Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 224 (2007) 560–586
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Stochastic spectral methods for efficient Bayesian solution of inverse problems

Youssef M. Marzouk *, Habib N. Najm, Larry A. Rahn

Sandia National Laboratories, Livermore, CA 94551, USA

Received 21 February 2006; received in revised form 18 September 2006; accepted 9 October 2006
Available online 22 November 2006

Abstract

We present a reformulation of the Bayesian approach to inverse problems, that seeks to accelerate Bayesian inference by using polynomial chaos (PC) expansions to represent random variables. Evaluation of integrals over the unknown parameter space is recast, more efficiently, as Monte Carlo sampling of the random variables underlying the PC expansion. We evaluate the utility of this technique on a transient diffusion problem arising in contaminant source inversion. The accuracy of posterior estimates is examined with respect to the order of the PC representation, the choice of PC basis, and the decomposition of the support of the prior. The computational cost of the new scheme shows significant gains over direct sampling.

Keywords: Inverse problems; Bayesian inference; Polynomial chaos; Monte Carlo; Markov chain Monte Carlo; Spectral methods; Galerkin projection; Diffusive transport

1. Introduction

Inverse problems, broadly defined, arise from indirect observations of a quantity of interest [1,2]. A physical system may be described by a *forward model*, which predicts some measurable features of the system given a set of parameters. The corresponding inverse problem consists of inferring these parameters from a set of observations of the features.

The simplicity of this definition belies many fundamental challenges. In realistic applications, data is almost always noisy or uncertain. Also, the forward model may have limitations on its predictive value; i.e. it may be an imperfect or imprecise model of the physical system. Furthermore, as highlighted in [2], inverse problems are often *non-local* and/or *non-causal*. In a forward model, solution values usually depend only on neighboring regions of space and affect only future values in time. Inverting these models, however, may (implicitly) require time-reversal or deconvolution. In mathematical terms, these properties render inverse problems ill-posed. No

* Corresponding author. Tel.: +1 925 294 1283; fax: +1 925 294 2595.
E-mail addresses: ymarz@alum.mit.edu, ymarzou@sandia.gov (Y.M. Marzouk), hnnajm@sandia.gov (H.N. Najm), rahn@sandia.gov (L.A. Rahn).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2006.10.010

feasible parameters may match the observed data (existence), or a multiplicity of model parameters may fit the data (uniqueness). Small errors in measurement can lead to enormous changes in the estimated model (stability).

The Bayesian setting for inverse problems offers a rigorous foundation for inference from noisy data and uncertain forward models, a natural mechanism for incorporating prior information, and a quantitative assessment of uncertainty in the inferred results [3,4]. Indeed, the output of Bayesian inference is not a single value for the model parameters, but a probability distribution that summarizes all available information about the parameters. From this *posterior* distribution, one may estimate means, modes, and higher-order moments, compute marginal distributions, or make additional predictions by averaging over the posterior.

Bayesian approaches to inverse problems have seen much recent interest [3,5,2], with applications ranging from geophysics [6,7] and climate modeling [8] to heat transfer [9,10]. In all of these applications, the primary computational challenge remains one of extracting information from the posterior density [11,12]. Most estimates take the form of integrals over the posterior, which may be computed with asymptotic methods, deterministic methods, or sampling. Deterministic quadrature or cubature [13,14] may be attractive alternatives to Monte Carlo at low to moderate dimensions, but Markov chain Monte Carlo (MCMC) [15–17] remains the most general and flexible method for complex and high-dimensional distributions. All of these methods, however, require evaluation of the likelihood or posterior at many values of the model parameters $\mathbf{m}$. In this setting, evaluating the likelihood requires solving the forward problem. With complex forward models, such as those described by partial differential equations, each single evaluation can be a computationally expensive undertaking [18]. For Monte Carlo simulations requiring 10^3–10^5 samples, the total cost of these forward evaluations quickly becomes prohibitive.

This paper presents a new formulation designed to accelerate evaluation of Bayesian integrals and other characterizations of the posterior. We develop methods to substantially reduce the cost of evaluating the posterior density, based on a stochastic spectral reformulation of the forward problem. These methods have their roots in uncertainty quantification (UQ) using polynomial chaos (PC) expansions [19–21].

The efficient *forward* propagation of uncertainty—i.e. from model parameters to model predictions—is a central challenge of uncertainty quantification. A simple approach is Monte Carlo simulation: sampling known distributions of the model parameters to obtain statistics or density estimates of the model predictions. Again, each sample requires a solution of the forward model, and with complex models, this sampling approach is computationally intensive. A useful alternative is to employ spectral representations of uncertain parameters and field quantities, specifically polynomial chaos (PC) expansions for random variables and stochastic processes. The polynomial chaos [19,22–26] was first defined by Wiener [22]; successive polynomial chaoses give rise to a functional basis consisting of Hermite polynomials of Gaussian random variables [27]. Ghanem and Spanos [19] describe the implementation of polynomial chaos in a finite element context. These stochastic finite element approaches have found numerous modeling applications, including transport in porous media [28], and solid [29,30] or structural [31] mechanics. Le Maître et al. [20,32] extended these techniques to thermo-fluid systems. Xiu et al. [33] used generalized polynomial chaos [34] for uncertainty quantification in fluid-structure interactions and in diffusion problems [35], while Debusschere et al. [36] used polynomial chaos to characterize uncertainty in electrochemical microfluid systems.

We will show that Bayesian estimation is intimately related to the forward propagation of uncertainty. In particular, using PC to propagate a wide range of uncertainty—e.g. prior uncertainty—through the forward problem and sampling the resulting spectral expansion enables a substantially more efficient Bayesian solution of the inverse problem. To this end, we employ an "intrusive" stochastic spectral methodology, in which polynomial chaos representations of the unknown parameters lead to a reformulation of the governing equations of the forward model. This process involves: (1) constructing PC expansions $\mathbf{g}(\boldsymbol{\xi})$ for each unknown parameter, according to probability distributions that include the support of the prior; (2) substituting these expansions into the governing equations and using Galerkin projection to obtain a coupled system of equations for the PC mode strengths; (3) solving this system; and (4) forming an expression for the posterior density based on the resulting PC expansions of forward model predictions, then exploring this posterior density with an appropriate sampling strategy. In this scheme, sampling can have negligible cost; nearly all the computational time is spent solving the system in step 3. Depending on model non-linearities and the necessary size of the PC

basis, this computational effort may be orders of magnitude less costly than exploring the posterior via direct sampling.

Other attempts at accelerating Bayesian inference in computationally intensive inverse problems have relied on reductions of the forward model. Wang and Zabaras [10] use proper orthogonal decomposition (POD) [37] and Galerkin projection to speed forward model calculations in a radiative source inversion problem. The empirical basis [38] used for model reduction is pre-constructed using full forward problem simulations. The choice of inputs to these simulations—in particular, how closely the inputs must resemble the inverse solution—can be important [10]. Balakrishnan et al. [39] introduce a PC representation of the forward model in a groundwater transport parameter identification problem, but obtain the PC coefficients by collocation; again, this process depends on a series of "snapshots" obtained from repeated forward simulations.

A different set of approaches, found in the statistical literature, retain the full forward model but use reduced models to guide and improve the efficiency of MCMC. Christen and Fox [40] use a local linear approximation of the forward model to improve the acceptance probability of proposed moves, reducing the number of times the likelihood must be evaluated with the full forward model. This "approximate MCMC" algorithm is shown to yield the same stationary distribution as a standard Metropolis–Hastings chain. Higdon et al. [18] focus on the estimation of spatially distributed inputs to a complex forward model. They introduce coarsened representations of the inputs and apply a Metropolis-coupled MCMC scheme [41] in which "swap proposals" allow information from the coarse-scale formulation, which may be computed more quickly, to influence the fine-scale chain. In contrast to the present formulation, however, all of the approaches cited above require repeated solutions of the full-scale deterministic forward model.

We will demonstrate our new formulation on a transient diffusion problem arising in contaminant source inversion, and compare the efficiency of the method and the accuracy of posterior estimates to direct evaluation of the posterior.

2. Formulation

2.1. Bayesian inference for inverse problems

Consider a forward problem defined as follows:

$$\mathbf{d} \approx \mathbf{G}(\mathbf{m}) \tag{1}$$

Here $\mathbf{m}$ is a vector of model parameters and $\mathbf{d}$ is a vector of observable quantities, or data. The forward model $\mathbf{G}$ yields predictions of the data as a function of the parameters. In the Bayesian setting, both $\mathbf{m}$ and $\mathbf{d}$ are random variables, and for the remainder of this paper we will take these random variables to be real-valued. We use Bayes' rule to define a posterior probability density for the model parameters $\mathbf{m}$, given an observation of the data $\mathbf{d}$:

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p_m(\mathbf{m})}{\int p(\mathbf{d}|\mathbf{m})p_m(\mathbf{m})\mathrm{d}\mathbf{m}} \tag{2}$$

In the Bayesian setting, probability is used to express knowledge about the true values of the parameters. In other words, prior and posterior probabilities represent *degrees of belief* about possible values of $\mathbf{m}$, before and after observing the data $\mathbf{d}$.

Data thus enters the formulation through the likelihood $p(\mathbf{d}|\mathbf{m})$, which may be viewed as a function of $\mathbf{m}$: $L(\mathbf{m}) \equiv p(\mathbf{d}|\mathbf{m})$. A simple model for the likelihood assumes that independent additive errors account for the deviation between predicted and observed values of $\mathbf{d}$:

$$\mathbf{d} = \mathbf{G}(\mathbf{m}) + \boldsymbol{\eta} \tag{3}$$

where components of $\boldsymbol{\eta}$ are i.i.d. random variables with density p_η. A typical assumption is $\eta_i \sim N(0, \sigma^2)$, in which case $p(\mathbf{d}|\mathbf{m})$ becomes $N(\mathbf{G}(\mathbf{m}), \sigma^2\mathbf{I})$. The likelihood is thus

$$L(\mathbf{m}) = \prod_i p_\eta(d_i - G_i(\mathbf{m})) \tag{4}$$

In this simple model, $\boldsymbol{\eta}$ may encompass both measurement error (e.g. sensor noise) and model error—the extent to which forward model predictions may differ from "true" values because of some unmodeled physics of the system.

Any additional information on the model parameters may enter the formulation through a suitably-defined prior density, $p_m(\mathbf{m})$. Prior models may embody simple constraints on $\mathbf{m}$ such as a range of feasible values, or may reflect more detailed knowledge about the parameters (shapes, correlations, smoothness, etc.). In the absence of additional information, one may simply choose a prior that is *uninformative*.

Bayesian estimation typically gives rise to integrals over the posterior density:

$$I[f] = \int f(\mathbf{m}) L(\mathbf{m}) p_m(\mathbf{m}) \mathrm{d}\mathbf{m} \tag{5}$$

The posterior expectation of a function f, for instance, is $\mathbb{E}_\pi f = I[f]/I[1]$. Though more sophisticated means of estimating such integrals will be discussed later, we note here that a conceptually simple method of obtaining posterior estimates is Monte Carlo sampling. If independent samples $\mathbf{m}^{(j)}$ can be drawn from the prior, then a Monte Carlo estimate of (5) is

$$\widehat{I}_n[f] = \frac{1}{n} \sum_{j=1}^{n} \left[f\left(\mathbf{m}^{(j)}\right) \prod_i p_\eta\left(d_i - G_i(\mathbf{m}^{(j)})\right) \right] \tag{6}$$

If parameters $\boldsymbol{\phi}_m$ of the prior density $p_m(\mathbf{m}|\boldsymbol{\phi}_m)$ or parameters $\boldsymbol{\phi}_\eta$ of the error model $p_\eta(\eta_i|\boldsymbol{\phi}_\eta)$ are not known *a priori*, they may become additional objects for Bayesian inference. In other words, these *hyperparameters* may themselves be endowed with priors and estimated from data [3]:

$$p(\mathbf{m}, \boldsymbol{\phi}_m, \boldsymbol{\phi}_\eta|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m}, \boldsymbol{\phi}_\eta) p_m(\mathbf{m}|\boldsymbol{\phi}_m) p(\boldsymbol{\phi}_\eta) p(\boldsymbol{\phi}_m) \tag{7}$$

The resulting joint posterior over model parameters and hyperparameters may then be interrogated in various ways—e.g. by marginalizing over the hyperparameters to obtain $p(\mathbf{m}|\mathbf{d})$; or first marginalizing over $\mathbf{m}$ and using the maximizer of this density as an estimate of the hyperparameters; or by seeking the joint maximum *a posteriori* estimate or posterior mean of $\mathbf{m}$, $\boldsymbol{\phi}_m$ and $\boldsymbol{\phi}_\eta$ [42,3].

2.2. Polynomial chaos expansions

Let $(\Omega, \mathscr{U}, P)$ be a probability space, where Ω is a sample space, $\mathscr{U}$ is a σ-algebra over Ω, and P is a probability measure on $\mathscr{U}$. Also, let $\{\xi_i(\omega)\}_{i=1}^{\infty}$ be a set of orthonormal standard Gaussian random variables on Ω. Then any square-integrable random variable $X : \Omega \to \mathbb{R}$ has the following representation:

$$X(\omega) = a_0 \Gamma_0 + \sum_{i_1=1}^{\infty} a_{i_1} \Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} a_{i_1 i_2} \Gamma_2(\xi_{i_1}, \xi_{i_2}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3} \Gamma_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + \cdots \tag{8}$$

where Γ_p is the Wiener polynomial chaos of order p [22,19,43]. This expansion may be re-written in a more compact form

$$X(\omega) = \sum_{k=0}^{\infty} \hat{a}_k \Psi_k(\xi_1, \xi_2, \ldots) \tag{9}$$

where there is a one-to-one correspondence between the coefficients and functionals in (8) and (9) [19]. For the standard normal random variables ξ_i chosen above, orthogonality of successive Γ_p requires that the Γ_p be multivariate Hermite polynomials; both these and the corresponding Ψ_k may be generated from univariate Hermite polynomials by taking tensor products.

