Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems

Youssef M. Marzouk [a,*], Habib N. Najm [b]

[a] Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[b] Sandia National Laboratories, Livermore, CA 94551, USA

ABSTRACT

We consider a Bayesian approach to nonlinear inverse problems in which the unknown quantity is a spatial or temporal field, endowed with a hierarchical Gaussian process prior. Computational challenges in this construction arise from the need for repeated evaluations of the forward model (e.g., in the context of Markov chain Monte Carlo) and are compounded by high dimensionality of the posterior. We address these challenges by introducing truncated Karhunen–Loève expansions, based on the prior distribution, to efficiently parameterize the unknown field and to specify a stochastic forward problem whose solution captures that of the deterministic forward model over the support of the prior. We seek a solution of this problem using Galerkin projection on a polynomial chaos basis, and use the solution to construct a reduced-dimensionality surrogate posterior density that is inexpensive to evaluate. We demonstrate the formulation on a transient diffusion equation with prescribed source terms, inferring the spatially-varying diffusivity of the medium from limited and noisy data.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Inverse problems arise from indirect observations of a quantity of interest. Observations may be limited in number relative to the dimension or complexity of the model space, and the action of the forward operator may include filtering or smoothing effects. These features typically render inverse problems ill-posed—in the sense that no solution may exist, multiple solutions may exist, or solutions may not depend continuously on the data. In practical settings, where observations are inevitably corrupted by noise, this presents numerous challenges.

Classical approaches to inverse problems have used regularization methods to impose well-posedness, and solved the resulting deterministic problems by optimization or other means [1]. However, important insights and methodologies emerge by casting inverse problems in the framework of statistical inference [2,3]. Here we focus on Bayesian approaches, which provide a foundation for inference from noisy and limited data, a natural mechanism for regularization in the form of prior information, and in very general cases—e.g., nonlinear forward operators, non-Gaussian errors—a quantitative assessment of uncertainty in the results [4,5]. Indeed, the output of Bayesian inference is not a single value for the quantity of interest, but a probability distribution that summarizes all available information about this quantity, be it a vector of parameters or a function (i.e., a signal or spatial field). Exploration of this posterior distribution—and thus estimating means, higher

---

* Corresponding author. Tel.: +1 617 253 1337.
  E-mail addresses: ymarz@mit.edu (Y.M. Marzouk), hnnajm@sandia.gov (H.N. Najm).

moments, and marginal densities of the inverse solution—may require repeated evaluations of the forward operator. For complex physical models and high-dimensional model spaces, this can be computationally prohibitive.

Our previous work [6] sought to accelerate the Bayesian solution of inverse problems through the use of stochastic spectral methods. Based on polynomial chaos (PC) representations of random variables and processes [7–12], stochastic spectral methods have been used extensively for *forward* uncertainty propagation—characterizing the probability distribution of the output of a model given a known distribution on the input. These methods constitute attractive alternatives to Monte Carlo simulation in numerous applications: transport in porous media [13], structural mechanics [14], thermo-fluid systems [15–17], electrochemical microfluid systems [18], and reacting flow [19]. In the *inverse* context, the Bayesian formulation in [6] constructs a stochastic forward problem whose random inputs span the support of the prior and seeks its solution using Galerkin methods. The prior support may be partitioned, but for each partition the stochastic forward problem is solved only once. The resulting spectral representations of the forward operator enter the likelihood function, and exploration of the posterior is recast as Monte Carlo sampling of the variables underlying the PC expansion. We used this scheme to infer parameters appearing nonlinearly in a transient diffusion equation, demonstrating exponential convergence to the true posterior and substantial speedup.

Other attempts at accelerating Bayesian inference in computationally intensive inverse problems have relied on reductions or surrogates for the forward model, constructed through repeated forward simulations. Wang and Zabaras [20] use proper orthogonal decomposition (POD) [21] to accelerate forward model calculations in a radiative source inversion problem. The empirical basis 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 [20]. Balakrishnan et al. [22] 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. In the statistical literature, under the headline of "Bayesian parameter calibration", Gaussian processes have been used extensively as surrogates for complex computational models [23]. These approaches treat the forward model as a black box, and thus require careful attention to experimental design and to modeling choices that specify the mean and covariance of the surrogate Gaussian process. A different set of approaches retain the full forward model but use simplified or coarsened models to guide and improve the efficiency of Markov chain Monte Carlo (MCMC). Christen and Fox [24] 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. Higdon et al. [25] 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 [26] in which "swap proposals" allow information from the coarse-scale formulation to influence the fine-scale chain. Efendiev et al. [27] also develop a two-stage MCMC algorithm, using a coarse-scale model based on multiscale finite volume methods to improve the acceptance rate of MCMC proposals.

This paper extends the stochastic spectral methodology of [6] to inverse problems whose solutions are unknown functions—i.e., spatial or temporal fields. In doing so, we also explore dimensionality reduction in the Bayesian formulation of inverse problems, and the dependence of dimensionality on both the prior and the data. Inverse problems involving fields are vital to applications ranging from geophysics to medical imaging. Spatial fields may correspond to inhomogeneous material properties, such as permeabilities, diffusivities, or densities, or may represent distributed source terms in transport equations.

Estimating fields rather than parameters typically increases the ill-posedness of the inverse problem, since one is recovering an infinite-dimensional object from finite amounts of data. Obtaining physically meaningful results requires the injection of additional information on the unknown field—i.e., regularization [3]. A standard Bayesian approach is to employ Gaussian process (GP) or Markov random field (MRF) priors [28,4,29]. Most studies then explore the value of the field on a finite set of grid points [30]; the dimension of the posterior is tied to the discretization of the field. This recipe presents difficulties for stochastic spectral approaches, however, as the size of a PC basis does not scale favorably with dimension [9]. Moreover, with any degree of smoothness, the value of the field at each grid point hardly represents an independent direction.

Ideally, one should employ a representation that reflects how much information is truly required to capture variation among realizations of the unknown field. To this end, we introduce a Karhunen–Loève (K–L) expansion based on the prior random process, transforming the inverse problem to inference on a truncated sequence of weights of the K–L modes. Other recent work has also employed K–L expansions in the context of statistical inverse problems. Li and Cirpka [31] emphasize the role of K–L expansions in enabling geostatistical inversion on unstructured grids. Efendiev et al. [27] use K–L expansions to parameterize the log-permeability field in their two-stage MCMC scheme, and introduce constraints among the weights in order to match known values of the permeability at selected spatial locations. In contrast to [31], we use a fully Bayesian approach, generating true conditional realizations from a non-Gaussian posterior.

