldsymbol{\eta}$ by a non-Gaussian random μ-vector $\hat{\boldsymbol{\eta}}$ such that $\hat{\boldsymbol{\eta}}$ and $\boldsymbol{\eta}$ have identical covariance matrices, namely $\boldsymbol{I}_{\mu\times\mu}$, and close marginal distributions. This is consistent with the observation that in many situations of practical interest not enough data is available to approximate the joint density of random vector $\boldsymbol{\eta}$. This assumption is furthermore consistent with the assumption of independence between components of

random vector $\hat{\boldsymbol{\eta}}$. Generally, this is clearly not the case for arbitrary stochastic processes. While relaxing this assumption presents no theoretical difficulties, it does, however, present a number of computational and experimental challenges [8]. It will moreover be assumed that the mean and the dominant eigenspace of the covariance matrix are invariant under a sequence of rank one perturbations associated with the assimilation of additional measurement. This assumption permits the use of the eigenvectors $\boldsymbol{\phi}_i$ and associated eigenvalues, obtained from initial measurements, to perform the foregoing analysis. This assumption can be readily relaxed by continually recomputing the dominant subspace as additional observations are obtained. Reminding that the invariance assumption is with regards to the dominant subspace, and not a particular eigenvector, it is noted that the stability of dominant subspaces under perturbations increases with their dimension.

For each $i = 1, \ldots, \mu$, let $F_{\eta^{(i)}}$ denotes the marginal cumulative distribution function of $\eta^{(i)}$, then $\hat{\eta}^{(i)} \sim F^{-1}_{\eta^{(i)}}(\Phi(\xi_i))$ with $\Phi(\xi_i) = \int_{-\infty}^{\xi_i} \frac{1}{\sqrt{2\pi}} e^{(-\frac{z^2}{2})}\, dz$ and $\xi_1, \ldots, \xi_\mu$ are independent standard normal random variables. Moreover, $\hat{\eta}^{(i)}$ can be represented in an orthogonal expansion in terms of one-dimensional normalized Hermite polynomials in ξ_i [33,31] resulting in,

$$\hat{\eta}^{(i)} = \sum_{j=1}^{\infty} \gamma_j^{(i)} \bar{\psi}_j(\xi_i), \tag{6}$$

$$\approx \sum_{j=1}^{p} \gamma_j^{(i)} \bar{\psi}_j(\xi_i), \quad i = 1, \ldots, \mu, \tag{7}$$

where

$$\bar{\psi}_j(\xi_i) := \psi_j(\xi_i) \Big/ \left(E\left[\psi_j^2(\xi_i)\right] \right)^{1/2}, \tag{8}$$

and

$$\begin{aligned} &\psi_0(\xi_i) = 1, \\ &\psi_1(\xi_i) = \xi_i, \\ &\psi_{j+1}(\xi_i) = \xi_i \psi_j(\xi_i) - j\psi_{j-1}(\xi_i) \end{aligned} \tag{9}$$

are the Hermite polynomials in ξ_i, and finally

$$E\left[\psi_j^2(\xi_i)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \psi_j^2(\xi_i) e^{\left(-\frac{\xi_i^2}{2}\right)}\, d\xi_i. \tag{10}$$

An appropriate value for p depends on the degree to which $\eta^{(i)}$ deviates from a Gaussian variable; however, in practice, a small p, e.g., $p = 6$, is sufficient for the relatively accurate representation of a large class of random variables. For cases where a large value of p is required to accurately represent each $\hat{\eta}^{(i)}$, e.g. multi-modal random variables [43], the above representation and accordingly the subsequent procedures may not be efficient in their current format. Moreover, $\gamma_j^{(i)}$, the coefficients of expansions in (7), are required to satisfy the following constraint associated with (5),

$$\sum_{j=1}^{p} (\gamma_j^{(i)})^2 = 1, \quad i = 1, \ldots, \mu. \tag{11}$$

As indicated previously, the present construction yields the components of random vector $\hat{\boldsymbol{\eta}} = (\hat{\eta}^{(1)} \ldots \hat{\eta}^{(\mu)})$ that are independent, in contrast with those of $\boldsymbol{\eta}$ which are, in general, only uncorrelated. Each $\hat{\eta}^{(i)}$ is distributed approximately with the marginal distribution of $\eta^{(i)}$, the corresponding component of $\boldsymbol{\eta}$. Moreover, the covariance matrix of the random vector $\hat{\boldsymbol{\eta}}$ is also $\boldsymbol{I}_{\mu\times\mu}$. Clearly, for the particular cases where only limited number of samples of $\boldsymbol{\eta}$ is available, the marginal distributions, $F_{\eta^{(i)}}$, cannot be approximated properly; therefore, the true coefficients, $\gamma_j^{(i)}$, cannot be obtained by already available techniques [31,33]. The next section describes a novel approach, that relies on Bayesian parameter estimation, to characterize these coefficients. Clearly, having a statistical estimate of $\gamma_j^{(i)}$, the stochastic function $a(x, \omega)$ can be reconstructed by simply replacing $\boldsymbol{\eta}$ with $\hat{\boldsymbol{\eta}}$ in the representation of a. More precisely, the reconstructed random field $\hat{a}(x, \omega)$ on the set $\boldsymbol{\Xi}$ is obtained by

$$\hat{\boldsymbol{a}}(\omega) = \bar{\boldsymbol{a}} + \sum_{i=1}^{\mu} \sqrt{\lambda_i}\hat{\eta}^{(i)}(\omega)\boldsymbol{\phi}_i. \tag{12}$$

Depending on the desirable degree of smoothness of a on $\mathscr{D}$, one might replace $\boldsymbol{\phi}_i$ and $\bar{\boldsymbol{a}}$ with their proper interpolants in order to have representation of a on any point inside $\mathscr{D}$.

It should be indicated at this point that other stochastic parameterizations of the variables $\hat{\eta}^{(i)}$ are possible. In particular, alternate representations [3,23,24,41] with respect to other measures (besides the Gaussian) are also possible. The rate of convergence of the associated expansions will clearly depend on the particular representation used. While, in this paper, the Gaussian chaos is exclusively used, the proposed method is quite general, and can be applied to general chaos expansions.

3. Bayesian inference with Markov chain Monte Carlo

Bayesian inference provides a robust procedure for estimating the unknown coefficients $\gamma_j^{(i)}$ in the above parametric non-Gaussian expansion, Eq. (7), from observations $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_M$. For the sake of simplicity, only one component of random vector $\hat{\boldsymbol{\eta}}$, namely $\hat{\eta}^{(k)}$, will be considered. Clearly all the following procedures apply to other components in an identical manner, since by construction, components of the vector $\hat{\boldsymbol{\eta}}$ are independent. Let $\boldsymbol{\gamma}_0 := (\gamma_1^{(k)}, \ldots, \gamma_p^{(k)}) \in \boldsymbol{\Lambda}$ be the true value of coefficients in the polynomial chaos expansion of $\eta^{(k)}$ and $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1^{(k)}, \ldots, \hat{\gamma}_p^{(k)}) \in \boldsymbol{\Lambda}$ its corresponding estimate. Since $\boldsymbol{\gamma}_0$ has the property that $\sum_{j=1}^{p} (\gamma_j^{(k)})^2 = 1$, it is an element of the so-called *Stiefel* manifold $\mathscr{O}(p,1)$. Therefore, the parameter set $\boldsymbol{\Lambda}$ is taken to be $\mathscr{O}(p,1)$. Consider a cost function $L : \boldsymbol{\Lambda} \times \boldsymbol{\Lambda} \to \mathbb{R}$ in the form,

$$L[\boldsymbol{\gamma}_0, \hat{\boldsymbol{\gamma}}] = \begin{cases} 1 & \text{if } \max\limits_{1 \leqslant i \leqslant p} |\gamma_i^{(k)} - \hat{\gamma}_i^{(k)}| > \Delta, \\ 0 & \text{if } \max\limits_{1 \leqslant i \leqslant p} |\gamma_i^{(k)} - \hat{\gamma}_i^{(k)}| \leqslant \Delta \end{cases} \tag{13}$$

for some small $\Delta > 0$. Then the Bayes estimate of $\boldsymbol{\gamma}_0$ is $\hat{\boldsymbol{\gamma}}$ for which the Bayes risk is minimum, i.e.,

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\zeta}} E\{L[\boldsymbol{\gamma}_0, \boldsymbol{\zeta}(\hat{\eta}_1^{(k)}, \ldots, \hat{\eta}_M^{(k)})]\}, \tag{14}$$

$$\approx \arg\min_{\boldsymbol{\zeta}} E\{L[\boldsymbol{\gamma}_0, \boldsymbol{\zeta}(\eta_1^{(k)}, \ldots, \eta_M^{(k)})]\}, \tag{15}$$

with $\boldsymbol{\zeta} \in \boldsymbol{\Lambda}$. Here, $\eta_1^{(k)}, \ldots, \eta_M^{(k)}$ are M observations of $\eta^{(k)}$. Since $\hat{\eta}^{(k)}$ is constructed to have approximately the same distribution as that of $\eta^{(k)}$, these observations are also assumed to correspond to $\hat{\eta}^{(k)}$. For the above situation, the Bayes estimate can be shown to be the conditional mode of $\boldsymbol{\gamma}$ given $\eta_1^{(k)}, \ldots, \eta_M^{(k)}$, and is called the *maximum a posteriori probability* estimate (MAP) of $\boldsymbol{\gamma}$ [13,30]. Since an analytical expression for the distribution of $\boldsymbol{\gamma}$ given $\eta_1^{(k)}, \ldots, \eta_M^{(k)}$ is usually not available a Markov chain Monte Carlo (MCMC) algorithm [12,11] is proposed to draw samples form the posterior distribution of $\boldsymbol{\gamma}$, given observations $\eta_1^{(k)}, \ldots, \eta_M^{(k)}$, thus avoiding complex numerical integrations in high dimensions to form the posterior distribution, as required if classical Bayesian inference were adopted. In particular, the Bayesian Inference with Metropolis-Hastings algorithm (BIMH) is more suitable for the particular case addressed in this work as compared to a Gibbs sampler (the most commonly used alternative to Metropolis-Hastings) version of MCMC [37]. The reason is partially due to the fact that, while being required in the Gibbs sampling method, the distribution of any component of $\boldsymbol{\gamma}$ conditioned on the other components is not available. The main idea behind an MCMC is to construct a Markov chain with a particular transition probability so that its stationary behavior is in accordance with a target distribution. In the above, and in accordance with the Bayesian approach, $\boldsymbol{\gamma}$ also refers to realizations of the parameters being estimated.