Of course, in computations it is not useful to represent a random variable with an infinite summation, and one truncates the expansion both in order p and in dimension n—i.e. by choosing a subset $\boldsymbol{\xi} = \{\xi_{\lambda_i}\}_{i=1}^{n}$ of the infinite set $\{\xi_i\}$, $\lambda_i \in \mathbb{N}$. The total number of terms P in the finite *polynomial chaos expansion*

$$X(\omega) = \sum_{k=0}^{P} x_k \Psi_k(\xi_1, \xi_2, \ldots, \xi_n) \tag{10}$$

is

$$P+1=\frac{(n+p)!}{n!p!} \tag{11}$$

Polynomial chaos (PC) expansions have been generalized to broader classes of orthogonal polynomials in the Askey scheme, each family resulting from a different choice of distribution for the ξ_i [34,44]. For each of these choices, orthogonality of the polynomials $\Psi_k(\boldsymbol{\xi})$ with respect to the inner product on $L^2(\Omega)$ is guaranteed:

$$\langle \Psi_i \Psi_j \rangle = \int \Psi_i(\boldsymbol{\xi}(\omega))\Psi_j(\boldsymbol{\xi}(\omega))\mathrm{d}P(\omega) = \int \Psi_i(\boldsymbol{\xi})\Psi_j(\boldsymbol{\xi})w(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} = \delta_{ij}\langle \Psi_i^2 \rangle \tag{12}$$

where, in the second (Riemann) integral, $w(\boldsymbol{\xi})$ denotes the probability density of $\boldsymbol{\xi}$. This property can be used to calculate the truncated PC representation of a random variable $f \in L^2(\Omega)$ by projecting onto the PC basis:

$$\tilde{f}(\omega) = \sum_{k=0}^{P} f_k \Psi_k(\boldsymbol{\xi}), \quad f_k = \frac{\langle f(X)\Psi_k \rangle}{\langle \Psi_k^2 \rangle} \tag{13}$$

This orthogonal projection minimizes the error $\|f - \tilde{f}\|$ on the space spanned by $\{\Psi_k\}_{k=0}^{P}$, where $\|\cdot\|$ is the inner-product norm on $L^2(\Omega)$.

Suppose that the behavior of f can be expressed as $\mathcal{O}(f, X) = 0$, where $\mathcal{O}$ is some deterministic operator and $X(\omega)$ is a random variable with a known PC expansion $X = \sum_i^P x_i \Psi_i$. Substituting PC expansions for f and X into this operator and requiring the residual to be orthogonal to Ψ_j for $j = 0 \ldots P$ yields a set of coupled, deterministic equations for the PC coefficients f_k:

$$\left\langle \mathcal{O}\left(\sum_k^P f_k \Psi_k, \sum_i^P x_i \Psi_i \right) \Psi_j \right\rangle = 0, \quad j = 0 \ldots P \tag{14}$$

This Galerkin approach is known as "intrusive" spectral projection [32], in contrast to "non-intrusive" approaches in which the inner product $\langle f(X)\Psi_k \rangle$ is evaluated by sampling or quadrature, thus requiring repeated evaluations of $f(X)$ corresponding to different realizations of $\boldsymbol{\xi}$ [45].

In practice, we employ a pseudospectral construction to perform intrusive projections efficiently for higher powers of random variables, e.g. $f(X) = X^j$, $j \geqslant 3$, and have developed additional techniques for non-polynomial functions f. These operations are incorporated into a library for "stochastic arithmetic," detailed in [21].

2.3. Efficient evaluation of the posterior

2.3.1. Sampling from the prior

We now connect ideas introduced in the two preceding sections to formulate a computationally efficient scheme for Bayesian inference. Suppose the model parameters have been endowed with a prior density $p_m(\mathbf{m})$. Knowing this density, one can construct a corresponding polynomial chaos expansion for each component m_i of the random vector $\mathbf{m}$:

$$m_i(\boldsymbol{\xi}) = \sum_{k=0}^{P} m_{ik} \Psi_k(\boldsymbol{\xi}) \tag{15}$$

where the dimension of $\boldsymbol{\xi}$ is at least equal to the dimension of $\mathbf{m}$.

Next, we introduce these PC expansions into the forward model and use Galerkin projection to obtain a PC representation for each component of the predicted data $G_i(\mathbf{m})$. Here $G_i(\mathbf{m})$ denotes the ith component of $\mathbf{G}(\mathbf{m})$ and $\widetilde{G}_i(\boldsymbol{\xi})$ is its (approximate) PC representation:

$$\widetilde{G}_i(\boldsymbol{\xi}) = \sum_{k=0}^{P} d_{ik} \Psi_k(\boldsymbol{\xi}) \tag{16}$$

Now consider a generic integral over the unnormalized posterior density, given in (5). Drawing samples $\boldsymbol{\xi}^{(j)}$ from the distribution of $\boldsymbol{\xi}$ will yield samples of $\mathbf{m}$ from the prior, calculated according to (15). But the corresponding forward model prediction, $\mathbf{G}(\mathbf{m})$, can now be computed very cheaply, simply by substituting the same

$\boldsymbol{\xi}^{(j)}$ into (16). In general, evaluating this $(P+1)$-term expansion will be significantly faster than solving the forward model for each sample. The prediction $\widetilde{\mathbf{G}}(\boldsymbol{\xi})$ obtained in this fashion then replaces $\mathbf{G}(\mathbf{m})$ in the likelihood $L(\mathbf{m})$. For the simple likelihood proposed in (4), an n-sample Monte Carlo estimate of the integral becomes:

$$\widehat{I}_n[f] = \frac{1}{n}\sum_{j=1}^{n}\left[f(\mathbf{m}(\boldsymbol{\xi}^{(j)}))\prod_i p_\eta\left(d_i - \widetilde{G}_i(\boldsymbol{\xi}^{(j)})\right)\right] \tag{17}$$

2.3.2. Sampling from alternate distributions

Polynomial chaos reformulations of a Bayesian estimate need not be limited to expansions for which $\mathbf{m}$ has density p_m, as specified in (15), however. Consider a different set of PC expansions, $\mathbf{m} = \mathbf{g}(\boldsymbol{\xi})$, where each component $m_i = g_i$ is given by

$$m_i = g_i(\boldsymbol{\xi}) = \sum_{k=0}^{P} g_{ik}\Psi_k(\boldsymbol{\xi}) \tag{18}$$

Let the g_{ik} be chosen such that $\mathbf{g}(\boldsymbol{\xi})$ has probability density q, where the support of q includes the support of the prior. Drawing samples $\boldsymbol{\xi}^{(j)}$ from the distribution of $\boldsymbol{\xi}$ will now yield samples of $\mathbf{m}$ drawn from q. A Monte Carlo estimate of the integral in (5) now becomes:

$$\widehat{I}_n[f] = \frac{1}{n}\sum_{j=1}^{n}\left[f(\mathbf{g}(\boldsymbol{\xi}^{(j)}))\prod_i p_\eta\left(d_i - \widetilde{G}_i(\boldsymbol{\xi}^{(j)})\right)\frac{p_m(\mathbf{g}(\boldsymbol{\xi}^{(j)}))}{q(\mathbf{g}(\boldsymbol{\xi}^{(j)}))}\right] \tag{19}$$

If q is chosen to sample from regions where $|f(\mathbf{m})|L(\mathbf{m})p_m(\mathbf{m})$ is relatively large, then the estimate in (19) amounts to importance sampling [46,47]. The variance of $f(\mathbf{m})L(\mathbf{m})p_m(\mathbf{m})/q(\mathbf{m})$, where $\mathbf{m}\sim q$, will be reduced by this sampling strategy and consequently the variance of the estimator $\widehat{I}_n[f]$ will be reduced as well. In the present context, of course, evaluating the likelihood of each sample via PC expansions is already inexpensive, so importance sampling may not yield significant gain. However, freedom in the choice of $\mathbf{g}(\boldsymbol{\xi})$ has a more immediate utility. If the prior distribution is such that it is difficult or inconvenient to write a PC expansion for $\mathbf{m}$ with density p_m, the ability to choose a density q that may be simpler than p_m ensures flexiblity in prior modeling.

2.3.3. Change of variables and MCMC

The essence of the two sampling schemes presented above is that the likelihood, when computed using PC expansions, becomes a function of $\boldsymbol{\xi}$ rather than of $\mathbf{m}$. Thus we sample $\boldsymbol{\xi}$ to generate samples of $\mathbf{m}$ from a specified distribution and simultaneously use $\boldsymbol{\xi}$ to compute the likelihood of each sample. Implicit in these schemes is a change of variables from $\mathbf{m}$ to $\boldsymbol{\xi}$, and it is fruitful to consider this change explicitly, as follows:

$$I[f] = \int_{\mathscr{M}} f(\mathbf{m})L(\mathbf{m})p_m(\mathbf{m})\mathrm{d}\mathbf{m} = \int_{\Xi_m} f(\mathbf{g}(\boldsymbol{\xi}))L(\mathbf{g}(\boldsymbol{\xi}))p_m(\mathbf{g}(\boldsymbol{\xi}))|\det D\mathbf{g}(\boldsymbol{\xi})|\mathrm{d}\boldsymbol{\xi} \tag{20}$$

Here, $D\mathbf{g}$ denotes the Jacobian of $\mathbf{g}$. Making this change of variables explicit imposes certain constraints on the transformation $\mathbf{g}$, namely (1) that $\Xi_m = \mathbf{g}^{-1}(\mathscr{M})$, the inverse image of the support of the prior, be contained within the range of $\boldsymbol{\xi}$, and (2) that $\mathbf{g}$ be a differentiable transformation from Ξ_m to $\mathscr{M}$ with a differentiable inverse (i.e. that $\mathbf{g}$ be a diffeomorphism from Ξ_m to $\mathscr{M}$).The first constraint is not new; indeed, in the preceding two sections, it is necessarily satisfied by a PC expansion that reproduces samples of $\mathbf{m}$ from the desired distribution, p_m or q. The latter constraint, however, limits $\mathbf{g}$ to rather simple PC expansions—for instance, linear transformations of $\boldsymbol{\xi}$.[1] But this limitation is not a great liability, as the transformed integral in (20) can now be

[1] This condition is not required by the sampling schemes in Sections 2.3.1 and 2.3.2. If, however, it *is* satisfied by $\mathbf{g}$ in Section 2.3.1 then we have

$$p_m(\mathbf{g}(\boldsymbol{\xi}))|\det D\mathbf{g}(\boldsymbol{\xi})| = w(\boldsymbol{\xi})$$

where w is defined in (12). An analogous condition holds true for q in Section 2.3.2, again when $\mathbf{g}$ is a diffeomorphism from Ξ_q to $\mathscr{M}$.

evaluated by any suitable sampling scheme in $\boldsymbol{\xi}$-space—in particular, by Markov chain Monte Carlo (MCMC).

MCMC encompasses a broad class of methods that simulate drawing samples from the posterior [17,48,46], and thus can be used to directly estimate the posterior expectation of f:

$$\mathbb{E}_{\pi_m} f = \int f(\mathbf{m})\pi_m(\mathbf{m})\mathrm{d}\mathbf{m} \tag{21}$$

where π_m denotes the normalized posterior density of $\mathbf{m}$

$$\pi_m(\mathbf{m}) \propto L_m(\mathbf{m})p_m(\mathbf{m}) \tag{22}$$

The subscript on L_m emphasizes that the likelihood is here a function of $\mathbf{m}$. The change of variables from $\mathbf{m}$ to $\boldsymbol{\xi}$ can be expressed compactly as

$$\mathbb{E}_{\pi_m} f = \mathbb{E}_{\pi_\xi}(f \circ \mathbf{g}) \tag{23}$$

where π_ξ is the posterior density in $\boldsymbol{\xi}$-space:

$$\pi_\xi(\boldsymbol{\xi}) \propto L_m(\mathbf{g}(\boldsymbol{\xi}))p_m(\mathbf{g}(\boldsymbol{\xi}))|\det D\mathbf{g}(\boldsymbol{\xi})| \tag{24}$$

As before, we would like to use the PC representation of $\mathbf{G}$ to accelerate evaluations of the posterior. We first introduce the following notational convention: any quantity computed by projection onto a finite PC basis—whether a forward model prediction, likelihood, or posterior density—is distinguished from its "direct" counterpart with a tilde. Thus, we seek samples from the surrogate posterior $\tilde{\pi}_\xi$:

$$\tilde{\pi}_\xi(\boldsymbol{\xi}) \propto \widetilde{L}_\xi(\boldsymbol{\xi})p_m(\mathbf{g}(\boldsymbol{\xi}))|\det D\mathbf{g}(\boldsymbol{\xi})| \propto \prod_i p_\eta\Big(d_i - \widetilde{G}_i(\boldsymbol{\xi})\Big)p_m(\mathbf{g}(\boldsymbol{\xi}))|\det D\mathbf{g}(\boldsymbol{\xi})| \tag{25}$$

The likelihood is now a function of $\boldsymbol{\xi}$. Since Metropolis–Hastings algorithms require knowledge only of the unnormalized posterior density, (25) is sufficient to simulate samples $\boldsymbol{\xi}^{(j)}$ from the posterior on $\boldsymbol{\xi}$-space. Eliminating b burn-in samples, the posterior expectation of f is estimated by an ergodic average:

$$\mathbb{E}_{\pi_m} f \approx \frac{1}{n-b}\sum_{j=b+1}^{n}(f \circ \mathbf{g})(\boldsymbol{\xi}^{(j)}) \tag{26}$$

Note that an MCMC chain could just as easily be run on $\mathbf{m}$-space, simulating samples from the surrogate posterior $\tilde{\pi}_m \propto (\widetilde{L}_\xi \circ \mathbf{g}^{-1}) \cdot p_m$. But this expression emphasizes why an invertible $\mathbf{g}$ must be used with MCMC; otherwise the argument to $\widetilde{L}_\xi$ corresponding to a given chain position $\mathbf{m}^{(j)}$ would be ill-defined.

Advantages of MCMC over the simple Monte Carlo schemes in Sections 2.3.1 and 2.3.2 are several. In many applications, a well-designed MCMC algorithm can offer far better sampling efficiency than sampling from the prior or from some alternate distribution $q \neq \pi_m$, despite the fact that MCMC samples are serially correlated [49]. In the present context, however, the value of improved sampling efficiency is tempered by the fact that samples are rendered inexpensive by the PC reformulation. However, MCMC offers additional benefits. Because MCMC directly simulates the posterior, it is simple to extract marginal densities for individual components of $\mathbf{m}$ with the aid of kernel density estimation [50]. Also, MCMC eliminates the need to calculate the posterior normalization factor I[1]. Using (25) to evaluate the posterior, further particulars of the MCMC algorithm we adopt are essentially independent of the PC formulation and thus we reserve their presention for Section 3.5.

The choice of $\mathbf{g}$ and of the PC basis $\{\Psi_k(\boldsymbol{\xi})\}_{k=0}^{P}$, on the other hand, will have a crucial impact on the accuracy and cost of PC-reformulated posterior estimates and predictions. Because it is the initial PC representation of $\mathbf{m}$—its distribution defining the stochastic forward problem—$\mathbf{g}(\boldsymbol{\xi})$ will directly affect $\widetilde{\mathbf{G}}$, as will the order and stochastic dimension of the PC basis used in Galerkin projections. The "surrogate" posterior $\tilde{\pi}$, obtained by replacing direct evaluation of the likelihood $L(\mathbf{m})$ with a likelihood written in terms of $\widetilde{\mathbf{G}}(\boldsymbol{\xi})$, is then at the heart of any PC-induced errors in the three sampling schemes discussed above. Whether we write this posterior in terms of $\boldsymbol{\xi}$ (e.g. $\tilde{\pi}_\xi$) or $\mathbf{m}$ (e.g. $\tilde{\pi}_m$, if $\mathbf{g}$ is invertible), the difference between π and $\tilde{\pi}$ completely captures the impact of the polynomial chaos representation of forward model predictions on the inverse

solution. We will explore the dependence of this posterior error on $\mathbf{g}$, p, and the distribution of $\boldsymbol{\xi}$—i.e. the type of PC basis—in Sections 3.3 and 3.6.