A more fundamental distinction of the present work is that we combine a K–L representation of the unknown field with spectral methods for uncertainty propagation. In particular, the Karhunen–Loève representation of a scaled Gaussian process prior *defines* the uncertainty that is propagated through the forward model with a stochastic Galerkin scheme. The deterministic forward model, originally specified by (a system of) partial differential equations, is thus replaced by stochastic PDEs; numerical approaches to such systems, in which random fields appear as boundary conditions or coefficients, have seen extensive development [9,16,32–35]. Uncertainty propagation yields a polynomial approximation of the forward operator over the support of the prior. This approximation then enters a reduced-dimensionality surrogate posterior, which we

explore with MCMC. The overall scheme avoids repeated forward simulations, and the computational cost per MCMC iteration becomes negligible compared to the cost of a full forward solution.

We demonstrate our scheme with a nonlinear forward model, and develop a fully Bayesian treatment of the problem in which hyperparameters describing the prior covariance are estimated simultaneously with the unknown field in a joint posterior distribution. To place the present Bayesian formulation in broader context, we recall connections between the K–L expansion and regularization penalties in the reproducing kernel Hilbert space (RKHS) norm corresponding to the prior covariance. We explore convergence both with respect to the number of K–L modes and the order of the PC basis. We also examine the efficiency of MCMC, quantify the limiting distribution of the K–L modes, and explore the impact of data resolution on the approach to this distribution. Significant gains in computational efficiency are achieved, with speedups of more than two orders of magnitude over a grid-based Bayesian scheme employing repeated forward solutions.

2. Formulation

We begin by introducing the essential building blocks of the present formulation: Gaussian processes (Section 2.1), the Karhunen–Loève (K–L) representation of a stochastic process (Section 2.2), and the Bayesian approach to inverse problems (Section 2.3). All three pieces are brought together in Section 2.4, where the K–L expansion of a Gaussian process prior enables a reduced-dimensionality approach to Bayesian inference. Stochastic spectral methods for forward uncertainty propagation are then introduced in Section 2.5, and used to accelerate the K–L/Bayesian approach in Section 2.5.2. Finally, connections among Gaussian priors, the K–L expansion, and regularization are reviewed in Section 2.6.

2.1. Gaussian processes

Let $(\Omega, \mathcal{U}, P)$ be a probability space, where Ω is a sample space, $\mathcal{U}$ is a σ-algebra over Ω, and P is a probability measure on $\mathcal{U}$. Also, let $D \subset \mathbb{R}^n$ be a bounded spatial domain. If $M(\mathbf{x}) : \Omega \to \mathbb{R}$ is a $\mathcal{U}$-measurable mapping for every $\mathbf{x} \in D$, then $M : \Omega \times D \to \mathbb{R}$ is a random field. $M(\mathbf{x}, \omega)$, for $\omega \in \Omega$, can thus be seen as a collection of real-valued random variables indexed by $\mathbf{x} \in D$. Alternatively, one can view $M(\cdot)$ as a random variable taking values in $\mathbb{R}^D$, the space of all real-valued functions on D. Though our presentation will focus on 'random fields' (typically signifying processes indexed by a spatial coordinate), the developments below are applicable to processes indexed by time or by both time and space.

If, for any $n \geqslant 1$ we have

$$(M(\mathbf{x}_1), \ldots, M(\mathbf{x}_n)) \stackrel{i.d.}{=} (M(\mathbf{x}_1 + \mathbf{s}), \ldots, M(\mathbf{x}_n + \mathbf{s})), \tag{1}$$

where $\stackrel{i.d.}{=}$ denotes equality in distribution, $\mathbf{s}$ is a spatial shift, and $\{\mathbf{x}_i, \mathbf{x}_i + \mathbf{s}\}_{i=1}^n \in D$, then M is said to be stationary [36]. If in addition, all finite-dimensional distributions of M are multivariate normal, then M is a *stationary Gaussian random field*, or simply a stationary Gaussian process (GP). Let $\mathbf{M}_{(n)} = (M(\mathbf{x}_1), \ldots, M(\mathbf{x}_n))$ denote the restriction of M to a finite set of indices. Then the characteristic function of $\mathbf{M}_{(n)}$ is [36]

$$\phi_M(\boldsymbol{\lambda}) \equiv \mathbb{E}[\exp(i\boldsymbol{\lambda}^T \mathbf{M}_{(n)})] = \exp\left(i\boldsymbol{\lambda}^T \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\lambda}^T \boldsymbol{\Sigma} \boldsymbol{\lambda}\right), \quad \boldsymbol{\lambda} \in \mathbb{R}^n, \tag{2}$$

where the mean is spatially invariant, $\boldsymbol{\mu} \equiv \mu \mathbf{1}_n$, and entries of $\boldsymbol{\Sigma}$ are values of the covariance function C:

$$\begin{aligned} \Sigma_{ij} &= C(\mathbf{x}_i, \mathbf{x}_j) \\ &\equiv \text{Cov}[M(\mathbf{x}_i), M(\mathbf{x}_j)] = \mathbb{E}[(M(\mathbf{x}_i) - \mu)(M(\mathbf{x}_j) - \mu)] \tag{3} \\ &= \widetilde{C}(\mathbf{x}_i - \mathbf{x}_j). \tag{4} \end{aligned}$$

Gaussian processes have finite second moments; that is, $M(\mathbf{x}) \in L^2(\Omega)$ for every $\mathbf{x}$ [37]. If $\boldsymbol{\Sigma}$ is invertible, the finite-dimensional density of order n of the Gaussian process is then

$$p(\mathbf{m}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{m} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{m} - \boldsymbol{\mu})\right), \tag{5}$$

where $\mathbf{m} = (m(\mathbf{x}_1), \ldots, m(\mathbf{x}_n))$. If we further restrict C to depend only on the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$, that is we put $\widetilde{C}(\mathbf{d}) = f(\|\mathbf{d}\|)$, then the stationary GP is called *isotropic* [38]. It is common to specify the covariance function with scale and range parameters θ_1 and θ_2, respectively [28]:

$$\widetilde{C}(\mathbf{d}) = \theta_1 \rho\left(\frac{\|\mathbf{d}\|}{\theta_2}\right). \tag{6}$$

Here $\rho(\cdot)$ is a correlation function, positive definite with $\rho(0) = 1$ [25], e.g., $\rho(d) = e^{-d}$ or $\rho(d) = e^{-d^2}$.

Gaussian processes are extensively employed as priors in Bayesian inference [38]. In particular, conceiving of the GP as a prior over functions motivates Gaussian process regression [39], also known as kriging in spatial statistics [40]; further applications include classification, with ties to support vector machines and other kernel methods [41]. Depending on the covariance kernel, realizations of a Gaussian process may be smooth or periodic, or for non-stationary kernels, capture certain

trends [42]. Gaussian process priors can thus inject regularity by assigning low probability to fields with undesirable properties. See Section 2.6 for a more formal discussion of Gaussian process priors and regularization.