3.1. BIMH Setup

Based on the Bayes rule one can write,

$$\pi[\boldsymbol{\gamma}|\eta_1^{(k)}, \ldots, \eta_M^{(k)}] \propto p_{\hat{\eta}^{(k)}}(\eta_1^{(k)}, \ldots, \eta_M^{(k)}|\boldsymbol{\gamma}) \times \pi(\boldsymbol{\gamma}), \tag{16}$$

where $\pi(\boldsymbol{\gamma})$ is a prior density for $\boldsymbol{\gamma}$, which in the absence of prior information, is taken to be the uniform distribution on $\boldsymbol{\Lambda}$ [9]. Also, $\pi[\boldsymbol{\gamma}|\eta_1^{(k)},\ldots,\eta_M^{(k)}]$ is the conditional or posterior density of $\boldsymbol{\gamma}$ given the observations $\eta_1^{(k)},\ldots,\eta_M^{(k)}$, and $p_{\hat{\eta}^{(k)}}(\eta_1^{(k)},\ldots,\eta_M^{(k)}|\boldsymbol{\gamma})$ is the likelihood of data or joint density of the observations, given $\boldsymbol{\gamma}$, which is equal to $\prod_{j=1}^{M} p_{\hat{\eta}^{(k)}}(\eta_j^{(k)}|\boldsymbol{\gamma})$ for the case of independent observations $\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_M$. Let $q(\boldsymbol{\theta}|\boldsymbol{\gamma}_i)$ be the *proposal distribution* to generate a candidate state $\boldsymbol{\theta}$ given that the underlying Markov chain is currently at state i, a state defined by $\boldsymbol{\gamma}_i$, and let it be equal to $\pi(\boldsymbol{\theta})$ as a possible proper choice. Therefore, the probability of moving to the candidate state $\boldsymbol{\theta}$ simplifies to

$$\rho(\boldsymbol{\theta},\boldsymbol{\gamma}_i) = \min\left\{\frac{\pi[\boldsymbol{\theta}|\eta_1^{(k)},\ldots,\eta_M^{(k)}]q(\boldsymbol{\gamma}_i|\boldsymbol{\theta})}{\pi[\boldsymbol{\gamma}_i|\eta_1^{(k)},\ldots,\eta_M^{(k)}]q(\boldsymbol{\theta}|\boldsymbol{\gamma}_i)},1\right\}, \tag{17}$$

$$= \min\left\{\frac{\Pi_{j=1}^{M} p_{\hat{\eta}^{(k)}}(\eta_j^{(k)}|\boldsymbol{\theta})}{\Pi_{j=1}^{M} p_{\hat{\eta}^{(k)}}(\eta_j^{(k)}|\boldsymbol{\gamma}_i)},1\right\}. \tag{18}$$

Remark 1. Having the above setting for the proposal distribution, the underlying Markov chain is known as an *independent Metropolis chain*. Since the proposal distribution $q(\cdot)$ is, almost everywhere, strictly positive on the corresponding Stiefel manifold, the associated kernel is *irreducible* and *aperiodic*, [29,39]. Thus one can show that the underlying Markov chain will be stationary, and therefore the proposed algorithm samples from the posterior distribution $\pi[\boldsymbol{\gamma}|\eta_1^{(k)},\ldots,\eta_M^{(k)}]$, provided that a long enough Markov chain is constructed. The following algorithm is proposed for sampling from the posterior distribution of γ, $\pi[\boldsymbol{\gamma}|\eta_1^{(k)},\ldots,\eta_M^{(k)}]$.

3.2. Algorithm of BIMH

(1) Choose the length of the burn-in period $t_b \in \mathbb{N}$ and an initial state $\boldsymbol{\gamma}_1$. Set $j = 1$.
(2) Estimate $\Pi_{j=1}^{M} p_{\hat{\eta}^{(k)}}(\eta_j^{(k)}|\boldsymbol{\gamma}_1)$.
(3) Generate candidate state $\boldsymbol{\theta}$ according to $\pi(\cdot)$.
(4) Estimate $\Pi_{j=1}^{M} p_{\hat{\eta}^{(k)}}(\eta_j^{(k)}|\boldsymbol{\theta})$.
(5) Generate U from $U(0,1)$ distribution. Set $\boldsymbol{\gamma}_{j+1} = \boldsymbol{\theta}$ if $U \leqslant \rho(\boldsymbol{\theta},\boldsymbol{\gamma}_j)$. Otherwise, set $\boldsymbol{\gamma}_{j+1} = \boldsymbol{\gamma}_j$.
(6) Repeat steps 2–5 sufficiently large, say $s \in \mathbb{N}$, number of times with $s > t_b$.

Notice that the burn-in period, t_b, ensures the dissipation of the initial condition effects. Having $s - t_b$ realizations $\boldsymbol{\gamma}_j$ from posterior distribution of $\boldsymbol{\gamma}$ as output of the above algorithm, one can easily obtain an estimate for the mode of distribution of $\boldsymbol{\gamma}_j$'s as an approximation of true $\boldsymbol{\gamma}_0$.

Having enough samples from the posterior distribution of $\boldsymbol{\gamma}$ obtained from the stationary part of the Markov chain, one can claim the convergence of the above estimate. The following theorem supports this claim for the case of continuous parameter space [13]:

Theorem 1. *If $\hat{\boldsymbol{\gamma}}$ is defined on a compact set and A is a neighborhood of $\boldsymbol{\gamma}_0$, the true vector of parameters, with non-zero prior probability, then $P(\hat{\boldsymbol{\gamma}} \in A|\eta_1^{(k)},\ldots,\eta_M^{(k)}) \to 1$ as $M \to \infty$.*

3.3. Estimation of likelihood function

Clearly, in the above procedure, the likelihood function, $\Pi_{j=1}^{M} p_{\hat{\eta}^{(k)}}(\eta_j^{(k)}|\boldsymbol{\theta})$, needs to be estimated. A naive estimate would be to sample the chaos basis and evaluate the likelihood at each $\eta_j^{(k)}$. Alternatively, a more efficient algorithm would be a *kernel density estimation*[14]. More specifically, let $K : \mathbb{R} \to \mathbb{R}$ be a kernel function, e.g. $K(u) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{u^2}{2})$ known as a Gaussian kernel. Let $\breve{\eta}_1^{(k)},\breve{\eta}_2^{(k)},\ldots,\breve{\eta}_n^{(k)}$ be n independent samples from $p_{\eta^{(k)}}$ conditioned on $\boldsymbol{\theta}$ obtained by sampling the chaos basis associated with $\hat{\eta}^{(k)}$. Then the likelihood function for each $\eta_j^{(k)}$, $j = 1,\ldots,M$, can be approximated as

$$p_{\hat{\eta}^{(k)}}(\eta_j^{(k)}|\boldsymbol{\theta}) \approx \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{\eta_j^{(k)} - \breve{\eta}_i^{(k)}}{h}\right). \tag{19}$$

Clearly the accuracy of the above approximation depends on the choice of bandwidth h. For the above case, based on Scott's *rule-of-thumb* [34], a suitable value for h is $h \approx 1.06n^{-0.2}$.

4. Uncertainty quantification for Bayesian estimates

As mentioned previously, the Bayes estimate, $\hat{\boldsymbol{\gamma}}$, is a function of $\eta_1^{(k)}, \ldots, \eta_M^{(k)}$, and thus, for a finite M, varies with the observation set. Accordingly, it is desirable to investigate the performance of the estimator in the presence of such dependence. This can be achieved by estimating the standard error or confidence intervals in the parameter estimates using resampling techniques (e.g. Bootstrap or Jackknife) when the estimate is computationally inexpensive [22,38]. Alternatively, large sample asymptotic behavior of the estimator can be relied upon, when it exists, to approximate its behavior. In the present work, for moderate values of p, i.e. the dimension of unknown parameter vector $\boldsymbol{\gamma}_0$, the estimate $\hat{\boldsymbol{\gamma}}$ could be computationally expensive; therefore, the latter approach is followed in order to measure the variability of the chaos coefficient estimates of (7).