2.4. Decomposition of parameter space

When a component of the forward solution depends very steeply on an uncertain parameter, a PC basis of smooth, global polynomials may require increasingly high order to provide an accurate solution of the stochastic forward problem. In the diffusion-driven forward problem to be considered here, many terms will be required to represent a sharply-localized source with broad prior uncertainty in its location. In the limit, when solutions are *discontinuous* with respect to an uncertain parameter, e.g. in the neighborhood of a critical point, a global PC basis may be unsuitable [51].

Several methods have been proposed to address this difficulty [51–53]. A PC basis of Haar wavelets was constructed in [51], then generalized to a multi-wavelet basis in [52]; both of these methods effectively resolved stochastic systems with multiple bifurcations. Computational efficiency of the latter scheme was improved by block-partitioning of the uncertain parameter space. We adopt a similar, but non-adaptive, partitioning scheme here. The support $\mathcal{M}$ of the prior, or equivalently the range of $\mathbf{m}$, is decomposed into N_b non-overlapping domains

$$\mathcal{M} = \bigcup_b^{N_b} \mathcal{M}^b, \quad \mathcal{M}^b \cap \mathcal{M}^{b'} = \emptyset \quad \text{if } b \neq b' \tag{27}$$

We seek a corresponding decomposition of the prior density $p_m(\mathbf{m})$, and thus of the Bayesian integral $I[f]$ in (5), as follows:

$$\begin{aligned} p^b(\mathbf{m}) &= \begin{cases} p_m(\mathbf{m}) & \mathbf{m} \in \mathcal{M}^b \\ 0 & \mathbf{m} \notin \mathcal{M}^b \end{cases} \\ p_m(\mathbf{m}) &= \sum_b^{N_b} p^b(\mathbf{m}) \end{aligned} \tag{28}$$

Note that the densities $p^b(\mathbf{m})$ are unnormalized on each domain $\mathcal{M}^b$, so that

$$I[f] = \sum_b^{N_b} I^b[f] = \sum_b^{N_b} \int f(\mathbf{m}) L(\mathbf{m}) p^b(\mathbf{m}) \mathrm{d}\mathbf{m} \tag{29}$$

This partitioning allows the construction of a separate polynomial expansion $\mathbf{g}^b(\boldsymbol{\xi})$ for $\mathbf{m}$ on each block, and thus a different version of the likelihood $\widetilde{L}_\xi$ in each integral $I^b[f]$ contributing to (5). What does this imply for $\mathbf{g}^b$? Again, we identify three cases, corresponding to the sampling schemes in Sections 2.3.1, 2.3.2, and 2.3.3.

When sampling from the prior, we seek $\mathbf{g}^b$ so that $\mathbf{m}$ has density proportional to $p^b(\mathbf{m})$. Sampling $\boldsymbol{\xi}$ will then yield samples from the prior on each block.

When sampling from an alternate distribution, e.g. from N_b densities $q^b(\mathbf{m})$, we simply require that the support of each q^b contain $\mathcal{M}^b$. In particular, the supports of each q^b need not be disjoint; the definition of $p^b(\mathbf{m})$ above ensures zero posterior density outside of each domain.

Finally, in an MCMC scheme, we work directly with the partition of Ξ and allow the chain to transition from block to block according to the proposal distribution.

Le Maître et al. [52] provide criteria for adaptively refining the partitions based on the local variance of the solution. Here, because the inverse problem to be considered has a simple symmetry in both its priors and its computational domain, we will take a fixed partition of $\mathcal{M} = [0,1] \times [0,1]$ into four equal quadrants.

3. Results

We demonstrate the stochastic spectral formulation of Bayesian inference by inverting for the source field in a transient diffusion problem.

One practical context of this inverse problem lies in contaminant source inversion [54]. Given a sparse set of concentration measurements—from sensors scattered throughout some space, for instance—one would like to find the sources of a toxin that has spread through the ambient medium. Specific parameters to infer include the number of sources and their possibly time-dependent strengths and locations. Though convective transport can play a role in many practical source inversion problems, we will limit our attention here to a purely diffusion-driven inverse problem in order to focus on the demonstration and analysis of the new formulation. Diffusive source inversion problems themselves arise in the context of porous media flows [55] and heat conduction [56–60].

3.1. Source inversion under diffusive transport

We begin by defining the deterministic forward problem $\mathbf{G}(\mathbf{m})$, since this is the basis for the general Bayesian approach to inverse problems described in Section 2.1 and for the stochastic forward problem described in Section 2.3.

Consider a dimensionless diffusion equation on a square domain $S = [0,1]\times[0,1]$ with adiabatic boundaries:

$$\begin{aligned} &\frac{\partial u}{\partial t} = \nabla^2 u + \sum_{l=1}^{N} \frac{s_l}{2\pi\sigma_l^2} \exp\left(-\frac{|\boldsymbol{\chi}_l - \mathbf{x}|^2}{2\sigma_l^2}\right)[1 - H(t - T_l)] \\ &\nabla u \cdot \hat{\mathbf{n}} = 0 \quad \text{on } \partial S \\ &u(\mathbf{x}, 0) = 0 \end{aligned} \tag{30}$$

The source term in (30) describes N localized sources, each one active on the interval $t \in [0, T_l]$ and centered at $\boldsymbol{\chi}_l \in S$ with strength s_l and width σ_l. Note that the location, size, and shutoff time of each source enter the problem non-linearly.

For the purposes of an initial demonstration and to allow direct visualization of the posterior, we restrict our attention to an inverse problem in two dimensions. Thus we fix $N = 1$; prescribe T, s and σ; and leave the source location $\boldsymbol{\chi} = (m_0, m_1)$ unknown. For any given value of $\mathbf{m}$, we solve the PDE in (30) using a finite difference method. The u-field is described on an uniform grid with spacing $h = 0.025$. Second-order centered differences are used to discretize the diffusion terms. Time integration is via an explicit, second-order-accurate, Runge–Kutta–Chebyshev (RKC) scheme [61] with $\Delta t = 0.002$. The number of substeps in the RKC scheme is automatically determined by stability constraints upon setting ϵ, the damping parameter that controls the extent of the stability region, to 2/13 [62]. Numerical resolution studies were conducted to validate the present choices of h and Δt.

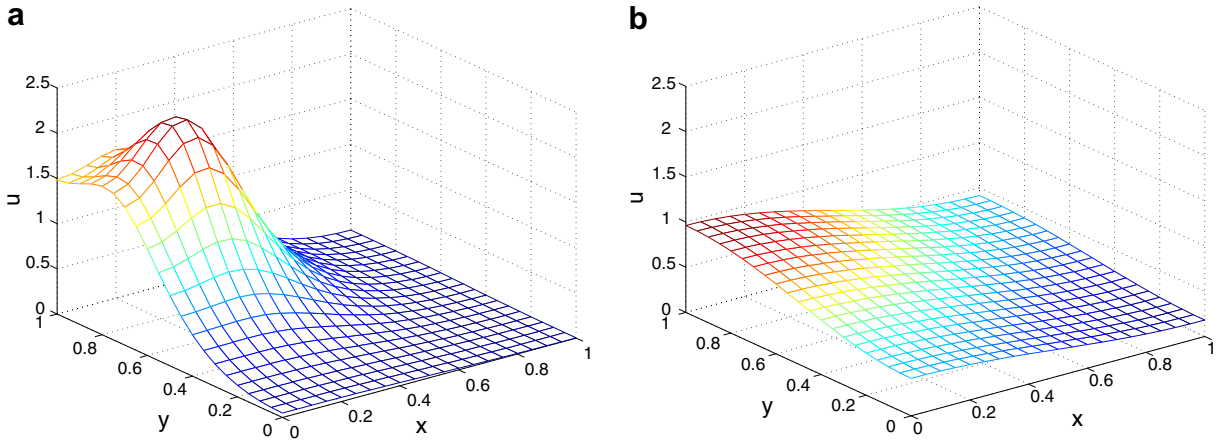


Fig. 1. Scalar field u in the deterministic forward problem, for $(m_0, m_1) = (0.25, 0.75)$. (a) $t = 0.05$ and (b) $t = 0.15$.

The forward model $\mathbf{G}(\mathbf{m})$ predicts the value of the field $u(\mathbf{x}, t)$ at specific locations and times. Below, unless otherwise specified, $\mathbf{G}$ will provide predictions on an uniform 3×3 grid covering the domain S at two successive times, $t = 0.05$ and $t = 0.15$. The *inverse problem* thus consists of inferring the source position from *noisy* measurements at these locations and times. We let independent zero-mean Gaussian random variables $\eta_i \sim N(0, \varsigma^2)$ express the difference between "real-world" measurements and model predictions, as specified in (3). In the examples below, we choose $\varsigma = 0.4$ unless otherwise specified. Priors simply constrain the source to lie in the domain S, i.e. $m_i \sim U(0,1)$. Again, in the interest of simplicity, we make no attempt at hyperparameter estimation for either the noise model or the prior distribution.

Fig. 1 shows the u-field resulting from a representative value of $\mathbf{m}$: $(m_0, m_1) = (0.25, 0.75)$, with $T = 0.05$, $s = 0.5$ and $\sigma = 0.1$. Though the solution is initially peaked around the source, note that it flattens at the later time once the source is no longer active. As the measurement time $(t > T) \to \infty$, the inverse problem becomes increasingly ill-conditioned. Measurement noise will overwhelm any residual variation in the u-field resulting from the particular location of the source.

3.2. Stochastic spectral solution of the forward problem

The accuracy, and computational expense, of the stochastic spectral solution of the forward problem depend on the order of the PC basis used. For computational efficiency, the requisite PC order can be reduced by partitioning the domain and solving an independent forward problem, with a smaller range of input uncertainty, on each subdomain. Because the present problem has a simple symmetry in both its priors and its computational domain, we take a fixed partition of the prior support $\mathscr{M} = [0,1] \times [0,1]$ into four equal quadrants $\mathscr{M}^b$. On each of these quadrants, we prescribe a PC expansion $\mathbf{m} = \mathbf{g}^b(\boldsymbol{\xi})$ consisting of multivariate Legendre polynomials and uniformly distributed $\xi_i \sim U(-1,1)$. In particular, we choose PC coefficients such that each $\mathbf{g}^b(\boldsymbol{\xi})$ has an uniform probability density on $\mathscr{M}^b$ and zero probability density elsewhere. On each quadrant, $\mathbf{m}$ is thus distributed according to the (normalized) prior density $p^b(\mathbf{m})$ given in (28), e.g.

$$\mathbf{g}^{b=1}(\boldsymbol{\xi}) = \begin{pmatrix} 1/4 \\ 1/4 \end{pmatrix} + \begin{pmatrix} 1/4 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \tag{31}$$

and so on for $b = 2\ldots4$.

The stochastic forward problem is then solved four times, once for each block of the prior support. On each block, we introduce the PC expansion $\boldsymbol{\chi} = \mathbf{m} = \mathbf{g}^b(\boldsymbol{\xi})$ into (30) with $N = 1$ and, using Galerkin projections and the same finite-difference/RKC scheme as in the deterministic problem, obtain a PC expansion for each prediction of the scalar field, $\widetilde{G}_i^b(\boldsymbol{\xi})$. These predictions are random variables $u(\mathbf{x}_i, t_i, \boldsymbol{\xi}(\omega))$, giving the value of the scalar field at each measurement location and time $(\mathbf{x}_i, t_i)$.

Note that the source term in (30) may be factored into a time-dependent component and a stationary component:

$$\sum_{l=1}^{N} \frac{s_l}{2\pi\sigma_l^2} \exp\left(-\frac{|\boldsymbol{\chi}_l - \mathbf{x}|^2}{2\sigma_l^2}\right)[1 - H(t - T_l)] = \sum_{l=1}^{N} q(\mathbf{x}, \boldsymbol{\chi}_l) s(t, T_l) \tag{32}$$

The stationary component $q(\mathbf{x}, \boldsymbol{\chi})$ contains the exponential of a random variable, and its PC representation is thus expensive to compute. In the interest of efficiency, we evaluate this term once at each grid point $\mathbf{x}_{mn}$ and store the resulting array of PC expansions for use at each subsequent timestep. So that comparisons of computational cost remain fair, we also factor the source term when solving the deterministic forward problem, again pre-computing the stationary part; in other words, the Gaussian source profile is evaluated only once during time integration of (30) for a given source location.

Solutions of the stochastic forward problem may be interrogated in several ways. Fig. 2 shows the predicted value of the scalar field at a particular measurement location and time, $u(x_{i=3} = 0.0, y_{i=3} = 0.5, t_{i=3} = 0.15)$, as a function of $\boldsymbol{\xi}$. This surface is a single component of the stochastic forward problem solution, $\widetilde{G}_i(\boldsymbol{\xi})$ for $i = 3$, and is obtained with PC bases of increasing order ($p = 3$, 6, and 9). Partition of the prior support into quadrants $b = 1\ldots4$ is indicated on each plot. Convergence is observed with increasing p. While there is no guarantee of continuity or smoothness of the solution between neighboring blocks, both seem to be achieved at

sufficient order. The pointwise error in these forward solutions is shown in Fig. 3, again for $i = 3$. Here, error is defined as the difference between the PC expansion of the forward solution and the exact solution of the deterministic forward model $\mathbf{G}(\mathbf{m})$, with argument $\mathbf{m}$ corresponding to the appropriate quadrant b and value of $\boldsymbol{\xi}$:

$$\mathrm{err}_i^b(\boldsymbol{\xi}) = \widetilde{G}_i^b(\boldsymbol{\xi}) - G_i(g^b(\boldsymbol{\xi})) \tag{33}$$

Once again, the error becomes negligible at sufficient order.

Since the input $\mathbf{m} = \mathbf{g}(\boldsymbol{\xi})$ to the forward model is a random variable, any forward model output $G_i(\mathbf{m})$ is also a random variable. The density of these forward model outputs is an useful diagnostic and may be estimated in one of two ways. A direct (and computationally expensive) method is to sample $\mathbf{m}$ and solve the forward problem for each sample, forming a normalized histogram from the resulting collection of forward model outputs. Alternatively, one can sample $\boldsymbol{\xi}$ and substitute it into the PC expansion $\widetilde{G}_i(\boldsymbol{\xi})$, again forming a histogram of the resulting values. This process essentially weighs the surface response in Fig. 2 according to the probability distribution of $\boldsymbol{\xi}$. The resulting density estimates for G_3 are shown in Fig. 4. While a lower-order PC basis ($p = 3$) results in a poor density estimate, the probability density converges to its true shape—obtained by the direct method—as p increases.