2.2. Karhunen–Loève expansion

Let $M(\mathbf{x}, \omega)$ be a real-valued random field with finite second moments, mean $\mu(\mathbf{x})$, and a covariance function that is continuous on $D \times D$, with D bounded. Then M has the following representation, termed a Karhunen–Loève (K–L) expansion [43]:

$$M(\mathbf{x}, \omega) = \mu(\mathbf{x}) + \sum_{k=1}^{\infty} \sqrt{\lambda_k} c_k(\omega) \phi_k(\mathbf{x}). \tag{7}$$

In general, this equality holds pointwise and in the mean square sense; that is, convergence is in $L^2(\Omega)$ for each $\mathbf{x} \in D$. If M is Gaussian and almost surely continuous, then convergence is uniform over D with probability one [44].[1] λ_k and $\phi_k(\mathbf{x})$ are eigenvalues and eigenfunctions of the linear operator corresponding to the covariance kernel C:

$$\int_D C(\mathbf{x}_1, \mathbf{x}_2) \phi_k(\mathbf{x}_2) d\mathbf{x}_2 = \lambda_k \phi_k(\mathbf{x}_1). \tag{8}$$

By the assumptions on M, the covariance kernel is symmetric and positive semidefinite, and thus by Mercer's theorem we have [36,46]

$$C(\mathbf{x}_1, \mathbf{x}_2) = \sum_{k=1}^{\infty} \lambda_k \phi_k(\mathbf{x}_1) \phi_k(\mathbf{x}_2), \tag{9}$$

where the eigenfunctions $\phi_k(\mathbf{x})$ are continuous and form a complete orthonormal system in $L^2(D)$. The random variables $c_k(\omega)$ are uncorrelated with zero mean and unit variance:

$$\mathbb{E}c_k = 0, \quad \mathbb{E}[c_j c_k] = \delta_{jk}. \tag{10}$$

These variables are in general non-Gaussian

$$c_k(\omega) = \frac{1}{\sqrt{\lambda_k}} \int_D (M(\mathbf{x}, \omega) - \mu(\mathbf{x})) \phi_k(\mathbf{x}) d\mathbf{x} \tag{11}$$

but if M is also a Gaussian process, the c_k are Gaussian and independent, $c_k \sim N(0, 1)$.

The K–L expansion is *optimal* in the following sense. Of all possible orthonormal bases for $L^2(D)$, the $\{\phi_k(\mathbf{x})\}$ satisfying (8) minimize the mean-squared error in a finite linear representation of $M(\cdot)$ [9]. That is, they minimize

$$\int_{\Omega \times D} \left(M(\mathbf{x}, \omega) - \mu(\mathbf{x}) - \sum_{k=1}^{K} \sqrt{\lambda_k} c_k(\omega) \phi_k(\mathbf{x}) \right)^2 dP(\omega) d\mathbf{x} \tag{12}$$

for any $K \geqslant 1$. As a result, the K–L expansion is an extremely useful tool for the concise representation of stochastic processes. It has close analogues in data reduction (i.e., principal components analysis), model reduction (proper orthogonal decomposition) [47], and linear algebra (SVD). If $M(\cdot)$ is approximated by a K-term K–L expansion,

$$M_K(\mathbf{x}, \omega) = \mu(\mathbf{x}) + \sum_{k=1}^{K} \sqrt{\lambda_k} c_k(\omega) \phi_k(\mathbf{x}), \tag{13}$$

the covariance function of M_K is simply

$$C_K(\mathbf{x}_1, \mathbf{x}_2) = \sum_{k=1}^{K} \lambda_k \phi_k(\mathbf{x}_1) \phi_k(\mathbf{x}_2), \tag{14}$$

which converges uniformly to (9) as $K \to \infty$ [44]. In particular, the total variance or "energy" of M_K is

$$\int_D \mathbb{E}[M_K(\mathbf{x}, \omega) - \mu(\mathbf{x})]^2 d\mathbf{x} = \int_D C_K(\mathbf{x}, \mathbf{x}) d\mathbf{x} = \sum_{k=1}^{K} \lambda_k \tag{15}$$

following from the orthonormality of the $\{\phi_k(\mathbf{x})\}$.

---

[1] Sufficient conditions for the continuity of Gaussian processes are detailed in Adler [44]. Abrahamsen [45] suggests that any Gaussian process on compact $D \in \mathbb{R}^n$ with a continuous mean and a continuous and "reasonable" covariance function will satisfy these conditions. Covariance functions that provably yield a.s. continuous Gaussian processes include Gaussian, exponential, spherical, Matérn, spline, and polynomial kernels, along with numerous others [44,38].

2.3. Bayesian approach to inverse problems

Bayesian approaches to inverse problems have received much recent interest [48,49,4], with applications ranging from geophysics [50,51] and climate modeling [52] to heat transfer [53,20]. We review this approach briefly below; for more extensive introductions, see [4,5,48].

Consider a forward problem defined as follows:

$$\mathbf{d} \approx \mathbf{G}(\mathbf{m}). \tag{16}$$

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

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})p_m(\mathbf{m}). \tag{17}$$

In the Bayesian paradigm, 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 function $p(\mathbf{d}|\mathbf{m})$, 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 $\mathbf{d}$:

$$\mathbf{d} = \mathbf{G}(\mathbf{m}) + \boldsymbol{\eta}, \tag{18}$$

where components of $\boldsymbol{\eta}$ are i.i.d. random variables with density p_η. The likelihood then takes the form

$$L(\mathbf{m}) = p_\eta(\mathbf{d} - \mathbf{G}(\mathbf{m})) = \prod_i p_\eta(d_i - G_i(\mathbf{m})). \tag{19}$$

Additional information on the model parameters may enter the formulation through the 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, such as correlations or smoothness. In the absence of additional information, one may choose a prior that is *uninformative*. Here we will focus on Gaussian process priors, which for finite-dimensional $\mathbf{m}$ take the form of (5).

If parameters ϕ_m of the prior density $p_m(\mathbf{m}|\phi_m)$ or parameters ϕ_η of the error model $p_\eta(\eta_i|\phi_\eta)$ 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 [48]:

$$p(\mathbf{m}, \phi_m, \phi_\eta|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m}, \phi_\eta)p_m(\mathbf{m}|\phi_m)p(\phi_\eta)p(\phi_m). \tag{20}$$

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 $\mathbf{m}$, ϕ_m, and ϕ_η [54,48]. In the present study, we will introduce hyperparameters to describe aspects of the prior covariance.

2.4. Dimensionality reduction in inverse problems

Now we integrate concepts from the previous three sections. Consider an inverse problem in which the unknown quantities comprise a real-valued field $M(\mathbf{x})$. In a computational setting, this field and the forward model must be discretized. If $M(\mathbf{x})$ can be adequately represented on a finite collection of points $\{\mathbf{x}_i\}_{i=1}^n \in D$, then we can write both the prior and posterior densities in terms of $\mathbf{m} = (M(\mathbf{x}_1), \ldots, M(\mathbf{x}_n))$. That is, we can directly apply the Bayesian formulation described in the preceding section and explore the posterior density of $\mathbf{m}$ with Markov chain Monte Carlo (MCMC) [55]. The vector $\mathbf{m}$, however, will likely be high-dimensional. High dimensionality not only renders MCMC exploration of the posterior more challenging and costly, but taxes the polynomial chaos formulation we introduce below to accelerate evaluations of the posterior density [6].