Under some regularity conditions, the fundamental result from the asymptotic, large-sample, Bayesian inference is that as more data is assimilated, the posterior distribution of the parameter vector approaches a multivariate normal distribution [13], as explained in the following theorem:

Theorem 2 (cf. [13,30]). *Under some regularity conditions* (*notably that the likelihood is a continuous function of* $\boldsymbol{\gamma}$ *and the true vector of parameters* $\boldsymbol{\gamma}_0$ *not be on the boundary of the set to which* $\boldsymbol{\gamma}$ *belongs*), *as* $M \to \infty$, *the posteriori distribution of* $\hat{\boldsymbol{\gamma}}$ *approaches normality with mean* $\boldsymbol{\gamma}_0$ *and covariance matrix* $(MJ(\boldsymbol{\gamma}_0))^{-1}$, *where* J *is the Fisher information matrix. In other words*,

$$\sqrt{M}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{\text{dist}} N(\mathbf{0}, J(\boldsymbol{\gamma}_0)^{-1}). \tag{20}$$

In practice, for M reasonably large, $\hat{\boldsymbol{\gamma}}$ is approximately $N(\boldsymbol{\gamma},(MJ(\boldsymbol{\gamma}))^{-1})$ distributed when $\boldsymbol{\gamma}$ is chosen close to the unknown $\boldsymbol{\gamma}_0$. Following Theorem 1, $\boldsymbol{\gamma}$ is usually approximated by $\hat{\boldsymbol{\gamma}}$ for the mean and also for evaluation of the Fisher information matrix $J(\boldsymbol{\gamma})$ [37,13]. An important corollary of the above theorem is that the posterior mode, $\hat{\boldsymbol{\gamma}}$, is *consistent* estimate for $\boldsymbol{\gamma}_0$; that is, $\hat{\boldsymbol{\gamma}}$ converges in probability to $\boldsymbol{\gamma}_0$ as $M \to \infty$.

In order to fully characterize the probabilistic behavior of the chaos coefficients, $\boldsymbol{\gamma}$, procedures for evaluating the Fisher information matrix, $J(\hat{\boldsymbol{\gamma}})$, must be developed. The next subsection addresses this issue.

4.1. Estimation of the Fisher information

Under suitable regularity conditions, the Fisher information matrix associated with the vector $\boldsymbol{\gamma}$ under the likelihood $p_{\hat{\eta}^{(k)}}(\hat{\eta}^{(k)}|\boldsymbol{\gamma})$ is defined as

$$J(\boldsymbol{\gamma}) = -E\left[\frac{\partial^2 l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}\, \partial \boldsymbol{\gamma}^{\mathrm{T}}}\right] = E\left[\frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}^{\mathrm{T}}\right], \tag{21}$$

where $l(\boldsymbol{\gamma})$ is the log-likelihood of data. In situations, including the present work, where the closed form of the likelihood is not available, the Fisher information cannot be computed analytically and thus has to be numerically approximated. For i.i.d. cases this approximation can be achieved using the so-called *empirical Fisher information* $\hat{\mathbf{J}}(\boldsymbol{\gamma})$ defined by [27,28,35]

$$\hat{J}(\boldsymbol{\gamma}) = \sum_{i=1}^{M} \boldsymbol{s}_i(\boldsymbol{\gamma})\boldsymbol{s}_i(\boldsymbol{\gamma})^{\mathrm{T}} - \frac{1}{M}\left(\sum_{i=1}^{M} \boldsymbol{s}_i(\boldsymbol{\gamma})\right)\left(\sum_{i=1}^{M} \boldsymbol{s}_i(\boldsymbol{\gamma})\right)^{\mathrm{T}}, \tag{22}$$

where $\boldsymbol{s}_i(\boldsymbol{\gamma}) = \frac{\partial l_i(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\big|_{\hat{\eta}^{(k)}=\hat{\eta}_i^{(k)}}$ is the derivative of the individual log-likelihood evaluated at $\hat{\eta}_i^{(k)}$ and is called the *score* function. While being a *consistent* estimator of $J(\boldsymbol{\gamma})$, $\hat{J}(\boldsymbol{\gamma})$ is obtained by a simple finite difference scheme for the estimation of score functions. To simplify the discussion, the asymptotic behavior of the Bayes estimates of the chaos coefficients is approximated as

$$\sqrt{M}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{\text{dist}} N(\mathbf{0}, \hat{J}(\hat{\boldsymbol{\gamma}})^{-1}). \tag{23}$$

4.2. Uncertainty modelling for the system parameter

Following the above probabilistic representation of the chaos coefficients, one can reconstruct each random variable $\hat{\eta}^{(i)}$ simply by replacing $\hat{\gamma}_j^{(i)}$ with $\tilde{\gamma}_j^{(i)}$ where, for each i, $\tilde{\gamma}_j^{(i)}$ are jointly Gaussian random variables with mean $\hat{\gamma}_j^{(i)}$ and covariance matrix $(M\hat{J}(\hat{\boldsymbol{\gamma}}))^{-1}$ and are independent from $\{\xi_i\}_{i=1}^{\mu}$, i.e.,

$$\hat{\eta}^{(i)} = \sum_{j=1}^{p} \tilde{\gamma}_j^{(i)} \bar{\psi}_j(\xi_i). \tag{24}$$

Accordingly, the final representation of â on the set $\boldsymbol{\Xi}$ is obtained by substituting Eq. (24) into Eq. (12),

$$\hat{\boldsymbol{a}} = \bar{\boldsymbol{a}} + \sum_{i=1}^{\mu} \sqrt{\lambda_i} \left(\sum_{j=1}^{p} \tilde{\gamma}_j^{(i)} \bar{\psi}_j(\xi_i) \right) \boldsymbol{\phi}_i. \tag{25}$$

It is worth noting that for different indices i, the random vectors $\tilde{\boldsymbol{\gamma}}^{(i)} := [\tilde{\gamma}_1^{(i)}, \ldots, \tilde{\gamma}_p^{(i)}]$ are independent from each other, since the corresponding Markov chains are constructed independently. An interesting feature of Eq. (25) is that the ξ_i's capture the dimensions representing the intrinsic uncertainty of â, while $\tilde{\gamma}_j^{(i)}$'s reflect the uncertainty due to lack of data in the estimation of the true parameters $\boldsymbol{\gamma}_0$. As is recognized from (23), the latter uncertainty can be reduced by gathering more observations from â while the former uncertainty is irreducible. Approximation (25) can be extended to the whole domain $\mathscr{D}$. Specifically, and depending on the desirable degree of smoothness of the final representation, this can be simply done using a suitable interpolation, e.g. Lagrange interpolation, radial basis functions, B-Spline, etc., of â on the set $\boldsymbol{\Xi}$. This results in the general form,

$$\hat{a}(x,\omega) = \tilde{\bar{a}}(x) + \sum_{i=1}^{\mu} \sqrt{\lambda_i} \left(\sum_{j=1}^{p} \tilde{\gamma}_j^{(i)} \bar{\psi}_j(\xi_i) \right) \tilde{\phi}_i(x), \tag{26}$$

where $\tilde{\bar{a}}(x)$ and $\tilde{\phi}_i(x)$ are the corresponding interpolants of $\bar{\boldsymbol{a}}$ and $\boldsymbol{\phi}_i$, respectively.

The next section addresses the propagation of the uncertainties in the estimation of coefficients $\boldsymbol{\gamma}_0$'s to the response of a system governed by an SPDE with stochastic input and deterministic operator. While the extension to SPDE with stochastic parameters can be readily developed it risks to obscure the essential contribution of the present work, and will therefore not be presented here.

5. Uncertainty propagation

Consider the stochastic linear elliptic boundary value problem, with deterministic operator and stochastic input, which consists of finding a stochastic function $u(x,\omega) : \bar{\mathscr{D}} \times \Omega \to \mathbb{R}$, such that the following equation holds almost surely in Ω,

$$\begin{aligned} -\nabla.(k(x)\nabla u(x,\omega)) &= a(x,\omega), \quad x \in \mathscr{D}, \\ u(x,\omega) &= 0, \quad x \in \partial\mathscr{D}, \end{aligned} \tag{27}$$

where

$$0 < k_{\min} \leqslant k(x) \leqslant k_{\max} < \infty \tag{28}$$

$a : \mathscr{D} \times \Omega \to \mathbb{R}$ is a stochastic function with continuous and square-integrable covariance function, Ω is the set of elementary events, and $\omega \in \Omega$. For the case of physical systems discussed in previous sections, one has $a(x,\omega) := \hat{a}(x,\omega) = a(x,\xi_1(\omega), \ldots, \xi_q(\omega))$ where $\{\xi_i(\omega)\}_{i=1}^{q}$ is the set of real mutually orthonormal Gaussian random variables with mean zero. Here q is the number of effective coordinates in the stochastic dimension of the problem.

For the sake of completeness the fundamental concepts of the *Stochastic Projection* approach are briefly reviewed next [20].