The probability densities computed above represent the propagation of prior uncertainty through the forward problem. Accordingly, we may endow them with an additional interpretation. Fig. 5 shows the probability density of u at a single measurement location ($x = 0.0$, $y = 0.0$) but at two successive times: $t = 0.05$ and $t = 0.15$. As in Fig. 4, we observe convergence of the PC-obtained density to its "direct" counterpart with

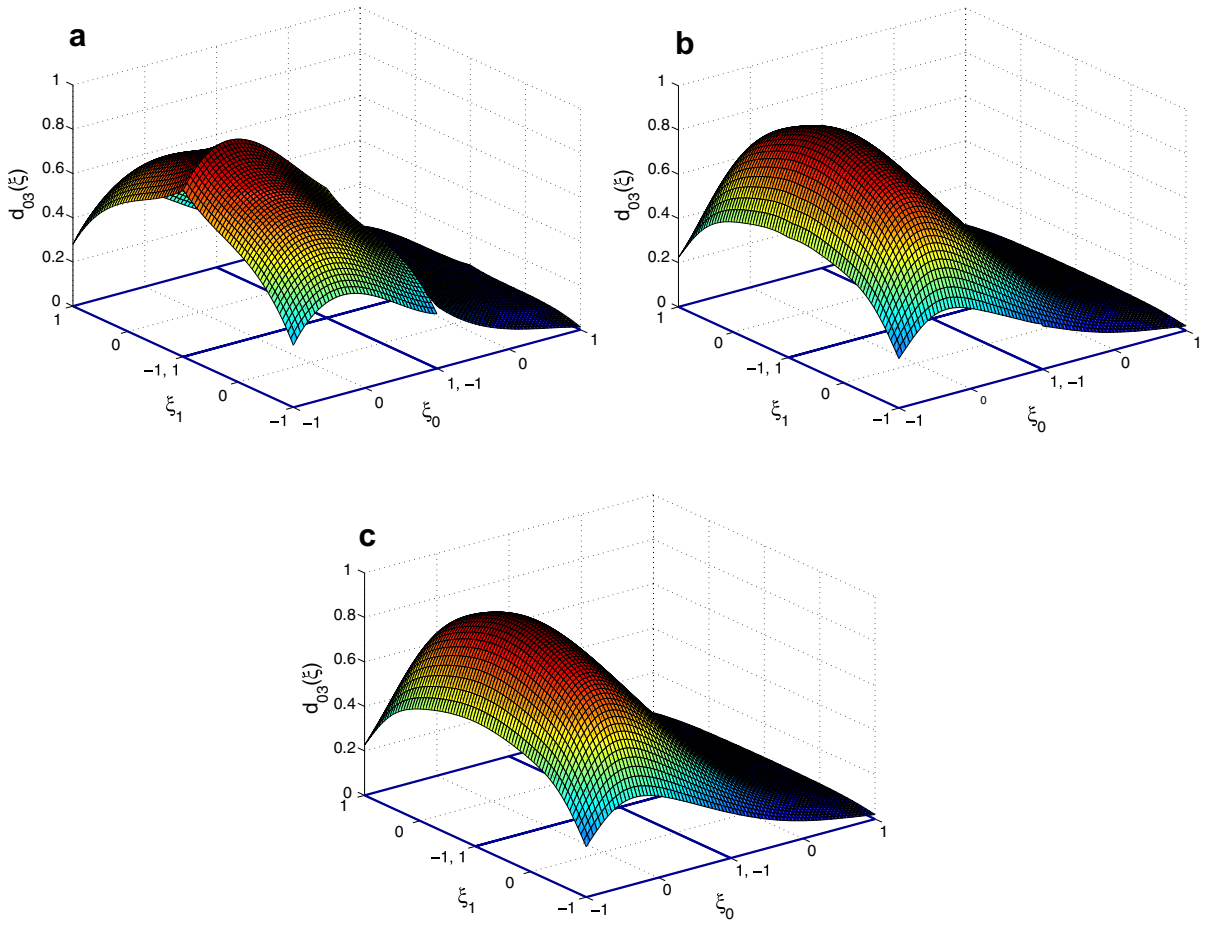


Fig. 2. A single component of the stochastic forward problem solution, $\widetilde{G}_3(\boldsymbol{\xi})$, shown as a surface response on four quadrants of prior support. (a) $p = 3$, (b) $p = 6$ and (c) $p = 9$.

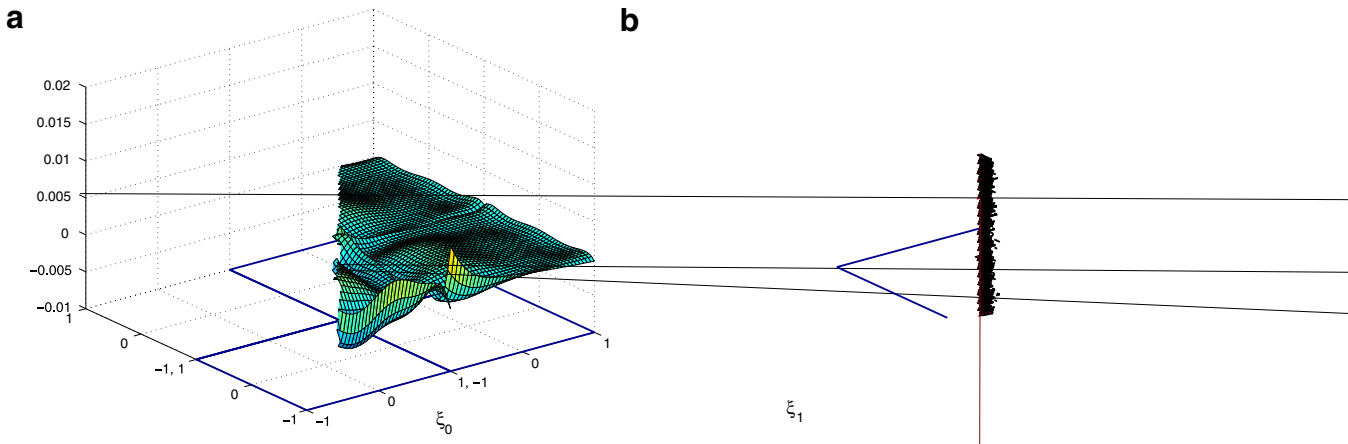


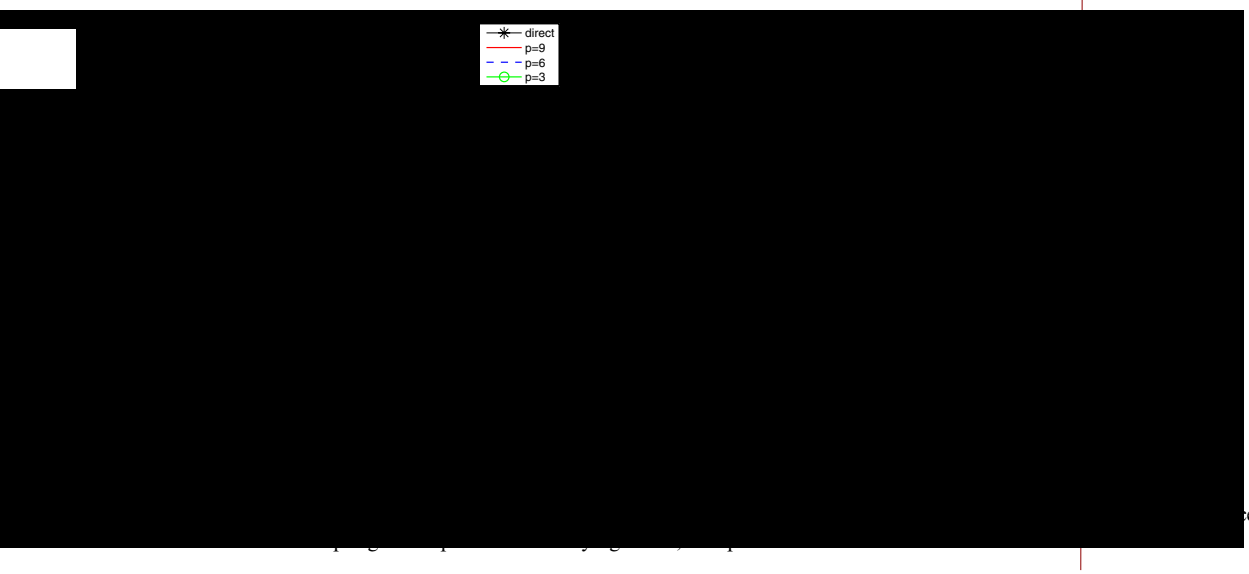
Fig. 5. $p = 6$ and (b) $p = 9$ in the solution to the stochastic forward problem, defined in

increasing order. But we also observe that the probability density in Fig. 5(a) is significantly broader than in Fig. 5(b). Under the prior uncertainty, the earlier-time measurement takes a wider range of values than the later-time measurement, and in this sense, the earlier measurement is more informative; it will allow the likelihood to discriminate more clearly among possible values of **m**. In the inverse problem setting, this information may be useful in choosing when and where to collect data.

3.3. Posterior densities

We now examine solutions of the inverse problem using polynomial chaos. A noisy data vector **d** is generated by solving the deterministic forward problem for a "true" model $\mathbf{m} = (0.25, 0.75)$, then perturbing the value of u at each sensor location/time with independent samples of Gaussian noise $\eta_i \sim N(0, \varsigma^2)$.

Fig. 6 shows contours of the posterior density conditioned on **d**. Solid lines are obtained via direct evaluations of the forward problem—i.e. they represent the posterior π_m in (22)—while dashed lines represent the posterior density computed with PC expansions on four partitions of the support of the prior. These are computed on $\boldsymbol{\xi}$-space, using (24), but since **g** is invertible it is simple to transform them back to **m**-space: $\tilde{\pi}_m \propto (\widetilde{L}_\xi \circ \mathbf{g}^{-1}) p_m$. Very close agreement between π_m and $\tilde{\pi}_m$ is observed with increasing order.



ce location.

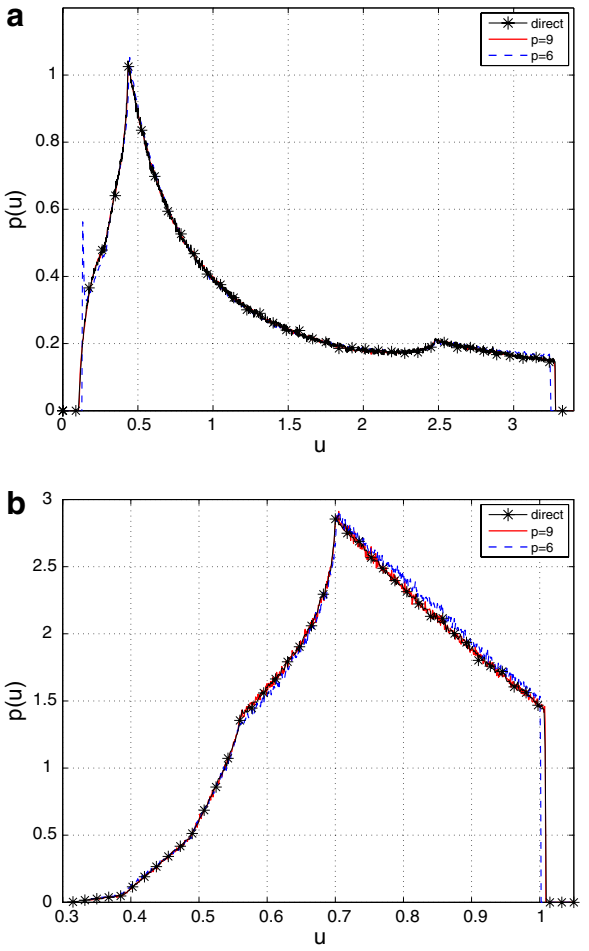


Fig. 5. Probability density of the scalar $u(x = 0.0, y = 0.0)$ at two successive times, given prior uncertainty in the source location. (a) $t = 0.05$ and (b) $t = 0.15$.

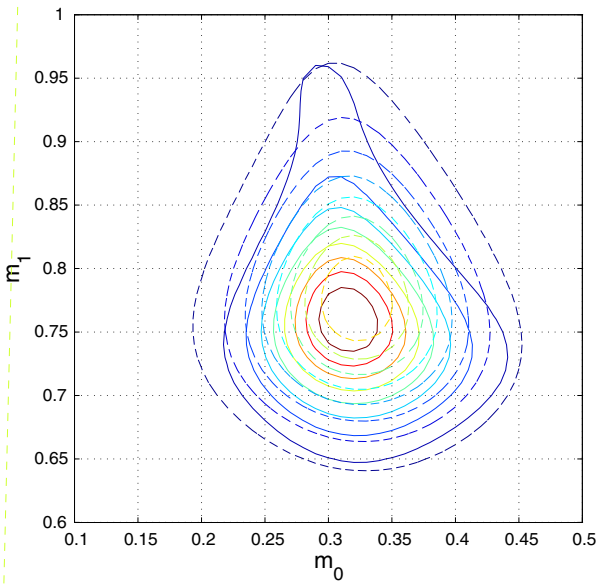
A quantitative assessment of the error in the posterior density is obtained by considering the Kullback–Leibler (KL) divergence of π_m from $\tilde{\pi}_m$:

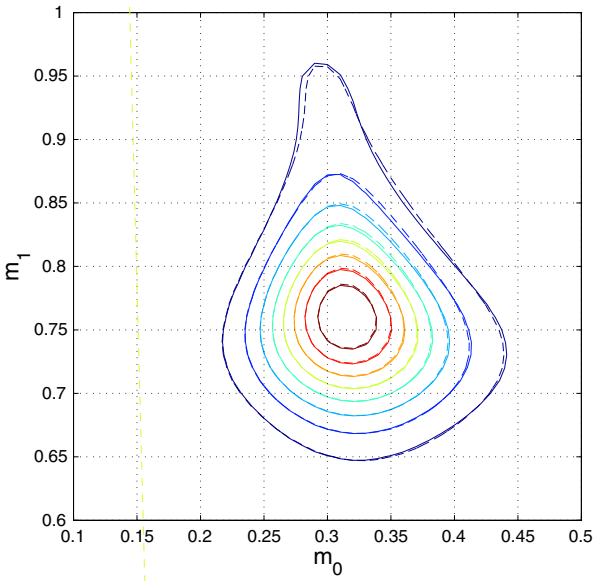
$$D(\tilde{\pi}\|\pi) = \int_{\mathcal{M}} \tilde{\pi}(\mathbf{m}) \log \frac{\tilde{\pi}(\mathbf{m})}{\pi(\mathbf{m})} \mathrm{d}\mathbf{m} \tag{34}$$

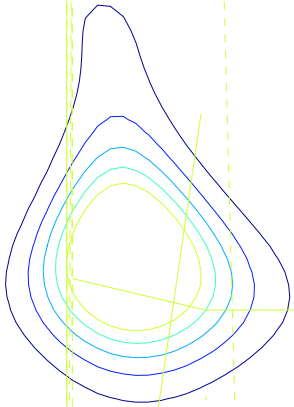
Fig. 7 plots $D(\tilde{\pi}_m\|\pi_m)$ for PC bases of increasing order p. In terms of KL divergence, we observe an exponential rate of convergence of the surrogate posterior $\tilde{\pi}_m$ to the true posterior.