Instead of exploring the value of $M(\mathbf{x})$ on each of n index points, we appeal to the K–L expansion. Let $M(\mathbf{x})$ be endowed with a Gaussian process prior with mean $\mu(\mathbf{x})$ and covariance kernel $C(\mathbf{x}_1, \mathbf{x}_2)$; we denote this as $M \sim \mathcal{GP}(\mu, C)$. Introduce the corresponding K-term K–L representation of $M(\mathbf{x})$ (13), with eigenvalues λ_k and eigenfunctions $\phi_k(\mathbf{x})$ satisfying (8). In general, $M(\mathbf{x}, \omega)$ is approached pointwise in mean square (and therefore in distribution) by $M_K(\mathbf{x}, \omega)$ as $K \to \infty$. For M a.s. continuous (see Section 2.2), realizations $M(\mathbf{x}, \omega)$ can be uniformly approximated as closely as desired by $M_K(\mathbf{x}, \omega)$—implying a corresponding realization $\mathbf{c}(\omega) \equiv (c_1(\omega), \ldots, c_K(\omega))$—with probability one. Updating distributions of M, by conditioning on the data, is thus equivalent to updating the joint distribution of the mode strengths c_k. We emphasize this viewpoint by writing $M_K(\mathbf{x}, \omega) = M_K(\mathbf{x}, \mathbf{c}(\omega)) = M_K(\mathbf{c})$, parameterizing M by the vector of weights $\mathbf{c}$. Components c_k are independent under the Gaussian process prior, with $c_k \sim N(0, 1)$. We thus truncate the K–L expansion at K terms and write a posterior density for $\mathbf{c}$:

$$p(\mathbf{c}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{c}) \prod_{k=1}^{K} p(c_k)$$

$$\propto p_\eta(\mathbf{d} - \mathbf{G}(M_K(\mathbf{c}))) \prod_{k=1}^{K} \exp(-c_k^2/2). \tag{21}$$

The inverse problem has been transformed to an inference problem on the weights c_k of a finite number of K–L modes. Note that the spatial discretization of $M(\mathbf{x})$ and of the forward model is now independent of the dimension of the posterior distribution. Here we have assumed the prior covariance to be completely known, thus ignoring hyperparameters in the expression for the posterior; we will relax this assumption in Section 3.

Truncating the K–L expansion in this context amounts to using a "modified" prior covariance kernel given by (14). Since the eigenvalues λ_k decay—exponentially fast for a smooth covariance kernel [34], algebraically fast in other cases—a small number of terms may be sufficient to capture almost all of the prior covariance. The linear operator corresponding to the modified covariance kernel now has finite rank; $\phi_k(\mathbf{x})$ that are not eigenfunctions of this operator cannot contribute to the inverse solution. The impact of this truncation is explored in Section 3.3.

2.5. Polynomial chaos acceleration

MCMC exploration of the reduced-dimensionality posterior (21) still requires repeated solutions of the forward model, once for each proposed move of the Markov chain. While dimensionality reduction may reduce the number of such evaluations, depending on the details of the MCMC sampler, it is desirable to avoid repeated forward solutions altogether.

Our previous work [6] introduced methods for accelerating Bayesian inference in this context, by using stochastic spectral methods to propagate prior uncertainty through the forward problem. These methods effectively create a "surrogate" posterior containing polynomial chaos (PC) representations of the forward model outputs. This density may be evaluated orders of magnitude more quickly than the "direct" posterior containing the full forward problem. Here, we will use the Gaussian process prior on M (and thus the prior distribution on $\mathbf{c}$) to define an appropriate stochastic forward problem. The K–L expansion of M ensures that we have chosen a concise representation of the prior uncertainty that is yet suitable for reconstructing inverse solutions. Beginning with polynomial chaos expansions, elements of this approach are described below.

2.5.1. Forward propagation of uncertainty

Once again let $(\Omega, \mathcal{U}, P)$ be a probability space on which we define a random process $X : \Omega \rightarrow \mathbb{R}^D$ with index set $D \subseteq \mathbb{R}^N$. Also, let $\{\xi_i(\omega)\}_{i=1}^{\infty}$ be i.i.d. standard normal random variables on Ω. Then any square-integrable X has the following representation:

$$X(\omega) = a_0 \Gamma_0 + \sum_{i_1=1}^{\infty} a_{i_1} \Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} a_{i_1 i_2} \Gamma_2(\xi_{i_1}, \xi_{i_2}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3} \Gamma_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + \cdots, \tag{22}$$

where Γ_p is the Wiener polynomial chaos of order p [7,9,56] and the $a_{i_1 i_2 \ldots}$ may be functions on D. This expansion can be rewritten in a more compact form

$$X(\omega) = \sum_{k=0}^{\infty} \hat{a}_k \Psi_k(\xi_1, \xi_2, \ldots), \tag{23}$$

where there is a one-to-one correspondence between the coefficients and functionals in (22) and in (23) [9]. For the standard normal 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 retain infinite summations, 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*

$$X(\omega) = \sum_{k=0}^{P} x_k \Psi_k(\xi_1, \xi_2, \ldots, \xi_n) \tag{24}$$

is

$$P + 1 = \frac{(n+p)!}{n!p!}. \tag{25}$$

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 [10,57]. For each of these choices, orthogonality of the polynomials $\Psi_k(\xi)$ with respect to the inner product on $L^2(\Omega)$ is maintained:

$$\langle \Psi_i \Psi_j \rangle = \int \Psi_i(\xi(\omega)) \Psi_j(\xi(\omega)) dP(\omega) = \int \Psi_i(\xi) \Psi_j(\xi) \rho(\xi) d\xi = \delta_{ij} \langle \Psi_i^2 \rangle, \tag{26}$$

where $\rho(\xi)$ denotes the probability density of ξ. 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:

$$\tilde{f}(\omega) = \sum_{k=0}^{P} f_k \Psi_k(\xi), \quad f_k = \frac{\langle f(X) \Psi_k \rangle}{\langle \Psi_k^2 \rangle}. \tag{27}$$

This orthogonal projection minimizes the error $\|f - \tilde{f}\|_2$ on the space spanned by $\{\Psi_k\}_{k=0}^{P}$, where $\| \cdot \|_2$ 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 is a random variable or process with a known PC expansion $X = \sum_{i=0}^{P} x_i \Psi_i(\xi)$. 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:

$$\left\langle \mathcal{O}\left(\sum_{k}^{P} f_k \Psi_k, \sum_{i}^{P} x_i \Psi_i\right) \Psi_j \right\rangle = 0, \quad j = 0 \cdots P. \tag{28}$$

This Galerkin approach is known as "intrusive" spectral projection [9,16], 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 ξ [14,58].

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 nonpolynomial functions f. These operations are incorporated into a library for "stochastic arithmetic," detailed in [11].