5.1. Spectral stochastic finite element method (SSFEM)

5.1.1. Variational formulation

The analysis of the numerical approximation of stochastic functions can be greatly facilitated by using a tensor product framework. Consider two domains $x \in \mathscr{D}$, $\omega \in \Omega$, let $H_0^1(\mathscr{D})$ be the subspace of $H^1(\mathscr{D})$ consisting of functions which vanish on $\partial\mathscr{D}$ and are equipped with the norm $\|v\|_{H_0^1(\mathscr{D})} = \{\int_{\mathscr{D}} |\nabla v|^2 \, dx\}^{\frac{1}{2}}$, also let $L_2(\Omega)$ be the space of random variables with finite variance defined on Ω, then the tensor space $H_0^1(\mathscr{D}) \otimes L_2(\Omega)$ is the completion of the formal sums $u(x,\omega) = \sum_{i,j=1,\ldots,n} v_i(x) w_j(\omega)$, $\{v_i(x)\} \subset H_0^1(\mathscr{D})$, $\{w_j(\omega)\} \subset L_2(\Omega)$, with respect to the inner product

$$(u, \tilde{u})_{H_0^1(\mathscr{D}) \otimes L_2(\Omega)} = \sum_{i,j} (v_i, \tilde{v}_i)_{H_0^1(\mathscr{D})} (w_j, \tilde{w}_j)_{L_2(\Omega)}. \tag{29}$$

Consider the tensor product Hilbert space $H = H_0^1(\mathscr{D}) \otimes L_2(\Omega)$ equipped with inner product $(v,u)_H = E[\int_{\mathscr{D}} \nabla v \cdot \nabla u \, dx]$. Construct the bilinear form $\mathscr{B} : H \times H \to \mathbb{R}$ by

$$\mathscr{B}(v,w) := E\left[\int_{\mathscr{D}} k \nabla v \cdot \nabla w \, dx\right] \quad \forall v, w \in H. \tag{30}$$

By the assumption of (28) on $k(x)$, the continuity and coercivity of the bilinear form $\mathscr{B}$ are guaranteed, therefore by the Lax–Milgram lemma, [6], the following variational formulation has a unique solution in H:

$$\mathscr{B}(u,v) = \mathscr{L}(v) \quad \forall v \in H, \tag{31}$$

where

$$\mathscr{L}(v) := E\left[\int_{\mathscr{D}} av \, dx\right] \quad \forall v \in H \tag{32}$$

is a bounded linear functional.

5.1.2. Polynomial chaos expansion

A second-order random variable $\chi(\omega)$ can be represented as [5]

$$\begin{aligned}\chi(\omega) = x_0 H_0 &+ \sum_{i_1=1}^{\infty} x_{i_1} H_1(\xi_{i_1}(\omega)) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} x_{i_1 i_2} H_2(\xi_{i_1}(\omega), \xi_{i_2}(\omega)) \\ &+ \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} x_{i_1 i_2 i_3} H_3(\xi_{i_1}(\omega), \xi_{i_2}(\omega), \xi_{i_3}(\omega)) + \cdots,\end{aligned} \tag{33}$$

where $H_n(\xi_{i_1}, \ldots, \xi_{i_n})$ is a Hermite polynomial of order n in variables $(\xi_{i_1}, \ldots, \xi_{i_n})$. It will be notationally more convenient to rewrite the above equation in the form

$$\chi(\omega) = \sum_{j=0}^{\infty} \hat{x}_j \psi_j(\omega), \tag{34}$$

where $\psi_j(\omega) := \psi_j(\boldsymbol{\xi}(\omega))$ and $\boldsymbol{\xi}(\omega)$ is the vector of independent Gaussian random variables $(\xi_{i_1}, \ldots, \xi_{i_n})$. Also there is a one-to-one correspondence between the functionals $\psi_j(\cdot)$ and $H(\cdot)$ and also between the associated coefficients. For computational purposes, the above series should be truncated with respect to both the dimension of vector $\boldsymbol{\xi}$ and also the order of Hermite polynomials. An important property of the above polynomials which will be employed in the following sections is their orthogonality with respect to the Gaussian probability measure, namely

$$E[\psi_i \psi_j] = E[\psi_i^2] \delta_{ij}, \tag{35}$$

where δ_{ij} is the Kronecker delta.

5.1.3. Galerkin projection

The SSFEM seeks the solution of the variational problem of (31) on the tensor product space consisting of the finite-dimensional space of piecewise continuous polynomials corresponding to a partition $\mathscr{T}_h$ of $\mathscr{D}$ as the spatial discretization and the space of the random variables spanned by polynomial chaos [20] of order up to p as the discretization of the random dimension of $u(x,\omega)$. More specifically, consider a family of finite element approximation spaces, $X := X_h \subset H_0^1(\mathscr{D})$, consisting of piecewise linear continuous functions on the corresponding triangulation of $\mathscr{D}$, and $Y := Y_p = \oplus_{i=1}^{p} \mathscr{H}_i \subset L_2(\Omega)$, where $\mathscr{H}_i$ represents the ith Homogeneous chaos, the space spanned by all ith order Hermite polynomial chaos, constructed from the set $\{\xi_i(\omega)\}_{i=1}^{q}$, [20]. Then the finite-dimensional Galerkin approximation of the exact solution $u(x,\omega)$ on the tensor product space of X and Y is obtained by solving

$$\mathscr{B}(u_{X,Y}, v) = \mathscr{L}(v) \quad \forall v \in X \otimes Y. \tag{36}$$

More clearly, one can write the approximation solution

$$u_{X,Y}(x,\omega) = \sum_{i,j} u_{ij} N_i(x) \psi_j(\omega), \tag{37}$$

and use the test function $v(x,\omega) = N_k(x)\psi_l(\omega)$ to find the coefficients u_{ij}. Then (36) gives the following system of linear equations for u_{ij}

$$\sum_{i,j} (E[\psi_l \psi_j (k\nabla N_i, \nabla N_k)_{L_2(\mathscr{D})}]) u_{ij} = E[\psi_l (a(.,.), N_k)_{L_2(\mathscr{D})}] \quad \forall k, l, \tag{38}$$

where $\{N_i(x)\}_{i=1}^{N}$ and $\{\psi_i(\omega)\}_{i=0}^{P}$ are the basis for X and Y, respectively, and $a(x,\omega)$ is substituted from (26).

Thanks to the orthogonality of the polynomial chaos basis, the above system of equations, for the case of deterministic operator, simplifies to the solution of $P+1$ linear systems of equations each of size $N \times N$.

5.2. Propagation for the case of $\hat{a}$

The solution of Eq. (27) when the input is modeled by Eq. (26) is obtained by the procedure introduced in the previous subsection, which is summarized in the following proposition.

Proposition 1. *Let*

$$u(x,\omega) = \bar{u}(x) + \sum_{i=1}^{\mu} \sum_{j=1}^{p} \sum_{k=1}^{N} u_{jk}^{(i)} \bar{\psi}_j(\xi_i) N_k(x) \tag{39}$$

be the discretized solution of Eq. (27) when the input is

$$\hat{a}(x,\omega) = \tilde{\bar{a}}(x) + \sum_{i=1}^{\mu} \sqrt{\lambda_i} \left(\sum_{j=1}^{p} \hat{\gamma}_j^{(i)} \bar{\psi}_j(\xi_i) \right) \tilde{\phi}_i(x), \tag{40}$$

then the discretized solution of Eq. (27) under Eq. (26) is given by

$$u(x,\omega) = \bar{u}(x) + \sum_{i=1}^{\mu} \sum_{j=1}^{p} \sum_{k=1}^{N} (1 + \beta_j^{(i)}) u_{jk}^{(i)} \bar{\psi}_j(\xi_i) N_k(x), \tag{41}$$

where $\beta_j^{(i)} := \frac{\tilde{\gamma}_j^{(i)} - \hat{\gamma}_j^{(i)}}{\hat{\gamma}_j^{(i)}}$.

Proof. First recall that $\tilde{\boldsymbol{\gamma}}^{(i)}$, $i = 1, \ldots, \mu$, are independent and asymptotically Gaussian. The derivation of (41) is then obtained by first applying the linear transformation $(M\hat{J}(\hat{\boldsymbol{\gamma}}^{(i)}))^{-1}$, on the vector random variable, $\tilde{\boldsymbol{\gamma}}^{(i)}$, leading to $\mu \times p$ independent standard Gaussian random variables, $\{\zeta_i(\omega)\}_{i=1}^{\mu \times p}$. Following that, the polynomial chaos expansion of the response as described in Section (5.1.2) is developed in terms of the random variables $\{\zeta_i(\omega)\}_{i=1}^{\mu \times p}$ and $\{\xi_i(\omega)\}_{i=1}^{\mu}$. Finally, the orthogonality of chaos polynomials along with the linearity of the operator with respect to the terms in $\hat{a}$ can be used to infer the result. □

Remark 1. The bases $\bar{\psi}_j(\xi_i)$ are themselves among the polynomial chaos basis constructed from $\{\xi_i(\omega)\}_{i=1}^{\mu}$, when normalized.

Remark 2. There is only need to solve $\mu + 1$ system of equations with size $N \times N$ to fully characterize Eq. (39) and hence Eq. (41).

Remark 3. Following the asymptotic approximation of Eq. (23), the variability in estimates of the polynomial chaos coefficient of (7) does not affect the response mean of (27). In other words only $\hat{\gamma}^{(i)}$'s are sufficient to estimate $\bar{u}(x)$.

Remark 4. Based on Proposition 1, the variance of $u(x,\omega)$ at each finite element node $l = 1,\ldots,N$ is given by

$$E\left[(u(x_l,\omega) - \bar{u}(x_l))^2\right] = \sum_{i=1}^{\mu}\sum_{j=1}^{p}\left(u_{jl}^{(i)}\right)^2\left(1 + \left(\rho_j^{(i)}\right)^2\right), \tag{42}$$

where

$$\rho_j^{(i)} := \frac{\left[\left(M\hat{J}(\hat{\gamma}^{(i)})\right)^{-1}\right]_{jj}^{\frac{1}{2}}}{\hat{\gamma}_j^{(i)}} \tag{43}$$

is the coefficient of variation of the jth entry of the random vector $\tilde{\gamma}^{(i)}$. Having the above representation of the response variance, one readily has the contribution of variability of estimates of the polynomial chaos coefficient towards the overall variance at any point on $\mathscr{D}$.