3.4. Posterior sampling and speedup

Practical Bayesian computations must explore the posterior by sampling, and it is here that we expect the PC approach to achieve significant speedup over its direct counterpart. Fig. 8 shows the computational time for Monte Carlo estimation of the posterior mean as a function of the number of samples n. Since $g^b(\boldsymbol{\xi})$ is chosen to have density proportional to $p^b(\mathbf{m})$, our Monte Carlo estimator uses samples from the prior, as described in Section 2.3.1. In other words, the posterior mean is evaluated using (17), with $\mathbb{E}_\pi \mathbf{m} = \widehat{I}_n[\mathbf{m}]/\widehat{I}_n[1]$. We use a sixth-order PC basis for the stochastic spectral forward solution and compare the computational cost to that of direct sampling.







Speedup over the direct method is quite dramatic. The initial cost of the PC approach is offset by the computation of stochastic forward solutions, but then grows very slowly. Indeed, the per-sample cost is three orders of magnitude smaller for PC evaluations than for direct evaluations, and thus for even a moderate number of samples the gain in efficiency is significant. The cost of the initial stochastic forward solutions is recouped for $n \approx 200$, and thereafter the computational time of direct sampling rapidly eclipses that of the PC approach. For more complex forward models, the ratio of these per-sample costs may widen and the cost of the stochastic forward solutions may be recouped at even smaller n.

Another measure of speedup is to compare the computational times required to achieve a certain error in the Monte Carlo estimate of $I[\mathbf{m}]$. The estimate $\widehat{I}_n[f(\mathbf{m})]$ is a random variable with variance

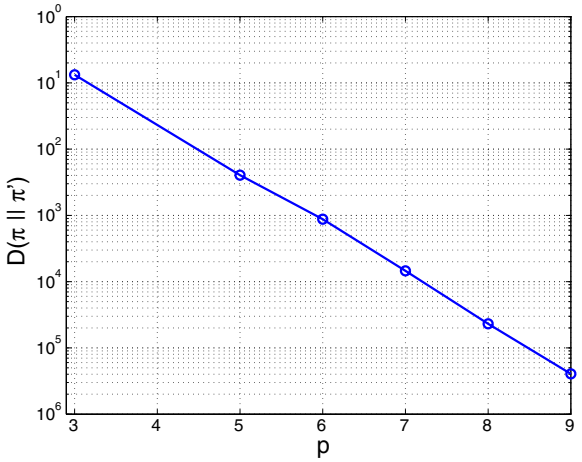


Fig. 7. Kullback–Leibler distance $D(\tilde{\pi}\|\pi)$ from the direct posterior to the PC-reformulated posterior, versus p.

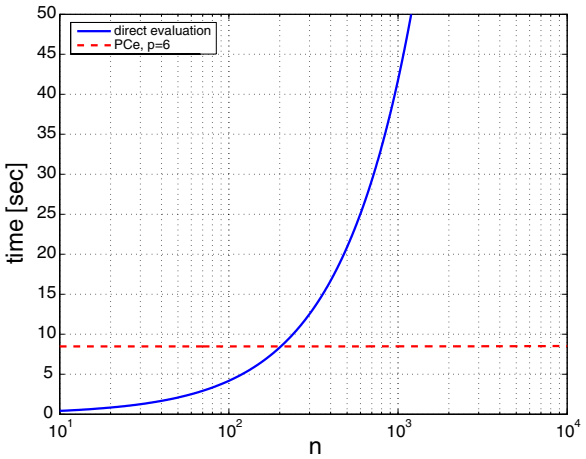


Fig. 8. Computational time for Monte Carlo simulation of the posterior mean; direct evaluation versus PCe.

$$\sigma^2\left(\widehat{I}_n[f(\mathbf{m})]\right) = \frac{1}{n}\mathrm{Var}_{\mathbf{m}\sim p_m}[f(\mathbf{m})L(\mathbf{m})] \tag{35}$$

where the subscript $\mathbf{m} \sim p_m$ reflects the drawing of samples from the prior distribution. In practice, we estimate the variance of $\widehat{I}_n$ from the Monte Carlo samples using the recursive formula given in [47]. The Monte Carlo "standard error" is then simply $\sigma(\widehat{I}_n[\mathbf{m}])$. Fig. 9 shows the error thus computed and normalized by $\widehat{I}_n[\mathbf{m}]$, versus computational time, for both the direct and PC approaches. Since error decreases as $n^{-1/2}$, the number of samples required to reduce the error grows rapidly. Because PC-obtained samples are inexpensive, however, very small relative errors are achievable at negligible cost. This certainly is not true for direct sampling, as the solid line in Fig. 9 indicates.

3.5. Markov chain Monte Carlo

Next we demonstrate the use of MCMC to simulate samples from the surrogate posterior $\tilde{\pi}_\xi$ given in (25). Since the $\mathbf{g}^b$ chosen above are invertible linear transformations from Ξ to $\mathscr{M}^b$, the conditions following the change of variables in (20) are satisfied. Therefore, the posterior expectation of any function f of $\mathbf{m}$ can be computed in $\boldsymbol{\xi}$-space, with samples from $\tilde{\pi}_\xi$, according to (23) and (26).

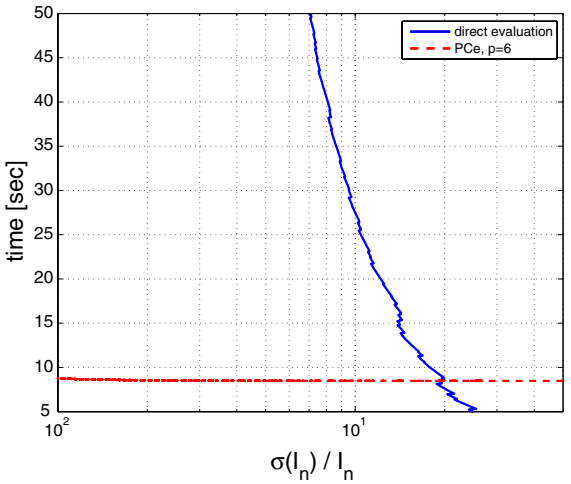


Fig. 9. Relative error in Monte Carlo estimates of the posterior mean versus computational time, sampling from the prior; direct evaluation versus PCe.

We employ a random-walk Metropolis algorithm [17] for MCMC, in which the proposal distribution $q(\cdot|\cdot)$ is a bivariate normal centered on the current position of the chain. The standard deviation of the proposal distribution is $\sigma_q = 0.4$. Results showing the chain position over 10,000 iterations are in Fig. 10. Visual inspection suggests that the chain mixes—i.e. moves within the support of $\tilde{\pi}_\xi$—rapidly. The two-dimensional view in Fig. 10(a) is reminiscent of the posterior contours in Fig. 6; this is not surprising, as $\boldsymbol{\xi}$ is just a diagonal linear transformation of $\mathbf{m}$.

A Metropolis–Hastings algorithm, such as the random-walk Metropolis sampler used here, provides for the construction of a Markov chain with stationary distribution $\tilde{\pi}_\xi$. Under certain additional conditions [15,63,49], one can establish a central limit theorem for ergodic averages $\bar{f}_n$ [64]:[2]

$$\bar{f}_n = \frac{1}{n}\sum_{j=1}^{n}(f \circ \mathbf{g})(\boldsymbol{\xi}^{(j)}) \tag{36}$$

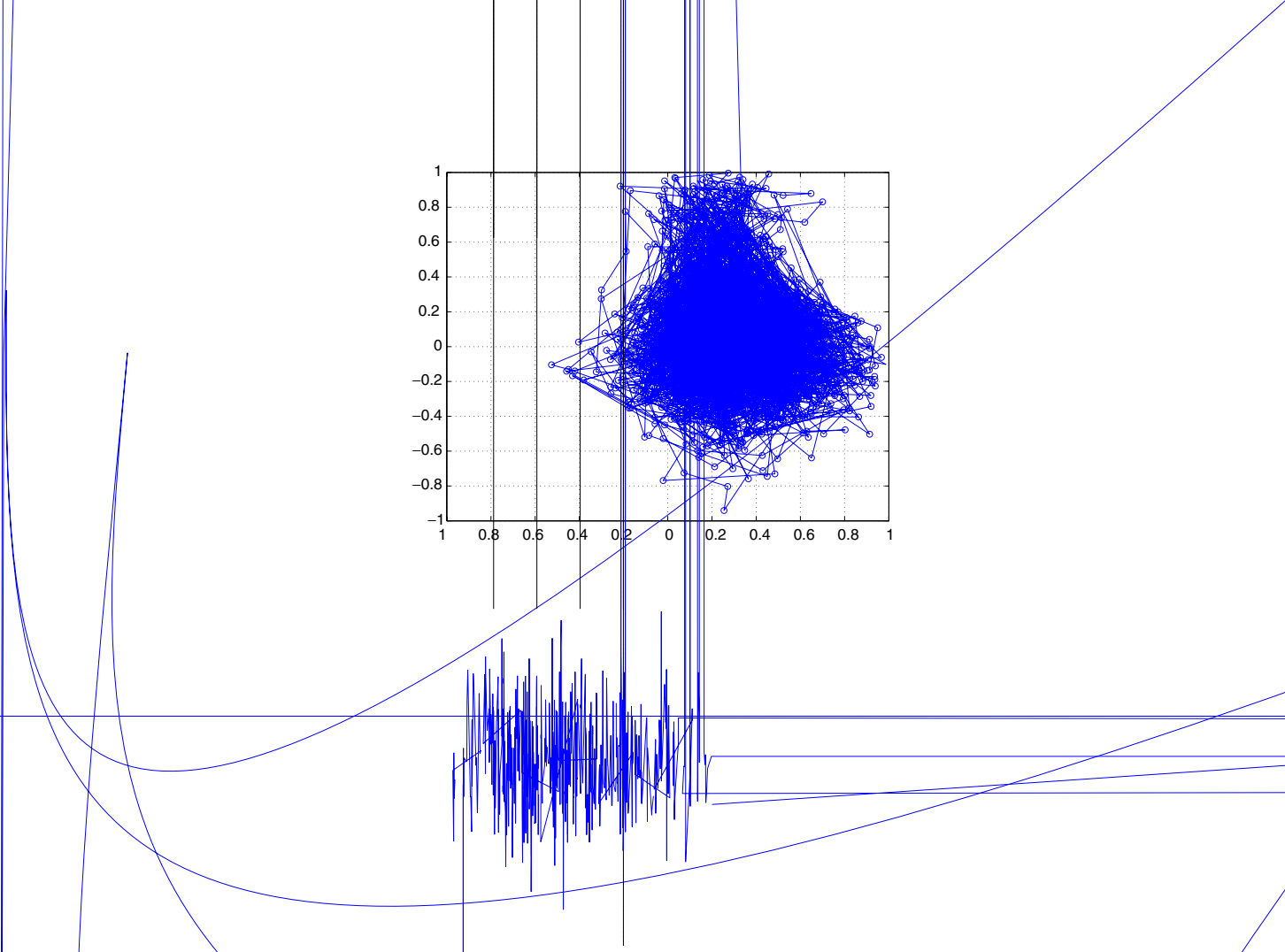
$$\sqrt{n}\left(\bar{f}_n - \mathbb{E}_{\pi_\xi}(f \circ \mathbf{g})\right) \xrightarrow{i.d.} N(0, \sigma_f^2) \tag{37}$$

where $\xrightarrow{i.d.}$ denotes convergence in distribution and

$$\sigma_f^2 = \mathrm{Var}_{\pi_\xi}(f \circ \mathbf{g}) + 2\sum_{s=1}^{\infty}\mathrm{Cov}_{\pi_\xi}[(f \circ \mathbf{g})(\boldsymbol{\xi}^{(0)}), (f \circ \mathbf{g})(\boldsymbol{\xi}^{(s)})] \tag{38}$$

The asymptotic variance in (38) thus reflects the correlation between successive samples. Even if a central limit theorem does not strictly hold in the form of (37), stronger correlations lead to larger variance of the MCMC estimate at a given n and thus less efficient sampling. We plot $\gamma(s)$, the empirical autocovariance at lag s, in Fig. 11 for several random-walk Metropolis samplers, varying the scale parameter of the proposal distribution σ_q. If σ_q is too large, a great proportion of the proposed moves will be rejected, and the chain will not move very often. If σ_q is too small, most proposed moves will be accepted but the chain will move very slowly through the posterior support. Both of these situations are reflected in long correlations and poor mixing. With $\sigma_q = 0.4$, however, we observe that autocovariance decays relatively quickly with lag along the chain,

[2] For chains on continuous state spaces, uniform or geometric ergodicity provide for central limit theorems [15,48]. Weaker conditions are sufficient to establish a law of large numbers. While uniform or geometric ergodicity have not been shown for Metropolis–Hastings samplers on general state spaces, many of these samplers are conjectured to be geometrically ergodic. The chain resulting from an independence sampler with bounded π/q is known to be uniformly ergodic [63].



consistent with the good mixing in Fig. 10. We also show, in black, the autocovariance of a chain resulting from an independence sampler [15], using the prior as a proposal distribution. This sampler also appears to be relatively efficient at exploring the simple posterior here.