2.5.2. Stochastic spectral formulation of Bayesian inference

In [6], we described three accelerated schemes for computing posterior estimates, all based on spectral solutions of a stochastic forward problem: Monte Carlo sampling from the prior distribution, importance sampling, and MCMC. Here we focus on the latter case. The essential idea is to construct a stochastic forward problem whose solution approximates the deterministic forward model over the support of the prior.

Let us begin with (i) a finite-dimensional representation of the unknown quantity that is the object of inference, and (ii) a prior distribution on the parameters of this representation. For instance, if the unknown quantity is a field $M(x)$ endowed with a Gaussian process prior, the finite representation may be a truncated K–L expansion with mode strengths $\mathbf{c}$ and priors $c_i \sim N(0, 1)$. The Bayesian formulation in Section 2.4 describes the inverse solution in terms of the posterior density of $\mathbf{c}$, which includes evaluations of the forward model $\mathbf{G}(M_K(\cdot))$. For simplicity, we shall abbreviate $\mathbf{G} \circ M_K$ as $\mathbf{G_c}$; inputs to this model are parameterized by $\mathbf{c}$. Also, let $\mathcal{C}$ denote the support of the prior.

Now define a random vector $\check{\mathbf{c}} = \mathbf{g}(\check{\xi})$, each component of which is given by a PC expansion

$$\check{c}_i = g_i(\check{\xi}) = \sum_{k=0}^{P} g_{ik} \Psi_k(\check{\xi}). \tag{29}$$

This vector will serve as input to $\mathbf{G_c}$, thus specifying a *stochastic forward problem*. Recall that the distribution of $\check{\xi}$ (e.g., standard normal) and the polynomial form of Ψ (e.g., multivariate Hermite) are intrinsic properties of the PC basis. We do not require that $\mathbf{g}$ be chosen such that $\check{\mathbf{c}}$ is distributed according to the prior on $\mathbf{c}$. Rather, we require only (1) that $\Xi_c = \mathbf{g}^{-1}(\mathcal{C})$, the inverse image of the support of the prior, be contained within the range of $\check{\xi}$, and (2) that $\mathbf{g}$ be a diffeomorphism from Ξ_c to $\mathcal{C}$.

Next, using Galerkin projection to solve the stochastic forward problem, we obtain a PC representation for each component of the model output. Here G_i is the ith component of $\mathbf{G_c}$, and $\widetilde{G}_i(\check{\xi})$ is its PC representation:

$$\widetilde{G}_i(\check{\xi}) = \sum_{k=0}^{P} d_{ik} \Psi_k(\check{\xi}). \tag{30}$$

The forward prediction $\widetilde{\mathbf{G}}$ obtained in this fashion is a function of $\check{\xi}$, and is a polynomial chaos approximation of $\mathbf{G_c}(\mathbf{g}(\check{\xi}))$. Note that both of these quantities are random variables, since $\check{\xi}$ is a random variable. But $\widetilde{\mathbf{G}}$ can also be evaluated with a deterministic argument[2]; in this sense, $\widetilde{\mathbf{G}}$ is a polynomial approximation of the deterministic forward model $\mathbf{G_c} \circ \mathbf{g}$.

We would like to use this approximation to replace $\mathbf{G_c}$ in the likelihood function $L(\mathbf{c}) \equiv p_\eta(\mathbf{d} - \mathbf{G_c}(\mathbf{c}))$:

$$L(\mathbf{g}(\xi)) \approx \widetilde{L}(\xi) \equiv p_\eta(\mathbf{d} - \widetilde{\mathbf{G}}(\xi)). \tag{31}$$

Implicit in this substitution is the change of variables $\mathbf{c} = \mathbf{g}(\xi)$, i.e., from the input parameterization of $\mathbf{G_c}$ to the input parameterization of $\widetilde{\mathbf{G}}$, enabled because $\mathbf{g}$ satisfies conditions (1) and (2) above.

We write the change of variables in terms of the posterior expectation of an arbitrary function f:

---

[2] In this exposition we have used ˘ to identify the random variables $\check{\mathbf{c}}$ and $\check{\xi}$ in order to avoid confusion with deterministic arguments to probability density functions, e.g., $\mathbf{c}$ and ξ below. Elsewhere, we revert to the usual notational convention and let context make clear the distinction between the two.

$$\mathbb{E}_{\pi_c} f = \mathbb{E}_{\pi_\xi} (f \circ \mathbf{g}), \tag{32}$$

where $\pi_c \equiv p(\mathbf{c}|\mathbf{d})$ is the posterior density on $\mathbf{c}$-space, and π_ξ is the corresponding posterior density on ξ-space:

$$\pi_\xi(\xi) \propto L(\mathbf{g}(\xi)) p_c(\mathbf{g}(\xi)) |\det D\mathbf{g}(\xi)|. \tag{33}$$

Here, $D\mathbf{g}$ denotes the Jacobian of $\mathbf{g}$ and p_c is the prior density of $\mathbf{c}$. Eliminating the forward model from the likelihood function via (31) finally yields the "surrogate" posterior density $\tilde{\pi}_\xi$:

$$\begin{aligned} \tilde{\pi}_\xi(\xi) &\propto \widetilde{L}(\xi) p_c(\mathbf{g}(\xi)) |\det D\mathbf{g}(\xi)| \\ &\propto p_{\boldsymbol{\eta}} (\mathbf{d} - \widetilde{\mathbf{G}}(\xi)) p_c(\mathbf{g}(\xi)) |\det D\mathbf{g}(\xi)|. \end{aligned} \tag{34}$$

This distribution may be explored with any suitable sampling strategy, in particular MCMC. Evaluating the density for purposes of sampling may have negligible cost; nearly all the computational time may be spent in intrusive spectral projection, obtaining the PC expansions in (30). Depending on model nonlinearities, the necessary size of the PC basis, and the number of posterior samples required, this computational effort may be orders of magnitude less costly than exploring the posterior via direct sampling. Accuracy of the surrogate posterior depends on the order and family of the PC basis, as well as on the choice of transformation $\mathbf{g}$—for instance, whether the distribution of $\check{\mathbf{c}}$ assigns sufficient probability to regions of $\mathcal{C}$ favored by the posterior on $\mathbf{c}$. A detailed discussion of these issues can be found in [6].

2.6. Gaussian processes, K–L expansions, RKHS, and regularization

There are important connections between Gaussian process priors and regularization penalties in the corresponding reproducing kernel Hilbert space (RKHS) norm. These connections can be understood in terms of the spectral expansion of the covariance kernel, and it is useful to review these relationships in the present context.