6. Numerical example

As an example, consider the following one-dimensional problem:

$$\begin{aligned} &-\frac{\mathrm{d}^2u(x,\omega)}{\mathrm{d}x^2} = a(x,\omega), \quad x \in \mathscr{D} = (-0.5,+0.5)\\ &u(\pm 0.5,\omega) = 0, \end{aligned} \tag{44}$$

where $a(x,\omega)$ is the underlying input function for which observations are available on the set $\Xi = \{-0.5,-0.25,0,0.25,0.5\}$. In order to evaluate the performance of the proposed procedures, a random field with known probabilistic structure is assumed. Limited number of realizations of the process on a set of observation points, Ξ, inside $\mathscr{D}$ are then recorded. It is assumed that the dominant eigenspace of the associated covariance matrix of the observations as well as the mean of the observations are close enough to the corresponding exact values, thus enabling the verification of the results of the proposed algorithm. Clearly, for practical cases this assumption can be relaxed, the mean and the dominant subspace are re-computed as observations are collected. The only other inputs to the algorithm is the recorded data along with the geometry of the domain. After estimating the unknown parameters, the marginal distribution of the original process and those of the reconstructed one are compared. Furthermore, the variability of the parameter estimates are propagated to the response of (44) and the effect of such variability on the statistics of the solution can then be analyzed.

6.1. Generation of artificial data

Let $G(x,\omega)$ be a random field on $\mathscr{D}$ given by

$$G(x,\omega) := \sum_{i=1}^{5}\sqrt{\lambda_i}\varphi_i(x)v_i(\omega). \tag{45}$$

Here $\{\lambda_i\}_{i=1}^{5}$ are the eigenvalues of the positive definite matrix $\boldsymbol{C}_G$ with $[\boldsymbol{C}_G]_{ij} = 0.3\exp(-0.25|i-j|/1.8)$ and $i,j = 1,\ldots,5$. The matrix $\boldsymbol{C}_G$ defines the covariance of random variables $G(x_i,\omega)$ with $x_i \in \Xi =$

$\{-0.5, -0.25, 0, 0.25, 0.5\}$. Moreover, $\varphi_i(x)$ is the piecewise linear interpolant of components of ith eigenvector of $\boldsymbol{C}_G$, and $v_1, \ldots, v_5$ are independent identically distributed standard normal random variables. Let

$$a(x,\omega) := \mathrm{e}^{G(x,\omega)}, \quad x \in \mathscr{D} \tag{46}$$

define the underlying random field representing the right-hand side (RHS) of Eq. (44). The reliability of the estimates are expected to increase as more observations of a are made available. To investigate this behavior, four cases of $M = 20, 30, 50, 100$ are considered. For each case the observations consist of randomly generated realizations of a on the set Ξ, as shown in Fig. (1) for the case of $M = 50$.

6.2. Characterization of a

Based on the eigenvalues of $\hat{\boldsymbol{C}}$ associated with observations of a, as shown in Fig. (2), a value of $\mu = 2$ is deemed suitable for the approximation of the random field a. Hence, only two random variables, namely $\eta^{(1)}$ and $\eta^{(2)}$, are sufficient for representing the variability in a. Having realizations of the field a, one can obtain the corresponding realizations of $\eta^{(1)}$ and $\eta^{(2)}$ according to Eq. (3). A fourth-order Hermite polynomial expansion for $\eta^{(1)}$ and $\eta^{(2)}$, in different dimensions, is assumed to well represent those random variables. More precisely,

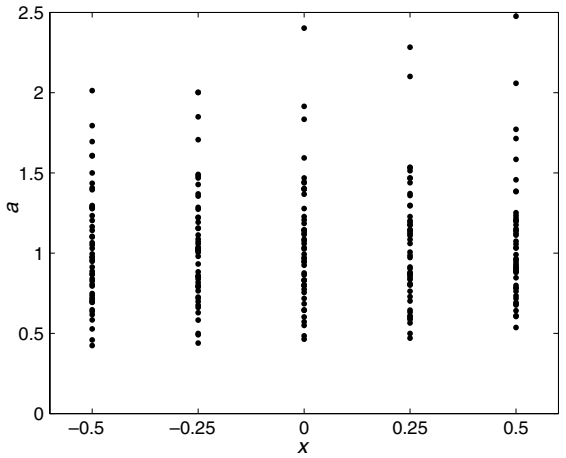

Fig. 1. Realizations of random field a on the set Ξ, $M = 50$.

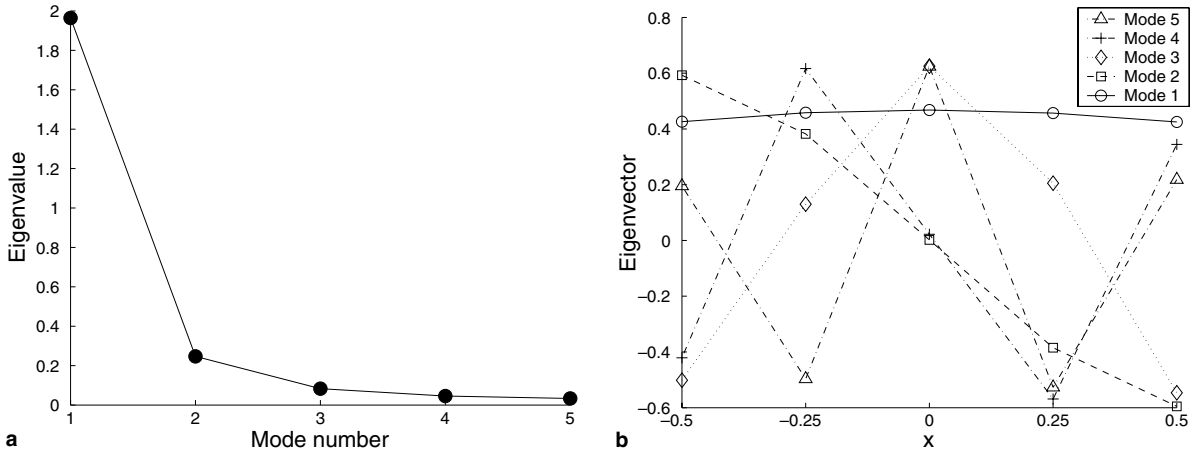

Fig. 2. Eigen pairs of the covariance matrix $\hat{C}$.

$$\hat{\eta}^{(1)} = \sum_{j=1}^{4} \gamma_j^{(1)} \bar{\psi}_j(\xi_1) \quad \text{and} \quad \hat{\eta}^{(2)} = \sum_{j=1}^{4} \gamma_j^{(2)} \bar{\psi}_j(\xi_2), \tag{47}$$

where ξ_1 and ξ_2 are independent standard normal random variables. Now the BIMH algorithm is implemented in order to estimate the unknown coefficients, $\gamma_j^{(1)}$ and $\gamma_j^{(2)}$, for $j = 1, \ldots, 4$ in the above expansions. For each expansion one Markov chain is run for 20,000 steps (i.e. $s = 20{,}000$) and the burning period, t_b, is taken to consist of 5,000. To preserve the variance of each random variable $\eta^{(1)}$ and $\eta^{(2)}$, the proposal candidates of $\gamma_j^{(1)}$, $j = 1, \ldots, 4$, and also $\gamma_j^{(2)}$, $j = 1, \ldots, 4$, are each sampled uniformly, following [9], on the Stiefel manifold $\mathcal{O}(4, 1)$. Based on the last 15,000 samples of each $\gamma_j^{(1)}$ and $\gamma_j^{(2)}$ conditioned on realizations of $\eta^{(1)}$ and $\eta^{(2)}$, the MAP estimate of $\gamma_j^{(1)}$ and $\gamma_j^{(2)}$ for $j = 1, \ldots, 4$ is obtained. As more data are used to estimate these parameters, the posterior distribution of the parameters approach that of a Gaussian random variable centered around the estimate. Fig. (3) shows the histogram of samples from the posterior distribution of $\gamma_2^{(1)}$ of the unknown parameters obtained from the MH algorithm of Section (3.2), for various values of M. Fig. (4) shows similar results for $\gamma_3^{(2)}$.