An useful feature of MCMC estimatio cting marginal distributions for components of **m** or $\boldsymbol{\xi}$. This is performed with kernel den

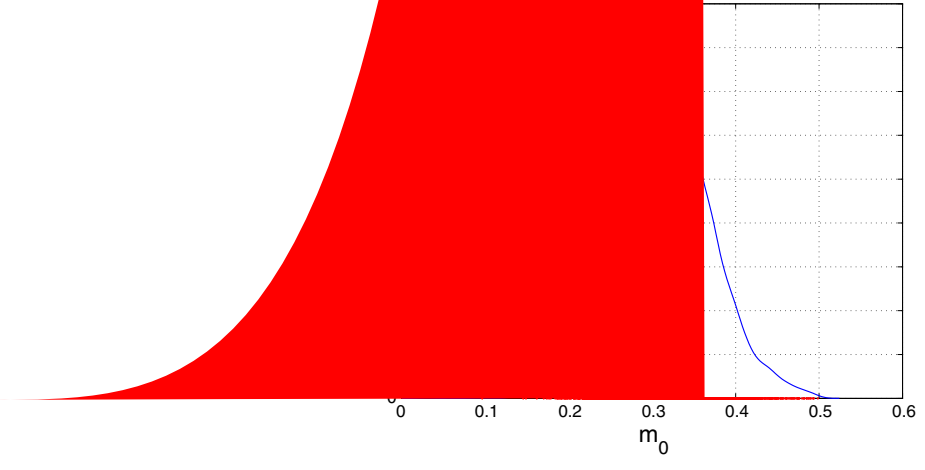
$$\pi(\xi_i) = \frac{1}{n-b} \sum_{j=b+1}^{n} K(\xi_i|\boldsymbol{\xi}^{(j)}) \tag{39}$$

where $K(\xi_i|\boldsymbol{\xi}^{(j)})$ is a density concentrat ere we use an one-dimensional Gaussian kernel, $K = N(\xi_i^{(j)}, \sigma_k^2)$ with bandwidth $\sigma_k =$ utions for the source coordinates, transformed back into **m**-space, are shown in Fig. are points at the bottom of the figure.

3.6. Choice of transformation and PC

Using MCMC to explore the pos onsiderable freedom in the choice of **g**, and further, in the choice of PC basis. All t e a diffeomorphism from a set Ξ_m to the support of the prior, where $\Xi_m \subseteq \Xi$ and Ξ is) is used to solve the stochastic forward problem on the chosen PC basis and an exp posterior density $\tilde{\pi}$ is formed.

It is reasonable, however, to exp influence the accuracy and cost of evaluating the posterior distribution. Increasing is certainly improves the accuracy of forward



problem solutions and reduces errors in the surrogate posterior, as we observed in Sections 3.2 and 3.3. But we must also address the larger question—essentially, what is the uncertainty that one should propagate through the forward problem?

We explore the impact of different transformations $\mathbf{g}$ and PC bases using the two-dimensional source inversion problem as before. We consider three "true" source locations, ranging from the center to the edge of the domain: $\mathbf{m} = (0.50, 0.50)$; $\mathbf{m} = (0.25, 0.75)$; and $\mathbf{m} = (0.10, 0.90)$. For each of these sources, we solve the deterministic forward problem, then perturb the value of u at each sensor with independent samples of Gaussian noise $\eta_i \sim N(0, \varsigma^2)$, thus generating three noisy data vectors $\mathbf{d}$ for inference. The sensor locations/times, the source strength and shutoff time, and the prior p_m are unchanged from previous sections.

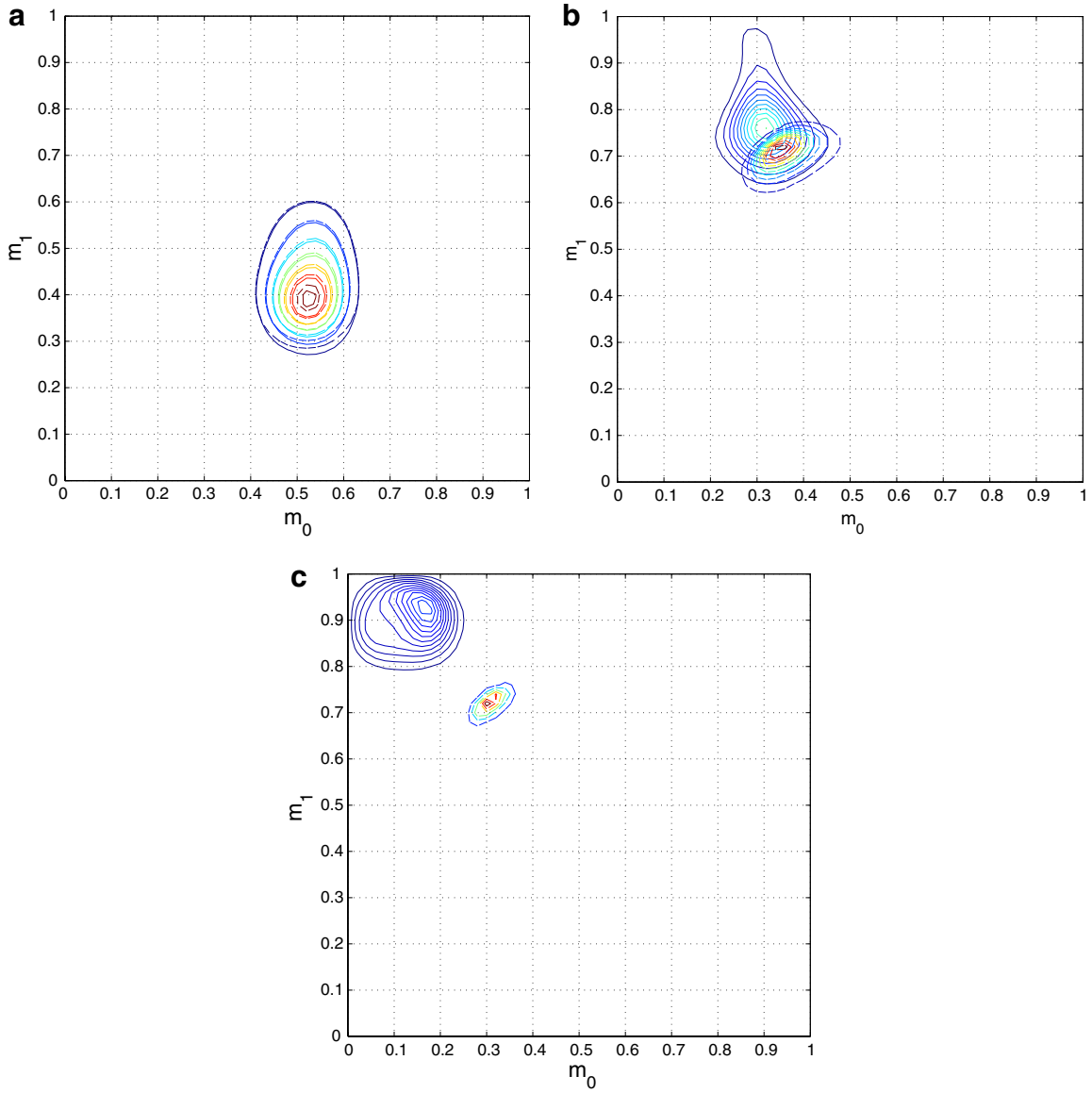


Fig. 13. Contours of posterior density. Gauss-Hermite PC, $p = 6$; $\sigma = 10^{-2}$. Solid lines are obtained with direct evaluations of the forward problem; dashed lines are obtained with polynomial chaos expansions. (a) $\mathbf{m}_{\text{true}} = (0.50, 0.50)$, (b) $\mathbf{m}_{\text{true}} = (0.25, 0.75)$ and (c) $\mathbf{m}_{\text{true}} = (0.10, 0.90)$.

For simplicity, we do not partition the prior support; i.e. we choose $b = 1$. To solve the inverse problem with polynomial chaos, we thus prescribe a single set of PC expansions $\mathbf{m} = \mathbf{g}(\boldsymbol{\xi})$. First, consider a Gauss-Hermite PC basis: bivariate Hermite polynomials $\Psi_k(\boldsymbol{\xi})$ with $\xi_i \sim N(0,1)$. We prescribe a PC expansion centered on $\mathbf{m} = (0.5, 0.5)$:

$$g_i(\boldsymbol{\xi}) = 0.5 + \sigma\xi_i; \quad i = 1, 2 \tag{40}$$

Since $\Xi = \mathbb{R}^2$, the inverse image of the support of the prior will be contained in Ξ for any σ.

Contours of the surrogate posterior density $\tilde{\pi}_m$ are shown in Fig. 13 for $\sigma = 10^{-2}$ and a sixth-order PC basis. These are compared to the "direct" posterior π_m for each source location. While agreement of the posteriors is relatively good for a source in the center of the domain (Fig. 13(a)), the accuracy of the surrogate

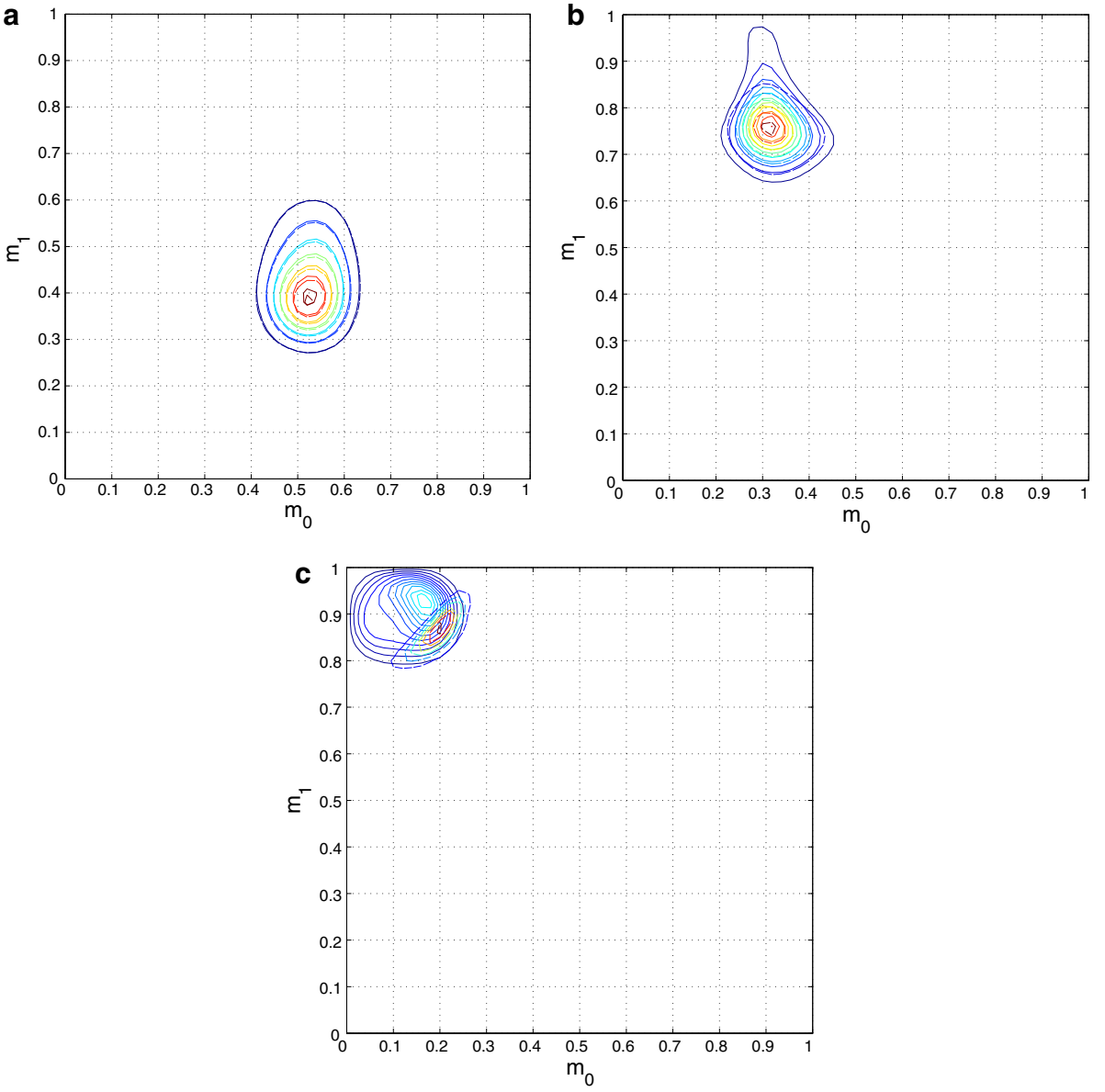


Fig. 14. Contours of posterior density. Gauss-Hermite PC, $p = 6$; $\sigma = 10^{-1}$. Solid lines are obtained with direct evaluations of the forward problem; dashed lines are obtained with polynomial chaos expansions. (a) $\mathbf{m}_{\text{true}} = (0.50, 0.50)$, (b) $\mathbf{m}_{\text{true}} = (0.25, 0.75)$ and (c) $\mathbf{m}_{\text{true}} = (0.10, 0.90)$.

posterior deteriorates rapidly as the source moves towards the upper-left corner. For $\mathbf{m}_{\text{true}} = (0.10, 0.90)$, the surrogate posterior is centered on an entirely different area of $\mathcal{M}$ than the direct posterior. This disagreement explicitly reflects errors in the forward problem solution $\tilde{G}(\boldsymbol{\xi})$ for $\boldsymbol{\xi}$ corresponding to values of $\mathbf{m}$ that are close to the boundary—i.e. for $|\boldsymbol{\xi}| = \mathcal{O}(1/\sigma)$.