The definition and properties of reproducing kernel Hilbert spaces are briefly reviewed in the Appendix. It is natural to think of a positive definite reproducing kernel K as a *covariance* kernel, and indeed any Gaussian process can be associated with a RKHS. Let $X(t), t \in \mathcal{T}$ be a centered Gaussian process with covariance kernel K. If K has more than a finite number of non-zero eigenvalues, realizations of $X(t)$ are almost surely *not* in the corresponding RKHS $H(K)$ [59].[3] However, there exists an isometry between the two. In particular, let $\mathcal{H}$ be the Hilbert space spanned by $X(t)$: $\mathcal{H} = \text{span}\{X_t, t \in \mathcal{T}\}$ with inner product $\langle Z_i, Z_j \rangle = \mathbb{E}[Z_i Z_j]$ for $Z_i, Z_j \in \mathcal{H}$. It can be shown that $\mathcal{H}$ is isometrically isomorphic to $H(K)$ [59,44].

Bayesian estimates with Gaussian process priors may lie in the corresponding RKHS, however [59,60]. Consider the case of an inverse problem with Gaussian process prior $\mathcal{GP}(0, K)$ on the unknown function f. Details of the likelihood function and forward model are unimportant here. For simplicity, we assume that the prior covariance is completely known. Let f_i denote the projection of f onto the ith K–L eigenfunction ϕ_i:

$$f_i = \int_D f(s) \phi_i(s) ds, \tag{35}$$

where ϕ_i and λ_i are, as usual, eigenfunctions and eigenvalues of the linear operator corresponding to K. According to (11), the prior distribution on each f_i is $N(0, \lambda_i)$. Then the posterior probability of the function f, $\pi(f)$, has the form:

$$J = -\log \pi(f) = -\text{log-likelihood} + \sum_i^\infty \frac{f_i^2}{2\lambda_i} + \text{const} = \cdots + \frac{1}{2} \|f\|_{H(K)}^2 + \cdots, \tag{36}$$

J is thus a RKHS-norm penalized cost functional, in which the Gaussian process prior provides the regularization penalty. Minimizing J to obtain the MAP estimate of f is equivalent to finding a Tikhonov-regularized solution to the inverse problem, with $\text{argmin} J = f_{\text{MAP}} \in H(K)$. Changing the prior covariance kernel amounts to changing the RKHS norm and thus the nature of the regularization penalty.

Moreover, there is an equivalence between the RKHS regularization functional $\|f\|_{H(K)}$ and a standard $L^2(D)$-norm penalty $\|Lf\|_2$ containing the differential operator L: the reproducing kernel K is the Green's function of the operator $L^* L$, where L^* denotes the adjoint of L [5,38]. Thus a Gaussian kernel leads to a penalty on derivatives of all orders; the exponential covariance kernel penalizes the square of the function value $f(s)$ and its derivative $\dot{f}(s)$ (i.e., a Sobolev H^1 norm); and the covariance kernel of Brownian motion, $K(s, t) = \min(s, t)$, leads to a penalty on the MAP estimate's squared derivatives.

Finally, we note that the present scheme of K–L based inversion (Section 2.4) recalls the "weight space" view of Gaussian process regression [41], in that we find weights on the set of feature vectors ϕ_k implied by the Gaussian process prior.

3. Numerical implementations and results

We explore the accuracy and efficiency of our dimensionality reduction approach by estimating inhomogeneous diffusivity fields in a transient diffusion problem. We pursue these inverse problems both with and without the added step of solving

---

[3] As an example of a GP whose realizations are not in the RKHS, consider standard Brownian motion. Sample paths are nowhere differentiable with probability one, but members of the RKHS are differentiable almost everywhere, with square-integrable derivatives.

the stochastic forward problem to construct a surrogate posterior (Section 2.5). In particular, we consider a diffusion equation on the unit interval $D = [0, 1]$ with adiabatic boundaries:

$$
\begin{aligned}
&\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(v(x)\frac{\partial u}{\partial x}\right) + \sum_{i=1}^{N} \frac{s_i}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{|l_i - x|^2}{2\sigma_i^2}\right)[1 - H(t - T_i)], \\
&\left.\frac{\partial u}{\partial x}\right|_{x=0} = \left.\frac{\partial u}{\partial x}\right|_{x=1} = 0, \\
&u(x, t = 0) = 0.
\end{aligned}
\tag{37}
$$

The source term in (37) involves N localized sources, each active on the interval $t \in [0, T_i]$ and centered at $l_i \in D$ with strength s_i and width σ_i, $i = 1, \ldots, N$. Source parameters are prescribed, and we infer $v(x)$ from noisy measurements of the u-field at a finite set of locations and times. This problem can be considered a prototype for the inverse estimation of an inhomogeneous conductivity field or any analogous material or transport property, such as the permeability field in a porous medium [28,61].

3.1. Inverse problem setup

The transient diffusion equation above may be cast as a forward model that predicts the value of the field at specific locations and times. Taking the diffusivity to be uniformly bounded away from zero, $v(x) > v_0 > 0$, we define the log-diffusivity $M(x) \equiv \log[v(x) - v_0]$; this function is the input to the forward model. We evaluate the field at mn points $\{u(x_i, t_j) : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\}$. The "sensor locations" $\{x_1, \ldots, x_m\}$ are uniformly spaced on D, including the endpoints, and the measurement times $\{t_1, \ldots, t_n\}$ are uniformly spaced on an arbitrary time interval. Below, unless otherwise specified, we use $m = 13$ sensors, $n = 9$ measurement times, and $N = 3$ sources. The source locations are staggered with respect to the sensors; i.e., they are placed at $l_i \in \{0.25, 0.50, 0.75\}$. We prescribe identical strengths $s_i = 100$, shutoff times $T_i = 0.01$, and widths $\sigma_i^2 = 10^{-3}$ for all three sources. We fix $v_0 = 0.1$. Measurements take place over the time interval $t \in [0.01, 0.03]$.

The u-field is described on a uniform grid with spacing $h = 1/48$. 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 [62] with $\Delta t = 10^{-4}$. For any input $v(x)$, 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 [63]. Numerical resolution studies were conducted to validate the present choices of h and Δt.

Note that the forward model is nonlinear in $v(x)$. Consider the simple case of a uniform diffusivity, $v(x) = \bar{v}$. Fig. 1 shows the resulting forward maps, from $\log[\bar{v} - v_0]$ to u, at a single measurement location and two successive times. The measurement location, $x^* = 1/6$, is adjacent to a source at $x = 1/4$. For very small diffusivities, the scalar introduced by the source does not diffuse towards the sensor in appreciable quantity, and hence u is small; in this regime, the magnitude of the scalar field rises with t and with $\bar{v}$. At very large diffusivities, the scalar introduced by all N sources rapidly diffuses towards all the sensors and the u-field quickly becomes uniform, approaching $u = \sum_i s_i T_i = 3$ as $\bar{v}t \to \infty$. For intermediate diffusivities, the measured value of u may decrease with rising $\bar{v}$: the scalar field, locally peaked at the nearby source, flattens as the diffusivity increases, until the influence of the remaining sources is felt at sufficiently high $\bar{v}$, raising the local value of u once again. The behavior of analogous forward maps in the case of nonuniform $v(x)$ is expected to be even more complicated.