To verify the accuracy of the above estimates, the marginal density of $\eta^{(1)}$ and $\eta^{(2)}$ and their corresponding approximates for $M = 20, 30, 50, 100$ are compared in Fig. (5). Notice that the exact marginal distribution of

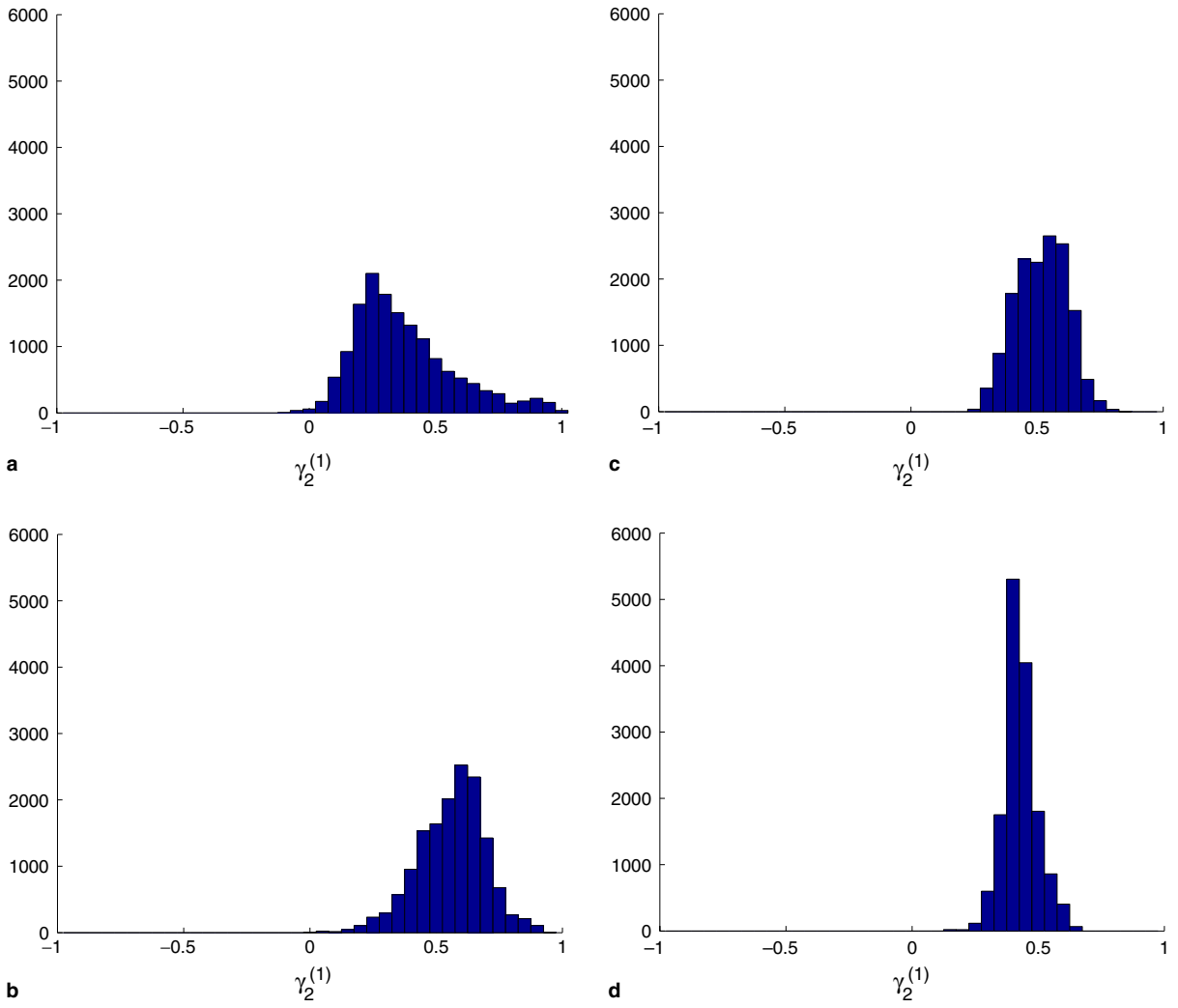

Fig. 3. Histogram of the posteriori samples of $\gamma_2^{(1)}$: (a) $M = 20$, (b) $M = 30$, (c) $M = 50$, (d) $M = 100$.

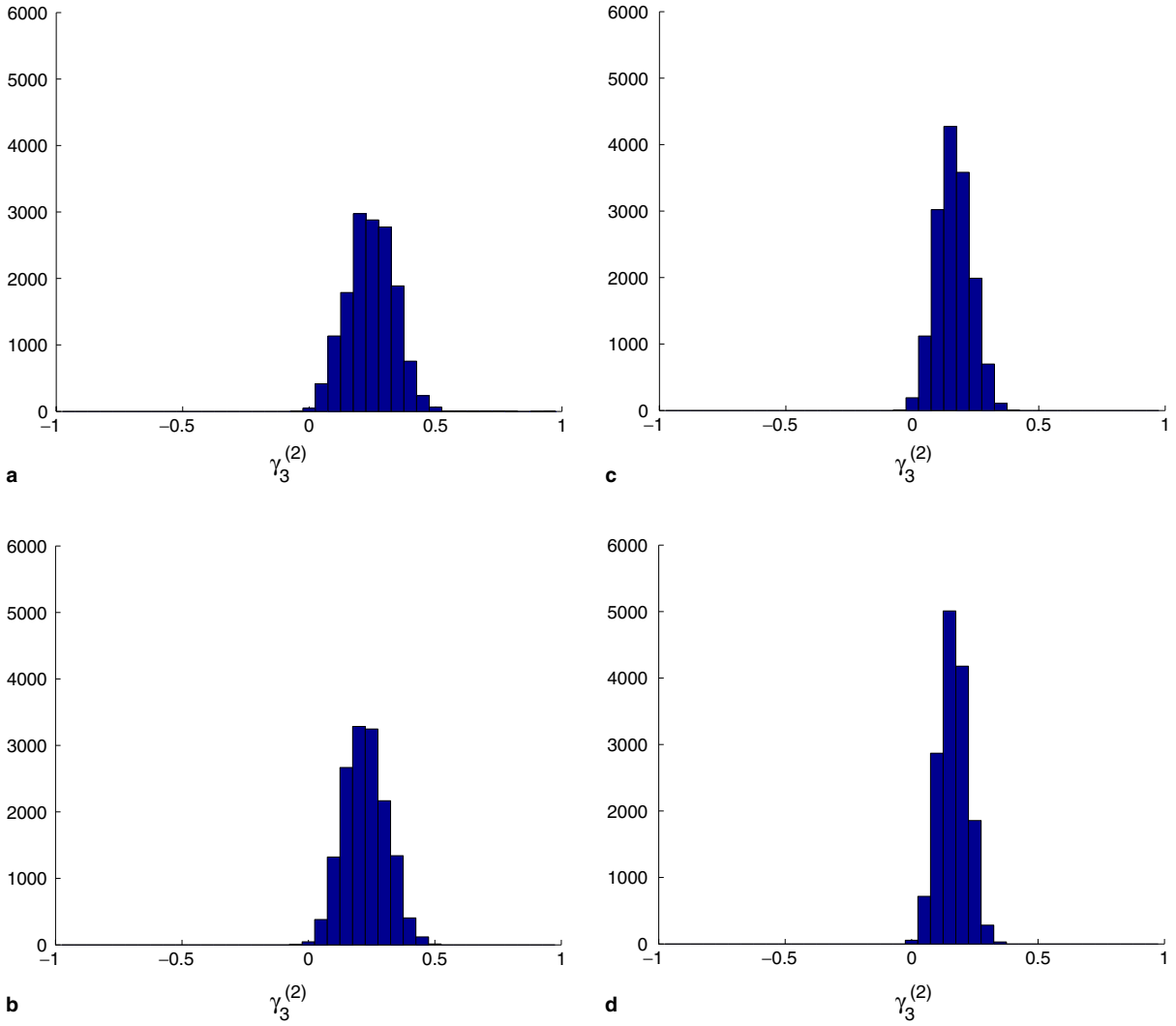

Fig. 4. Histogram of the posteriori samples of $\gamma_3^{(2)}$: (a) $M = 20$, (b) $M = 30$, (c) $M = 50$, (d) $M = 100$.

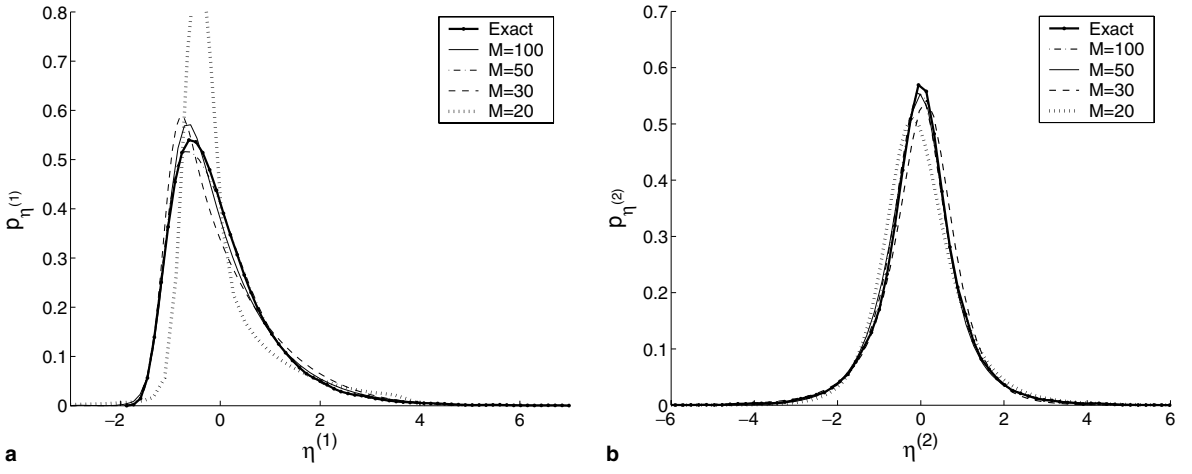

Fig. 5. Comparison of the exact and the corresponding estimates of $\eta^{(1)}$ and $\eta^{(2)}$: (a) $\eta^{(1)}$, (b) $\eta^{(2)}$.

$\eta^{(1)}$ and $\eta^{(2)}$ are easily obtained by drawing large number of samples from the original a and consequently from $\eta^{(1)}$ and $\eta^{(2)}$.

Moreover, since the underlying field a is known, the marginal distribution of $a(x_i,\omega)$ for each $x_i \in \Xi$ is exactly available and therefore can be compared with those of the reconstructed field. Based on Eq. (26), the effect of uncertainty of the estimates of the parameters are included in such comparison and shown in Fig. (6).