Widening the density of $\mathbf{g}(\boldsymbol{\xi})$ improves agreements significantly. Fig. 14 shows posterior contours for $\sigma = 10^{-1}$ and a sixth-order PC basis. Once again, the best agreement is obtained when the source is in the center of the domain, but reasonable overlap of the posteriors is achieved even for $\mathbf{m}_{\text{true}} = (0.25, 0.75)$. The mean of the surrogate posterior for $\mathbf{m}_{\text{true}} = (0.10, 0.90)$, while still misplaced, shows some improvement over the $\sigma = 10^{-2}$ case. Increasing the order of the PC basis to $p = 9$ sharpens agreement for all three source locations,

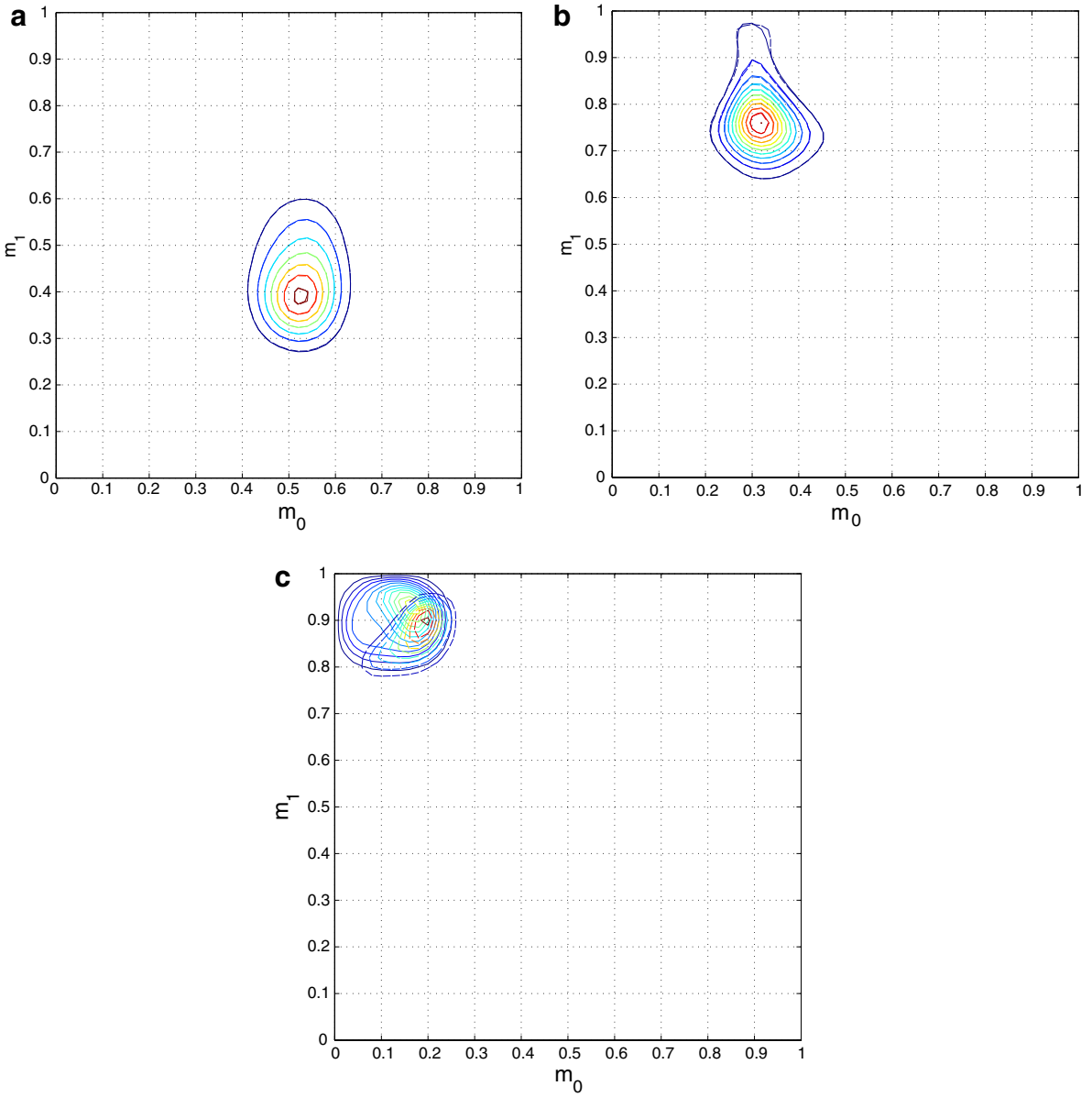


Fig. 15. Contours of posterior density. Gauss-Hermite PC, $p = 9$; $\sigma = 10^{-1}$. Solid lines are obtained with direct evaluations of the forward problem; dashed lines are obtained with polynomial chaos expansions. (a) $\mathbf{m}_{\text{true}} = (0.50, 0.50)$, (b) $\mathbf{m}_{\text{true}} = (0.25, 0.75)$ and (c) $\mathbf{m}_{\text{true}} = (0.10, 0.90)$.

as observed in Fig. 15. However, the key trend—deteriorating agreement as the source moves closer to the boundary—is preserved.

It is instructive to compare these results to those obtained with a uniform-Legendre PC basis. We again use a single partition of the prior support, with a PC expansion centered on $\mathbf{m} = (0.5,\ 0.5)$: $g_i = 0.5 + 0.5\xi_i$, $\xi_i \sim U(-1, 1)$. Results with $p = 9$ are shown in Fig. 16. Compared to the Gauss-Hermite basis at the same order, we observe slightly poorer agreement for sources in the center of the domain, but a more consistent level of error as the source is moved towards the boundary. Agreement of the surrogate and direct posteriors for $\mathbf{m}_{\text{true}} = (0.10,\ 0.90)$ is substantially better with uniform-Legendre PC than with any of the Gauss-Hermite bases.

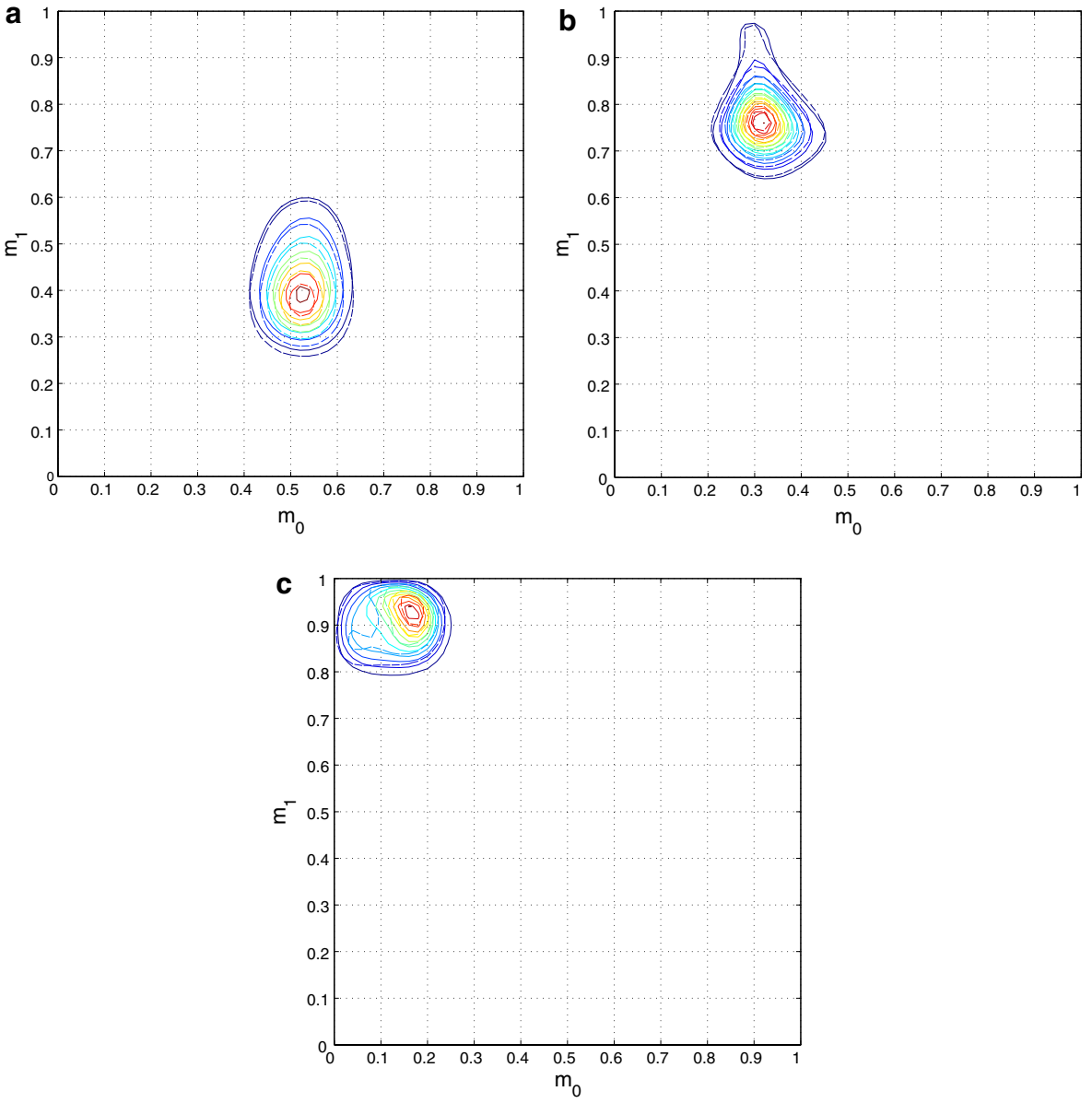


Fig. 16. Contours of posterior density. Uniform-Legendre PC, $p = 9$. Solid lines are obtained with direct evaluations of the forward problem; dashed lines are obtained with polynomial chaos expansions. (a) $\mathbf{m}_{\text{true}} = (0.50,\ 0.50)$, (b) $\mathbf{m}_{\text{true}} = (0.25,\ 0.75)$ and (c) $\mathbf{m}_{\text{true}} = (0.10,\ 0.90)$.

An explanation for these results may be found in the distribution of $\boldsymbol{\xi}$, which is of course a defining property of the PC basis. As noted above, error in the surrogate posterior at a particular value of $\mathbf{m}$ reflects error in the stochastic forward problem solution for $\boldsymbol{\xi} = \mathbf{g}^{-1}(\mathbf{m})$. Where is this error likely to occur? If, for simplicity, one ignores aliasing errors resulting from the pseudospectral construction [21], Galerkin projection of the forward problem outputs onto the PC basis minimizes $\|\mathbf{G}(\mathbf{g}(\boldsymbol{\xi})) - \widetilde{\mathbf{G}}(\boldsymbol{\xi})\|_{L^2(\Omega)}$, where this inner-product norm is defined by the probability measure P on $(\Omega, \mathscr{U})$. Let $\boldsymbol{\xi}$ and P induce a probability distribution on $\mathbb{R}^n$. The Gaussian distribution weighs errors near the origin much more strongly than errors at large $|\boldsymbol{\xi}|$, whereas the uniform distribution weighs errors equally over the entire (finite) range of $\boldsymbol{\xi}$.

This weighing is consistent with the error trends in Figs. 13–16, and its impact is described more extensively in Fig. 17. Here we compute the Kullback–Leibler divergence of π from $\tilde{\pi}$ for Gauss-Hermite bases of order $p = 6$ and 9, while varying the scale parameter σ in (40). Since $g_i(\boldsymbol{\xi}) \sim N(0.5, \sigma^2)$, the scale parameter controls the standard deviation of the input uncertainty to the forward problem. The Gauss-Hermite results exhibit a common dependence on σ for all three source locations. Posterior errors increase at small σ and large σ, but are minimized at intermediate values of σ, e.g. $\sigma = 10^{-1}$. If σ is very small, the input distribution is narrowly centered on (0.5, 0.5) and the posterior distribution favors values of $\mathbf{m} = \mathbf{g}(\boldsymbol{\xi})$ that lie on the edges of this input distribution. Errors in the forward solution at these values of $\boldsymbol{\xi}$ receive little weight in the $L^2(\Omega)$ norm and thus

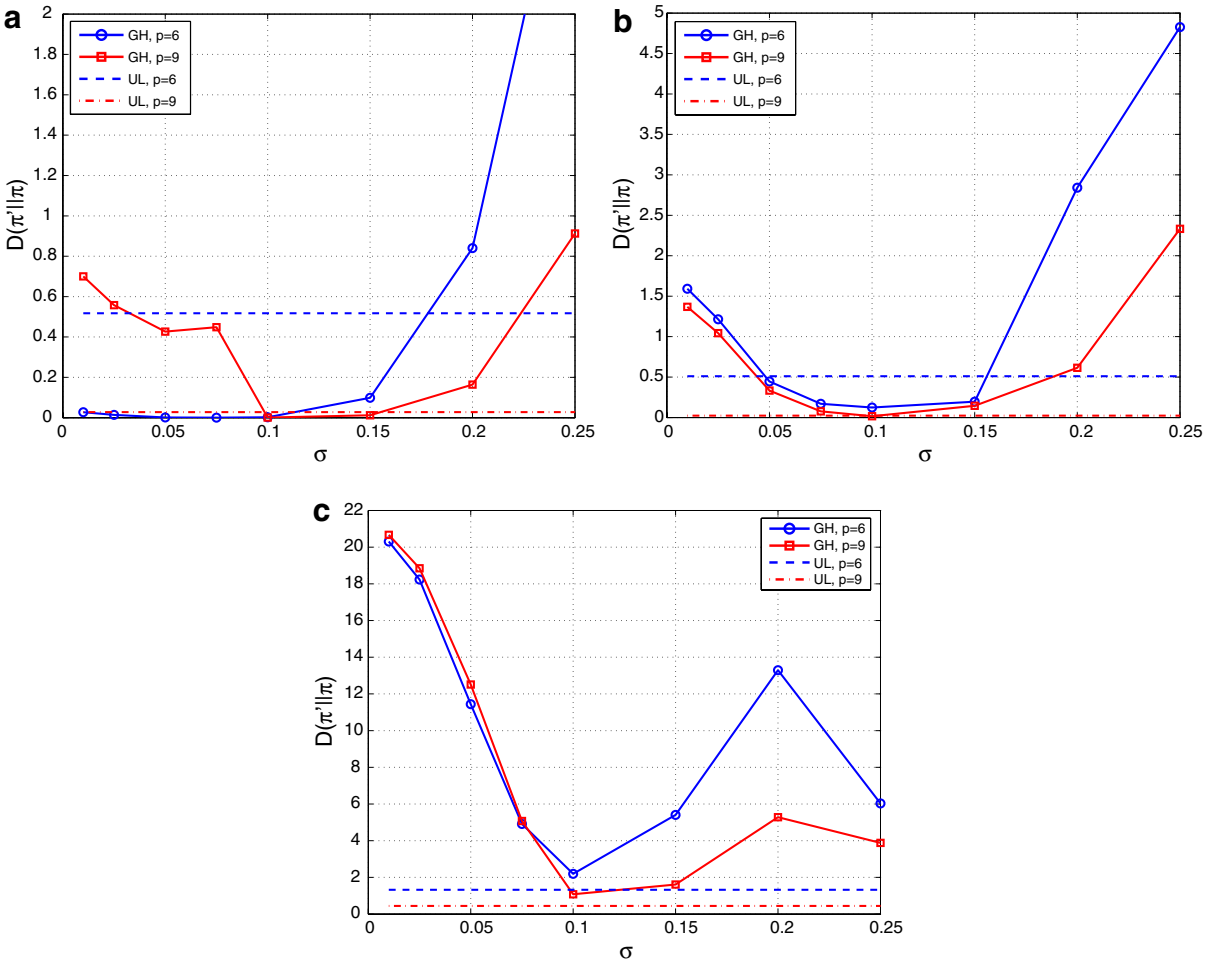


Fig. 17. $D(\tilde{\pi}\|\pi)$ with Gauss-Hermite and uniform-Legendre PC at varying orders and source locations. Scale parameter σ is the standard deviation of input uncertainty to the forward problem in the Gauss-Hermite case. (a) $\mathbf{m}_{\text{true}} = (0.50, 0.50)$, (b) $\mathbf{m}_{\text{true}} = (0.25, 0.75)$ and (c) $\mathbf{m}_{\text{true}} = (0.10, 0.90)$.

lead to errors in the posterior. On the other hand, with large σ (e.g. $\sigma > 0.1$) the input distribution of $\mathbf{g}(\boldsymbol{\xi})$ broadens to include appreciable tails outside the square domain. While values of $\boldsymbol{\chi} = \mathbf{m}$ outside the unit square S are perfectly feasible according to (30), the $L^2(\Omega)$ norm then penalizes errors in the stochastic forward solution for these values at the expense of errors inside the domain. As a result, posterior errors again increase.