The *inverse problem* thus consists of inferring $M(x) \equiv \log[v(x) - v_0]$ from *noisy* measurements of $u(x_i, t_j)$. In the Bayesian setting, we provide statistical information about the measurement process and about our prior knowledge of $M(x)$. 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 (18). In the examples below, we choose $\varsigma = 0.1$. We endow $M(x)$ with a zero-mean Gaussian process prior $M \sim \mathcal{GP}(0, C)$, where C is a stationary Gaussian covariance kernel:

$$
C_\theta(x_1, x_2) = \widetilde{C}_\theta(|x_1 - x_2|) = \theta\exp\left(-\frac{|x_1 - x_2|^2}{2L^2}\right).
\tag{38}
$$

For simplicity, we assume that the correlation length L is known; in practical applications, an estimate of L will often be available [28,61]. We do not, on the other hand, presume to know the scale θ of the prior covariance. Adopting a fully Bayesian approach, we let θ be a hyperparameter endowed with a conjugate inverse gamma hyperprior, $\theta \sim IG(\alpha, \beta)$ [64,65,53]:

$$
p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-\alpha-1}\exp\left(-\frac{\beta}{\theta}\right).
\tag{39}
$$

In the examples below, we fix the shape parameter $\alpha = 1$ and the scale parameter $\beta = 1$. This yields a proper but long-tailed prior for θ, with undefined mean and variance. The magnitude of the prior covariance θ joins the remaining parameters describing $M(x)$ in the joint posterior density; we can then marginalize over θ to obtain a posterior describing $M(x)$ alone (see Section 2.3). Note that when considering MAP estimates of M conditioned on θ, the ratio ς^2/θ is akin to the regularization parameter appearing in deterministic inversion; thus we are effectively estimating the strength of the regularization when conditioning on the data [4].
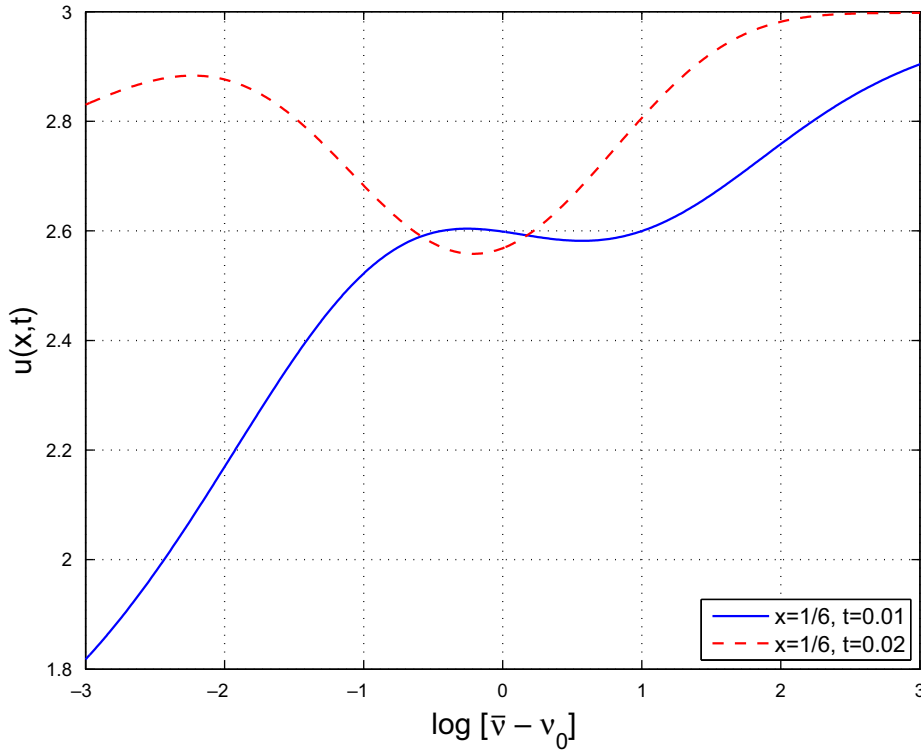
Fig. 1. Simple forward map, from a uniform diffusivity $\bar{v}$ to the value of the scalar field $u(\bar{v}; x, t)$ at two successive times. Source parameters are given in Section 3.1.

In Sections 3.2, 3.3, and 3.4, we will solve the inverse problem for four different "target profiles" $M(x)$: a simple linear profile, a sinusoidal profile, a profile randomly drawn from the Gaussian process prior with $L = 0.3, \theta = 1.0$, and a well-shaped profile. An additional target profile is introduced in Section 3.6. For each profile, a noisy data vector $\mathbf{d}$ is generated by solving the deterministic forward problem with the target log-diffusivity, then perturbing the resulting value of u at each sensor location/time with independent samples of Gaussian noise $\eta_i \sim N(0, \varsigma^2)$. To avoid an "inverse crime" [4], we generate the mn values of $u(x_i, t_j)$ by solving the forward problem at a much higher resolution than that used in the inversion, i.e., with $h = 1/408$ and a correspondingly finer Δt.

3.2. Grid-based inversion

We begin with a straightforward full-dimensional Bayesian approach to the inverse problem, as described at the start of Section 2.4. Let $M(x)$ be represented on a finite collection of points $\{x_i\}_{i=1}^n \in D$; an obvious choice with adequate resolution is the collection of grid points used to discretize the forward model, uniformly spaced on the unit interval with $x_{i+1} - x_i = h$. Then we can write both the prior and posterior densities in terms of $\mathbf{m} = (M(x_1), \ldots, M(x_n))$:

$$p(\mathbf{m}, \theta | \mathbf{d}) \propto p(\mathbf{d} | \mathbf{m}) \times p(\mathbf{m} | \theta) \times p(\theta) \propto p_\eta(\mathbf{d} - \mathbf{G}(\mathbf{m})) \times \theta^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \mathbf{m}^T \Sigma_\theta^{-1} \mathbf{m}\right) \times p(\theta)$$

$$\propto \exp\left(-\frac{[\mathbf{d} - \mathbf{G}(\mathbf{m})]^T [\mathbf{d} - \mathbf{G}(\mathbf{m})]}{2\varsigma^2}\right) \times \theta^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \mathbf{m}^T \Sigma_\theta^{-1} \mathbf{m}\right) \times \theta^{-\alpha-1} \exp\left(-\frac{\beta}{\theta}\right), \tag{40}$$

where $(\Sigma_\theta)_{ij} \equiv C_\theta(x_i, x_j)$.