As is clear from Fig. (6), as more data are incorporated in the estimation of parameters, not only do the estimates represent the corresponding exact densities more accurately, but also the confidence in the estimates increases. This is consistent with the results associated with the variance of a, at any spatial point $x_i \in \Xi$, as shown in Fig. (7).

6.3. Propagation of â

The SSFEM approach described in Section (5) is implemented to propagate â through Eq. (44). Particularly, a uniform mesh of linear elements with size $h = 0.125$ is used to partition $\mathscr{D}$. A fourth-order polynomial chaos in two dimensions is used to discretize the random dimension associated with $u(x,\omega)$. To illustrate the effect of the parameter variabilities of â on the response of (44), the variance of the exact solution, when the RHS is modeled by a, is compared with the variance of â for $M = 20, 30, 50, 100$.

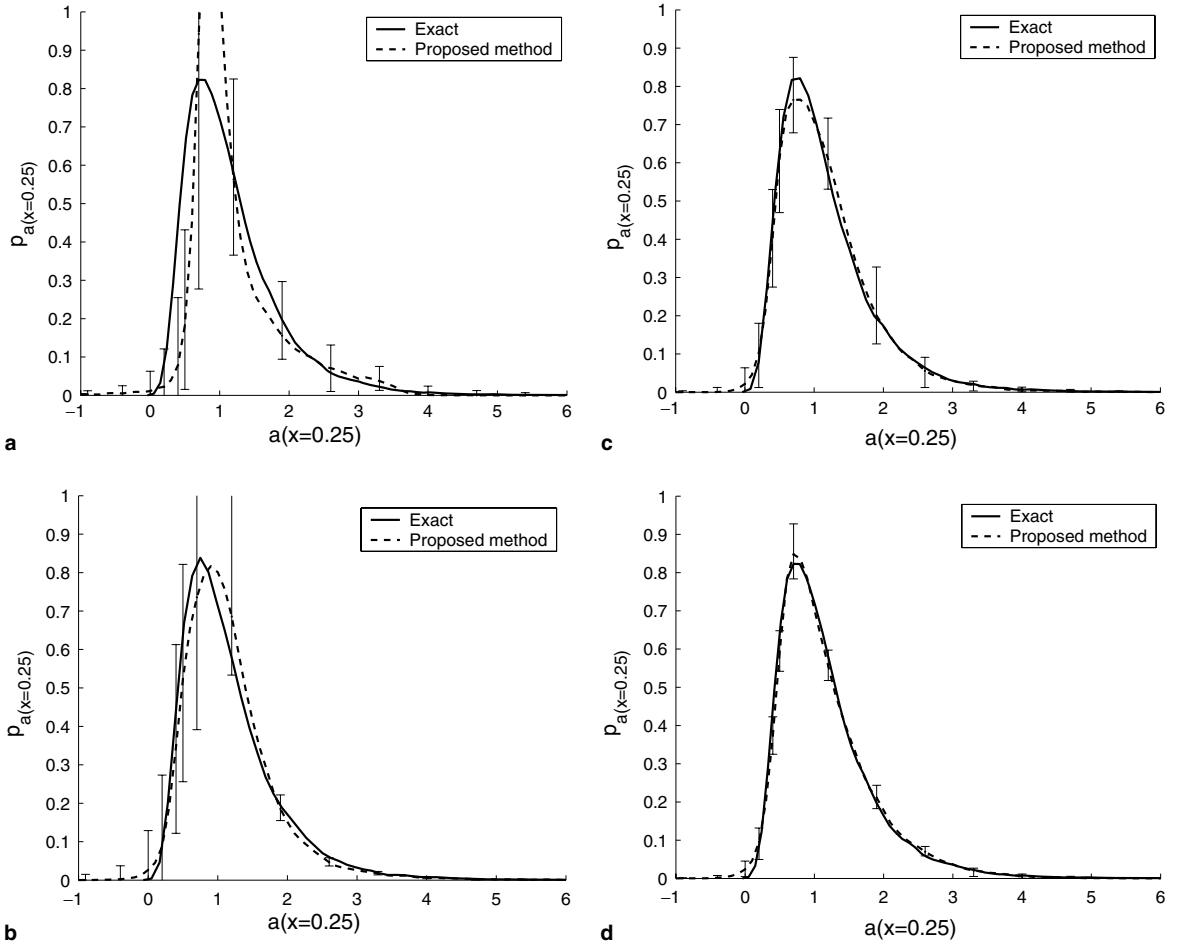

Fig. 6. Probability density function of $a(x = -0.25)$ with 95% confidence intervals around the estimated density: (a) $M = 20$, (b) $M = 30$, (c) $M = 50$, (d) $M = 100$.

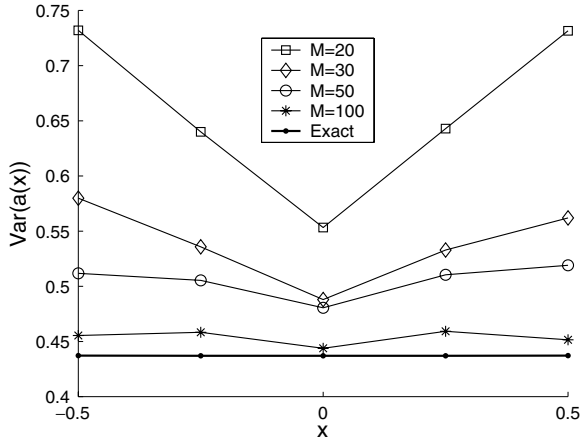

Fig. 7. Variance of $\hat{a}(x)$ considering the variability associated with estimates.

As observed from Fig. 8, as more data are used to characterize a, the variability of response associated with the estimates reduces to zero. For the particular choice of this example the largest variability corresponds to $x = 0$ (Fig. 8). This is more investigated by looking at the probability distribution of marginal of u at $x = 0$ in Fig. 9.

A careful observation of the above results highlights the significance of adaptivity with respect to the number of experimental data points used in calibrating the stochastic system parameters. More specifically, both the estimates and the associated statistical confidence are affected by the experimental sample size used in the estimation algorithm, and the accuracy of both quantities increases with more observations. This parallels the concept of error analysis in the finite element solution of deterministic partial differential equations where both the approximate solution and the associated a posteriori error estimates are obtained more accurately as the descritization spaces are enriched. It is noted that this parallelization between the deterministic and the stochastic cases is greatly facilitated by the product space constructions and associated projections adopted in conjunction with the polynomial chaos decompositions.

It is finally noted that recent efforts at modeling epistemic uncertainty [10] have constructed polynomial chaos representations that are consistent, in either a distributional or an almost sure sense, with some soft constraints such as bounds on the values of the probability density function, or the covariance function [32]. The present formulation provides a construction procedure for these representations starting directly from observations.

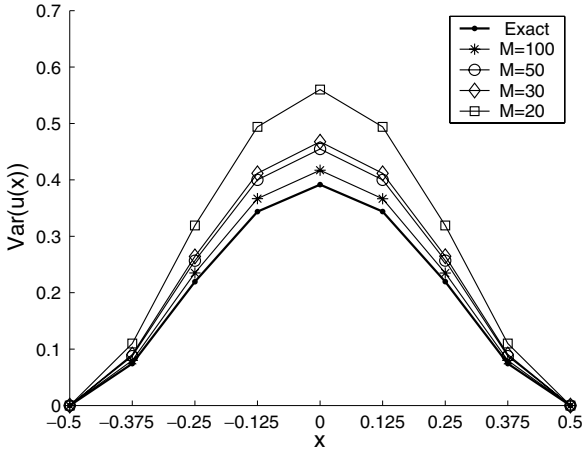

Fig. 8. Variance of $u(x)$ when the RHS is modeled with $\hat{a}$ (for $M = 20$, 30, 50 and 100) and a (exact solution).