In general, $p = 9$ results show smaller errors than those at $p = 6$. The errors at $p = 9$ and small σ with $\mathbf{m}_{\text{true}} = (0.50,\ 0.50)$ appear to be an exception, reflecting errors at very large $|\boldsymbol{\xi}|$. We find that the ratio of the $p = 9$ and $p = 6$ posterior divergences at a given σ varies with the realization of the measurement noise, and thus the trends in Fig. 17 should be generalized to other data in a qualitative sense.

Fig. 17 also compares Gauss-Hermite posterior divergences to those obtained with uniform-Legendre bases of the same order—shown as horizontal lines, since we fix $g_i(\boldsymbol{\xi}) \sim U(0,1)$ in the uniform-Legendre case. Again, these results show strong dependence on the source location. For a source (and thus a posterior density) in the center of the domain, it is possible to obtain lower posterior divergences with Gauss-Hermite PC at a given order than with uniform-Legendre PC. With an appropriately-chosen scale parameter, this is again possible for $\mathbf{m}_{\text{true}} = (0.25,\ 0.75)$ at $p = 6$, but the uniform-Legendre basis proves more accurate at higher order ($p = 9$). And for a source near the edge of the domain, the uniform-Legendre basis provides better accuracy than the Gauss-Hermite basis at all values of the scale parameter and both orders.

Implications of this exercise on the choice of transformation $\mathbf{g}(\boldsymbol{\xi})$ and on the choice of PC basis are as follows. One should avoid transformations or bases for which "true" value(s) of $\mathbf{m}$ correspond to values of $\boldsymbol{\xi}$ that have small probability. The true values of $\mathbf{m}$ are of course unknown *a priori*, so an useful guideline is to ensure that $\mathbf{g}(\boldsymbol{\xi})$ has distribution equal to the prior, approximately equal to the prior, or with longer tails than the prior. Then the posterior probability mass is likely to fall within regions where forward problem errors are penalized with sufficient weight in the $L^2(\Omega)$ norm. Of course, it is possible for very large quantities of data to overwhelm the prior. In this case, if it is apparent that the posterior is tending towards regions of $\boldsymbol{\xi}$ that have been accorded small probability—as in Figs. 13(c) and 14(c), for instance—a new choice of $\mathbf{g}$ or PC basis would be indicated.

4. Conclusions

Bayesian inference provides an attractive setting for the solution of inverse problems. Measurement errors, forward model uncertainties, and complex prior information can all be combined to yield a rigorous and quantitative assessment of uncertainty in the inverse solution. Obtaining useful information from this posterior density—e.g., computing expectations or marginal distributions of the unknown parameters—may be a computationally expensive undertaking, however. For complex forward models, such as those that arise in inverting systems of PDEs, the cost of likelihood evaluations may render the Bayesian approach prohibitive.

The theoretical developments in this paper fundamentally accelerate Bayesian inference in computationally intensive inverse problems. We present a reformulation of the Bayesian approach based on polynomial chaos representations of random variables and associated spectral methods for efficient uncertainty propagation. Uncertain inputs that span the range of the prior define a stochastic forward problem; a Galerkin solution of this problem with the PC basis yields a spectral representation of uncertain forward model predictions. Evaluation of integrals over the unknown parameter space is then recast as sampling of the random variables $\boldsymbol{\xi}$ underlying the PC expansion, with significant speedup. In particular, we introduce three schemes for exploring the posterior: Monte Carlo sampling from the prior distribution, Monte Carlo sampling from an alternate distribution that includes the support of the prior, and Markov chain Monte Carlo in $\boldsymbol{\xi}$-space. Each of these schemes is compatible with partitioning of the prior support.

The new approach is demonstrated on a transient diffusion problem arising in contaminant source inversion. Spectral representation is found to reduce the cost of each posterior evaluation by three orders of magnitude, so that sampling of the PC-reformulated problem has nearly negligible cost. Error in the surrogate posterior decreases rapidly with increasing order of the PC basis; in the present case, convergence is exponentially fast. MCMC sampling of the posterior offers considerable freedom in choosing the PC basis and the initial transformation defining the stochastic forward problem, but a detailed exploration of posterior errors suggests guidelines for ensuring accuracy and computational efficiency.

Ongoing work will extend the polynomial chaos approach to more complex inverse problems. For instance, forward models with additional parametric uncertainty—parameters that may be marginalized in the posterior—should be quite amenable to PC acceleration. We also plan to explore stochastic spectral approaches to significantly higher-dimensional inverse problems, e.g. with spatially extended input parameters. A further extension involves convective source inversion problems, with the associated challenges of spectral uncertainty propagation in non-linear advection equations.

Acknowledgments

This research was supported in part by an appointment to the Sandia National Laboratories Truman Fellowship in National Security Science and Engineering, sponsored by Sandia Corporation (a wholly owned subsidiary of Lockheed Martin Corporation) as operator of Sandia National Laboratories under US Department of Energy Contract No. DE-AC04-94AL85000. Support was also provided by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences.

References

[1] S.N. Evans, P.B. Stark, Inverse problems as statistics, Inverse Problems 18 (2002) R55–R97.
[2] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer, 2005.
[3] A. Mohammad-Djafari, Bayesian inference for inverse problems, in: Bayesian inference and Maximum Entropy Methods in Science and Engineering, vol. 21, 2002, pp. 477–496.
[4] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, 2005.
[5] R. Aster, B. Borchers, C. Thurber, Parameter Estimation and Inverse Problems, Academic Press, 2004.
[6] W.P. Gouveia, J.A. Scales, Resolution of seismic waveform inversion: Bayes versus Occam, Inverse Problems 13 (1997) 323–349.
[7] A. Malinverno, Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, Geophysical Journal International 151 (2002) 675–688.
[8] C. Jackson, M.K. Sen, P.L. Stoffa, An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions, Journal of Climate 17 (2004) 2828–2841.
[9] J. Wang, N. Zabaras, Hierarchical Bayesian models for inverse problems in heat conduction, Inverse Problems 21 (2005) 183–206.
[10] J. Wang, N. Zabaras, Using Bayesian statistics in the estimation of heat source in radiation, International Journal of Heat and Mass Transfer 48 (2005) 15–29.
[11] T. Loredo, Computational technology for Bayesian inference, in: D.M.R. Plante, D. Roberts (Eds.), ASP Conf. Ser., Astronomical Data Analysis Software and Systems VIII, vol. 172, 1999.
[12] K. Mosegaard, M. Sambridge, Monte Carlo analysis of inverse problems, Inverse Problems 18 (2002) R29–R54.
[13] M. Evans, T. Swartz, Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems, Statistical Science 10 (3) (1995) 254–272.
[14] R. Cools, P. Dellaportas, The role of embedded integration rules in Bayesian statistics, Statistics and Computing 6 (1996) 245–250.
[15] L. Tierney, Markov chains for exploring posterior distributions, The Annals of Statistics 22 (4) (1994) 1701–1728.
[16] J. Besag, P. Green, D. Higdon, K. Mengersen, Bayesian computation and stochastic systems, Statistical Science 10 (1) (1995) 3–41.
[17] W.R. Gilks, S. Richardson, D.J. Spiegelhalter (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, 1996.
[18] D. Higdon, H. Lee, C. Holloman, Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems, Bayesian Statistics 7 (2003) 181–197.
[19] R. Ghanem, P. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer Verlag, New York, 1991.
[20] O. Le Maître, O. Knio, H. Najm, R. Ghanem, A stochastic projection method for fluid flow I. Basic formulation, Journal of Computational Physics 173 (2001) 481–511.
[21] B. Debusschere, H. Najm, P. Pébay, O. Knio, R. Ghanem, O. Le Maître, Numerical challenges in the use of polynomial chaos representations for stochastic processes, SIAM Journal of Scientific Computing 26 (2) (2004) 698–719.
[22] N. Wiener, The homogenous chaos, American Journal of Mathematics 60 (1938) 897–936.
[23] A. Chorin, Hermite expansions in Monte Carlo computation, Journal of Computational Physics 8 (1971) 472–482.
[24] F. Maltz, D. Hitzl, Variance reduction in Monte Carlo computations using multi-dimensional Hermite polynomials, Journal of Computational Physics 32 (1979) 345–376.
[25] W. Meecham, D. Jeng, Use of the Wiener-Hermite expansion for nearly normal turbulence, Journal of Fluid Mechanics 32 (1968) 225–249.
[26] A. Chorin, Gaussian fields and random flow, Journal of Fluid Mechanics 63 (1974) 21–32.
[27] R. Cameron, W. Martin, The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals, Annals of Mathematics 48 (1947) 385–392.

[28] R. Ghanem, Probabilistic characterization of transport in heterogeneous media, Computational Methods in Applied Mechanics and Engineering 158 (1998) 199–220.
[29] R. Ghanem, Ingredients for a general purpose stochastic finite element formulation, Computational Methods in Applied Mechanics and Engineering 168 (1999) 19–34.
[30] R. Ghanem, Stochastic finite elements for heterogeneous media with multiple random non-gaussian properties, ASCE Journal of Engineering Mechanics 125 (1999) 26–40.
[31] R. Ghanem, J. Red-Horse, A. Sarkar, Modal properties of a space-frame with localized system uncertainties, in: Proceedings of the 8th ASCE Specialty Conference of Probabilistic Mechanics and Structural Reliability, No. PMC200-269, ASCE, 2000.
[32] O. Le Maître, M. Reagan, H. Najm, R. Ghanem, O. Knio, A stochastic projection method for fluid flow II. Random process, Journal of Computational Physics 181 (2002) 9–44.
[33] D. Xiu, D. Lucor, C.-H. Su, G. Karniadakis, Stochastic modeling of flow-structure interactions using generalized polynomial chaos, ASME Journal of Fluids Engineering 124 (2002) 51–59.
[34] D. Xiu, G. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM Journal of Scientific Computing 24 (2) (2002) 619–644.
[35] D. Xiu, G. Karniadakis, Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos, Computer Methods in Applied Mechanics and Engineering 191 (2002) 4927–4948.
[36] B. Debusschere, H. Najm, A. Matta, O. Knio, R. Ghanem, O. Le Maître, Protein labeling reactions in electrochemical microchannel flow: Numerical simulation and uncertainty propagation, Physics of Fluids 15 (8) (2003) 2238–2250.
[37] G. Berkooz, P. Holmes, J.L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, Annual Review of Fluid Mechanics 25 (1993) 539–575.
[38] L. Sirovich, Turbulence and the dynamics of coherent structures. Part 1: Coherent structures, Quarterly of Applied Mathematics 45 (3) (1987) 561–571.
[39] S. Balakrishnan, A. Roy, M.G. Ierapetritou, G.P. Flach, P.G. Georgopoulos, Uncertainty reduction and characterization for complex environmental fate and transport models: an empirical Bayesian framework incorporating the stochastic response surface method, Water Resources Research 39 (12) (2003) 1350.
[40] J.A. Christen, C. Fox, MCMC using an approximation, Journal of Computer and Graphical Statistics 14 (4) (2005) 795–810.
[41] C.J. Geyer, Markov chain Monte Carlo maximum likelihood, in: E.M. Keramidas (Ed.), Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, vol. 23, Interface Foundation of North America, 1991, pp. 156–163.
[42] D.J.C. MacKay, Comparison of approximate methods for handling hyperparameters, Neural Computation 11 (5) (1999) 1035–1068.
[43] S. Kakutani, Spectral analysis of stationary Gaussian processes, in: J. Neyman (Ed.), Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, University of California Press, 1961, pp. 239–247.
[44] W. Schoutens, Stochastic Processes and Orthogonal Polynomials, Springer, 2000.
[45] M. Reagan, H. Najm, R. Ghanem, O. Knio, Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection, Combustion and Flame 132 (2003) 545–555.
[46] C. Andrieu, N. de Freitas, A. Doucet, M.I. Jordan, An introduction to MCMC for machine learning, Machine Learning 50 (2003) 5–43.
[47] G.S. Fishman, Monte Carlo: Concepts, Algorithms, and Applications, Springer, 1996.
[48] D. Gamerman, Markov chain Monte Carlo: Stochastic simulation for Bayesian inference, Chapman & Hall, 1997.
[49] G.O. Roberts, Markov chain concepts related to sampling algorithms, in: W.R. Gilks, S. Richardson, D.J. Spiegelhalter (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, 1996, pp. 45–58.
[50] A.E. Gelfand, A.F.M. Smith, Sampling-based approaches to calculating marginal densities, Journal of the American Statistical Association 85 (1990) 398–409.
[51] O. Le Maître, R. Ghanem, O. Knio, H. Najm, Uncertainty propagation using Wiener–Haar expansions, Journal of Computational Physics 197 (1) (2004) 28–57.
[52] O. Le Maître, H. Najm, R. Ghanem, O. Knio, Multi-resolution analysis of Wiener-type uncertainty propagation schemes, Journal of Computational Physics 197 (2004) 502–531.
[53] X. Wan, G.E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, Journal of Computational Physics 209 (2005) 617–642.
[54] V. Akcelik, G. Biros, O. Ghattas, K. Long, B. van Bloemen Waanders, A variational finite element method for source inversion for convective-diffusive transport, Finite Elements in Analysis and Design 39 (2003) 683–705.
[55] U. Shavit, A. Furman, The location of deep salinity sources in the Israeli coastal aquifer, Journal of Hydrology 250 (2001) 63–77.
[56] J.V. Beck, B. Blackwell, C.R.S. Clair, Inverse Heat Conduction: Ill-Posed Problems, Wiley, 1985.
[57] O.M. Alifanov, Inverse Heat Transfer Problems, Springer, 1994.
[58] C.H. Lai, C.S. Ierotheou, C.J. Palansuriya, K.A. Pericleous, M.S. Espedal, X.C. Tai, Accuracy of a domain decomposition method for the recovering of discontinuous heat sources in metal sheet cutting, Computing and Visualization in Science 2 (1999) 149–152.
[59] V.A. Karkhin, V.V. Plochikhine, H.W. Bergmann, Solution of inverse heat conduction problem for determining heat input, weld shape, and grain structure during laser welding, Science and Technology of Welding and Joining 7 (4) (2002) 224–231.
[60] J. Wang, N. Zabaras, A Bayesian inference approach to the inverse heat conduction problem, International Journal of Heat and Mass Transfer 47 (2004) 3927–3941.
[61] B. Sommeijer, L. Shampine, J. Verwer, RKC: An explicit solver for parabolic PDEs, Journal of Computational and Applied Mathematics 88 (1997) 315–326.

[62] J.G. Verwer, B.P. Sommeijer, W. Hundsdorfer, RKC time-stepping for advection-diffusion-reaction problems, CWI Report MAS-E0405, CWI, Amsterdam, The Netherlands, 2004.
[63] L. Tierney, Introduction to general state-space Markov chain theory, in: W.R. Gilks, S. Richardson, D.J. Spiegelhalter (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, 1996, pp. 59–74.
[64] C.J. Geyer, Practical Markov chain Monte Carlo, Statistical Science 7 (4) (1992) 473–483.