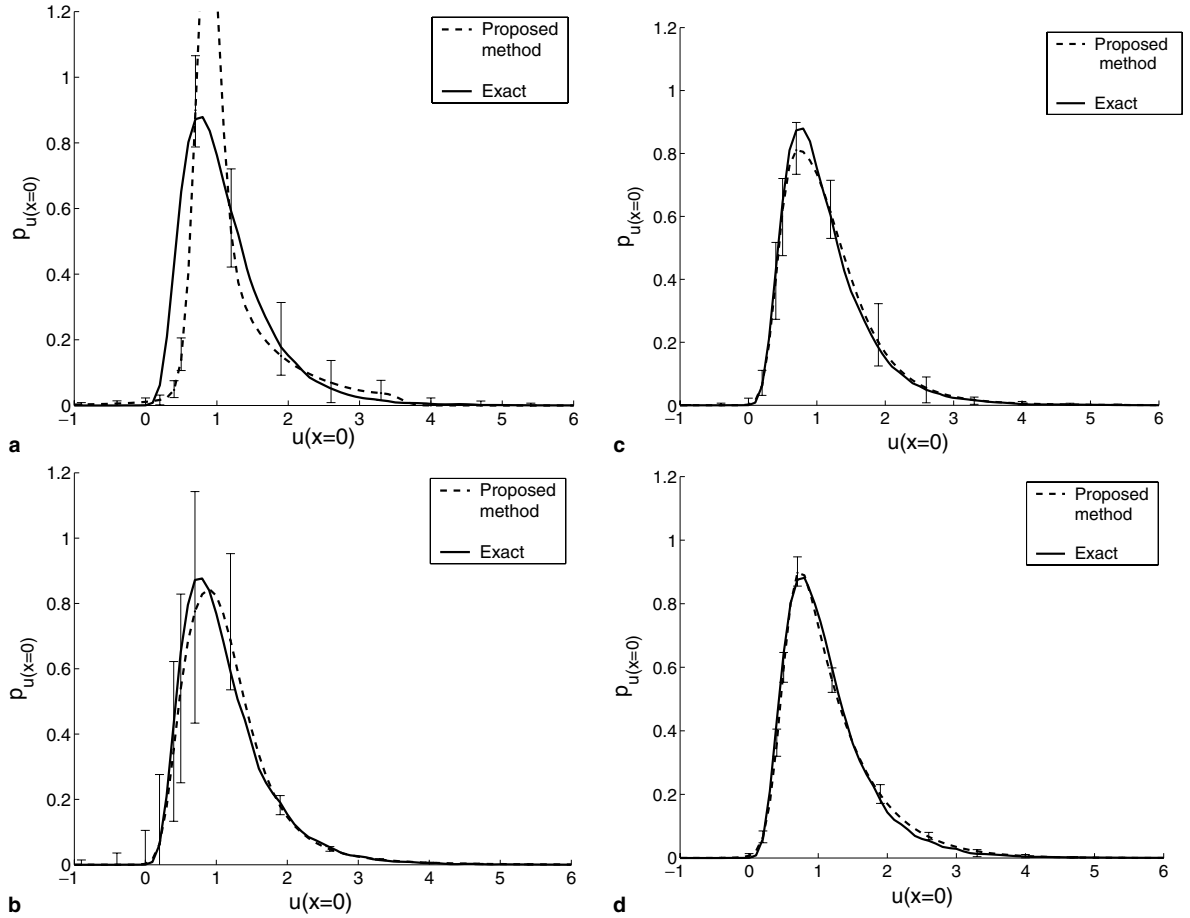

Fig. 9. Probability density function of $u(x = 0)$ with 95% confidence intervals around the estimated density: (a) $M = 20$, (b) $M = 30$, (c) $M = 50$, (d) $M = 100$.

7. Conclusion

A Bayesian parameter estimation approach has been proposed to construct a polynomial chaos model of random field for which only limited number of data observed at a limited number of locations on its domain is available. The performance of the proposed methodology has been evaluated based on a simple random field whose probabilistic structure is known. Cases corresponding to various amounts of available data are considered. For each one of these cases, the associated variability of the estimates is approximated based on the asymptotic behavior of the MAP estimator. In particular, fairly close match between the exact and estimated quantities of interest is observed. Furthermore, the effect of such variability on the response of a particular stochastic system has also been quantified through the well-developed stochastic projection scheme. Having such information helps one to develop adaptivity procedures, with respect to the input information to a system, for refinement of the model parameters. It is worth mentioning that the associated Markov chains for the various random variables, $\eta^{(i)}$, are independent from each other, and the whole procedures thus ends itself to very efficient parallelism.

Acknowledgments

This work was supported by the ONR, AFOSR, NSF and Sandia National Laboratories.

References

[1] I. Babuška, P. Chatzipantelidis, On solving elliptic stochastic partial differential equations, Comput. Meth. Appl. Mech. Eng. 191 (37–38) (2002) 4093–4122.
[2] I. Babuška, K. Liu, R. Tempone, Solving stochastic partial differential equations based on the experimental data, Math. Models Meth. Appl. Sci. 13 (3) (2003).
[3] I. Babuška, R. Tempone, G. Zouraris, Galerkin finite element approximations of stochastic elliptic partial differential equations, SIAM J. Numer. Anal. 42 (2) (2004) 800–825.
[4] F. Benth, J. Gjerde, Convergence rates for finite element approximations of stochastic partial differential equations, Stochastics Stochastic Rep. 63 (1998) 313–326.
[5] R.H. Cameron, W.T. Martin, The orthogonal development of nonlinear functionals in series of Fourier–Hermite functionals, Ann. Math. 48 (1947) 385–392.
[6] P. Ciarlet, The Finite Element Method For Elliptic Problems, North-Holland, Amsterdam, 1978.
[7] B. Debusschere, H. Najm, P. Pebay, O. Knio, R. Ghanem, O. Le Maitre, Numerical challenges in the use of polynomial chaos representations for stochastic processes, SIAM J. Sci. Comput. 26 (2) (2004) 698–719.
[8] C. Descelliers, R. Ghanem, C. Soize, Maximum likelihood estimation of stochastic chaos representation from experimental data, Int. J. Numer. Meth. Eng. (to appear).
[9] K.T. Fang, R.Z. Li, Some methods for generating both an nt-net and the uniform distribution on a Stiefel manifold and their applications, Comput. Stat. Data Anal. 24 (1997) 29–46.
[10] S. Ferson, C. Joslyn, J. Helton, W. Oberkampf, K. Sentz, Summary from the epistemic uncertainty workshop: consensus amid diversity, Reliab. Eng. Syst. Safety 85 (1–3) (2004) 355–369.
[11] D. Gamerman, Morkov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Chapman and Hall, London, 1997.
[12] A. Gelfand, A. Smith, Sampling-based approaches in calculating marginal densities, J. Am. Stat. Assoc. 85 (410) (1990) 398–409.
[13] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, second ed., Chapman and Hall, Boca Raton, FL, 2003.
[14] J.E. Gentle, Elements of Computational Statistics, Springer, New York, 2002.
[15] R. Ghanem, Probabilistic characterization of transport in heterogeneous porous media, Comput. Meth. Appl. Mech. Eng. 158 (3–4) (1998) 199–220.
[16] R. Ghanem, Ingredients for a general purpose stochastic finite elements formulation, Comput. Meth. Appl. Mech. Eng. 168 (1–4) (1999) 19–34.
[17] R. Ghanem, S. Dham, Stochastic finite element analysis for multiphase flow in heterogeneous porous media, Transp. Porous Media 32 (1998) 239–262.
[18] R. Ghanem, A. Sarkar, Mid-frequency structural dynamics with parameter uncertainty, Comput. Meth. Appl. Mech. Eng. 191 (2002) 5499–5513.
[19] R. Ghanem, A. Sarkar, Reduced models for the medium-frequency dynamics of stochastic systems, JASA 113 (2) (2003) 834–846.
[20] R. Ghanem, P. Spanos, Stochastic Finite Elements: A Spectral Approach, Dover, New York, 2002.
[21] L. Guadagnini, A. Guadagnini, D.M. Tartakovsky, Probabilistic reconstruction of geologic facies, J. Hydrol. 294 (2004) 57–67.
[22] P. Hall, The Bootstrap and Edgeworth Expanison, Springer, Berlin, 1992.
[23] O.P. Le Maitre, H. Najm, R. Ghanem, O. Knio, Multi-resolution analysis of Wiener-type uncertainty propagation schemes, J. Comput. Phys. 197 (2) (2004) 502–531.
[24] O.P. Le Maitre, O. Knio, H. Najm, R. Ghanem, Uncertainty propagation using Wiener–Haar expansions, J. Comput. Phys. 197 (2004) 28–57.
[25] M. Loeve, Probability Theory, fourth ed., Springer, New York, 1977.
[26] O. Le Maitre, Knio, M. Reagan, H. Najm, R. Ghanem, A stochastic projection method for fluid flow. i: Basic formulation, J. Comput. Phys. 173 (2001) 481–511.
[27] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, Wiley, New York, 1997.
[28] I. Meilijson, A fast improvement to the em algorithm on its own terms, J. R. Statist. Soc. B 51 (1989) 127–138.
[29] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equations of of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1109.
[30] H. Poor, Introduction to Signal Detection and Estimation, second ed., Springer, New York, 1994.
[31] B. Puig, C. Poirion, C. Soize, Non-gaussian simulation using Hermite polynomial expansion: convergences and algorithms, Probab. Eng. Mech. 17 (2002) 253–264.
[32] J.R. Red-Horse, A. Benjamin, A probabilistic approach to uncertainty quantification with limited information, Reliab. Eng. Syst. Safety 85 (1–3) (2004) 183–190.
[33] S. Sakamoto, R. Ghanem, Polynomial chaos decomposition for the simulation of non-gaussian non-stationary stochastic processes, ASCE J. Eng. Mech. 128 (2) (2002) 190–201.
[34] D.W. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization, Wiley, New York, 1992.
[35] W.A. Scott, Maximum likelihood estimation using the empirical fisher information matrix, J. Statist. Comput. Simul. 72 (8) (2002) 599–611.
[36] C. Soize, R. Ghanem, Physical systems with random uncertainties: chaos representations with arbitrary probability measure, SIAM J. Sci. Comput. 26 (2) (2005) 395–410.
[37] J. Spall, Introduction to Stochastic Search and Optimization, Wiley, New York, 2003.
[38] B. Efronand, R. Tibshirani, An Introduction to the Bootstrap, Chapman and Hall, London, 1993.

[39] L. Tierney, Markov chains for exploring posterior distributions, Ann. Stat. 22 (4) (1994) 1701–1762.
[40] N. Wiener, The homogeneous chaos, Am. J. Math. 60 (1938) 897–936.
[41] D. Xiu, G. Karniadakis, Modeling uncertainty in flow simulations via generalized polynomial chaos, J. Comput. Phys. 187 (1) (2003) 137–167.
[42] D. Xiu, G.M. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.
[43] D. Xiu, D.M. Tartakovsky, A two-scale nonperturbative approach to uncertainty analysis of diffusion in random composites, J. Multiscale Model. Simul. 2 (4) (2004) 662–674.