Directly applying a Metropolis–Hastings algorithm to this posterior, however, is not likely to be successful. Simple proposal distributions for $\mathbf{m}$, such as normal distributions centered at the current position of the chain, generate candidate points with very low acceptance probabilities—even when applied component-at-a-time [61]. These proposals do not account for correlations among neighboring components of $\mathbf{m}$. We surmount this issue with a change of variables, using the Cholesky factorization of the prior covariance matrix, using $\theta = 1$: $\Sigma_{(\theta=1)} = \mathbf{L}\mathbf{L}^T$. If $\mathbf{z}$ is vector of n i.i.d. standard normal random variables, then, $\forall \theta$,

$$\mathbf{m} = \sqrt{\theta}\mathbf{L}\mathbf{z} \tag{41}$$

will have a zero-mean multivariate normal distribution with covariance Σ_θ. (Multiplication by **L** is analogous to, in the continuous case, generating samples of a Gaussian process by convolution with white noise [66].) Thus the $N(\mathbf{0}, \Sigma_\theta)$ prior distribution on **m** reduces to a Gaussian prior on **z** with *diagonal* covariance, $N(\mathbf{0}, \mathbf{I})$. Equivalently, we can write $\mathbf{m} = \mathbf{Lz}$ and let the scale parameter θ control the prior variance of $\mathbf{z} \sim N(\mathbf{0}, \theta\mathbf{I})$, thus reparameterizing the posterior density as follows:

$$p(\mathbf{z}, \theta|\mathbf{d}) \propto \exp\left(-\frac{[\mathbf{d} - \mathbf{G}(\mathbf{Lz})]^T[\mathbf{d} - \mathbf{G}(\mathbf{Lz})]}{2\varsigma^2}\right) \times \theta^{-\frac{n}{2}}\exp\left(-\frac{\mathbf{z}^T\mathbf{z}}{2\theta}\right) \times \theta^{-\alpha-1}\exp\left(-\frac{\beta}{\theta}\right). \tag{42}$$

We use a Metropolis–Hastings algorithm to simulate samples from this distribution [55]. For the scale parameter θ, we apply Gibbs updates: the full conditional $p(\theta|\mathbf{z}, \mathbf{d})$ is proportional to $IG(\alpha + n/2, \beta + (\sum_{i=1}^{n} z_i^2)/2)$ [64], so we sample directly from this distribution with acceptance probability 1. For the remaining parameters **z**, we use single-component random-walk Metropolis updates: each proposal distribution $q(\cdot|\cdot)$ is a univariate normal centered on the current position of the chain. It may be possible to increase the efficiency of this sampler by using single-component updating for the first few components of **z** and block updating for the higher-index, less important components [28], but we do not pursue such fine-tuning here. MCMC yields a series of samples $\{(\mathbf{z}^{(s)}, \theta^{(s)})\}$, which are easily transformed to $\{(\mathbf{m}^{(s)}, \theta^{(s)})\}$. From these samples, we can estimate posterior expectations (e.g., means, variances, higher moments), extract marginal densities $p(M(x_i)|\mathbf{d})$, and estimate quantiles of the marginal distributions.

Figs. 2–4 show the results of grid-based inversion with several target profiles. Since we are not inferring the prior correlation length, in each case we have assumed a value that reflects characteristic length scales of the target: $L = 1.0$ for the linear profile (not shown), $L = 0.2$ for the sinusoidal profile, $L = 0.1$ for the well-shaped profile, and $L = 0.3$ for the profile corresponding to a random draw from the GP prior. Figs. 2–4(a) show the posterior mean and standard deviation, along with five samples from each posterior distribution. In part (b) of these figures, we extract one-dimensional marginal distributions of $M(x)$ at each grid point x_i, then plot the median and 5% and 95% quantiles of the distributions. Even though statistical dependence among different spatial locations has been marginalized away, these profiles reflect an envelope of uncertainty in the inverse solution. In all cases, uncertainty in the log-diffusivity is greatest near the boundaries, with some additional rise near the center of the domain. All of the results presented here are based on 6×10^5 MCMC samples; we find negligible change in the estimated moments and quantiles with further iterations.

In three of the four cases above (the linear, sinusoidal, and random-draw targets), the posterior mean and median are good estimates of the true profile; the true log-diffusivity is generally contained within the credibility intervals bounded by the marginal quantiles. Mismatch with the true profile may be ascribed to limited sensor resolution (in space and in time), noise in the data, and the interaction of these conditions with the physics of the forward model. Reducing the noise in the data or introducing more finely-spaced sensors then yields closer agreement with the true profile. This is demonstrated in Fig. 6, which compares the posterior mean, conditioned on several different data sets, to the true random-draw target profile.

With the well-shaped target (Fig. 4), however, the inferred profile is smoother than the true profile. While the location of the well ($0.4 < x < 0.7$) may be surmised from the posterior, the true profile does not fall entirely within the marginal quantiles. Here, information encoded in the prior is actually *inconsistent* with the well-shaped log-diffusivity. Even with a small correlation length, a GP prior with a Gaussian covariance encodes significant smoothness, assigning very small probability to sharp variations. The posterior distribution reflects this belief in the character of the log-diffusivity profile. To obtain more appropriate reconstructions and credibility intervals in this case, the prior distribution must be chosen more carefully. Tarantola [5] suggests that if discontinuities are expected, their geometric properties should enter explicitly into the parameterization of the inverse problem. One may also construct *structural priors*, typically Gaussian but not isotropic or stationary, that encode the location and geometry of non-smooth features [4,67,68].

Since the full posterior is a distribution on $n + 1$-dimensional space, it contains much more information than can be shown in Figs. 2–4. Consider, for instance, the change in the covariance of $M(x)$ from the prior to the posterior. Computing $\text{Var}[\mathbf{m}] = \text{Cov}(M(x_i), M(x_j))$ requires marginalizing over the hyperparameter θ. The prior marginal, with density $p(\mathbf{m}) = \int_0^\infty p(\mathbf{m}|\theta)p(\theta)d\theta$, is a multivariate t-distribution; its covariance can be obtained analytically as $\beta\Sigma/(\alpha - 1)$ for $\alpha > 1$. The posterior covariance is estimated numerically from the MCMC samples. Fig. 5(a) shows the prior covariance with $L = 0.3$; in this case only, we put $\alpha = 3$ and $\beta = 2$ so the magnitude of the marginal prior covariance is well-defined. Fig. 5(b) shows the corresponding posterior covariance, again with $\alpha = 3$ and $\beta = 2$, conditioned on the noisy data vector used to infer the random-draw target in Fig. 2. The posterior covariance clearly reflects a nonstationary process, and its overall scale is more than an order of magnitude smaller than the prior covariance. The diagonal of the posterior covariance is analogous to the square of the standard deviation in Fig. 2(a). Decay of the covariance away from the diagonal reflects the character of spatial variation in the log-diffusivity profiles comprising the inverse solution.

It is important to note that, because the forward operator G is nonlinear, the posterior distributions shown here (whether 1-D marginals or full joint distributions) are not in general Gaussian or even symmetric.

3.3. Reduced-dimensionality inversion

Now we pursue a reduced-dimensionality solution of the inverse problem by exploring the posterior distribution of the weights c_k of a finite number of K–L modes, as described in Section 2.4. First, we must determine the eigenfunctions and eigenvalues appearing in the K–L expansions. For the Gaussian covariance kernel (38) on $D = [0, 1]$, there is no analytical
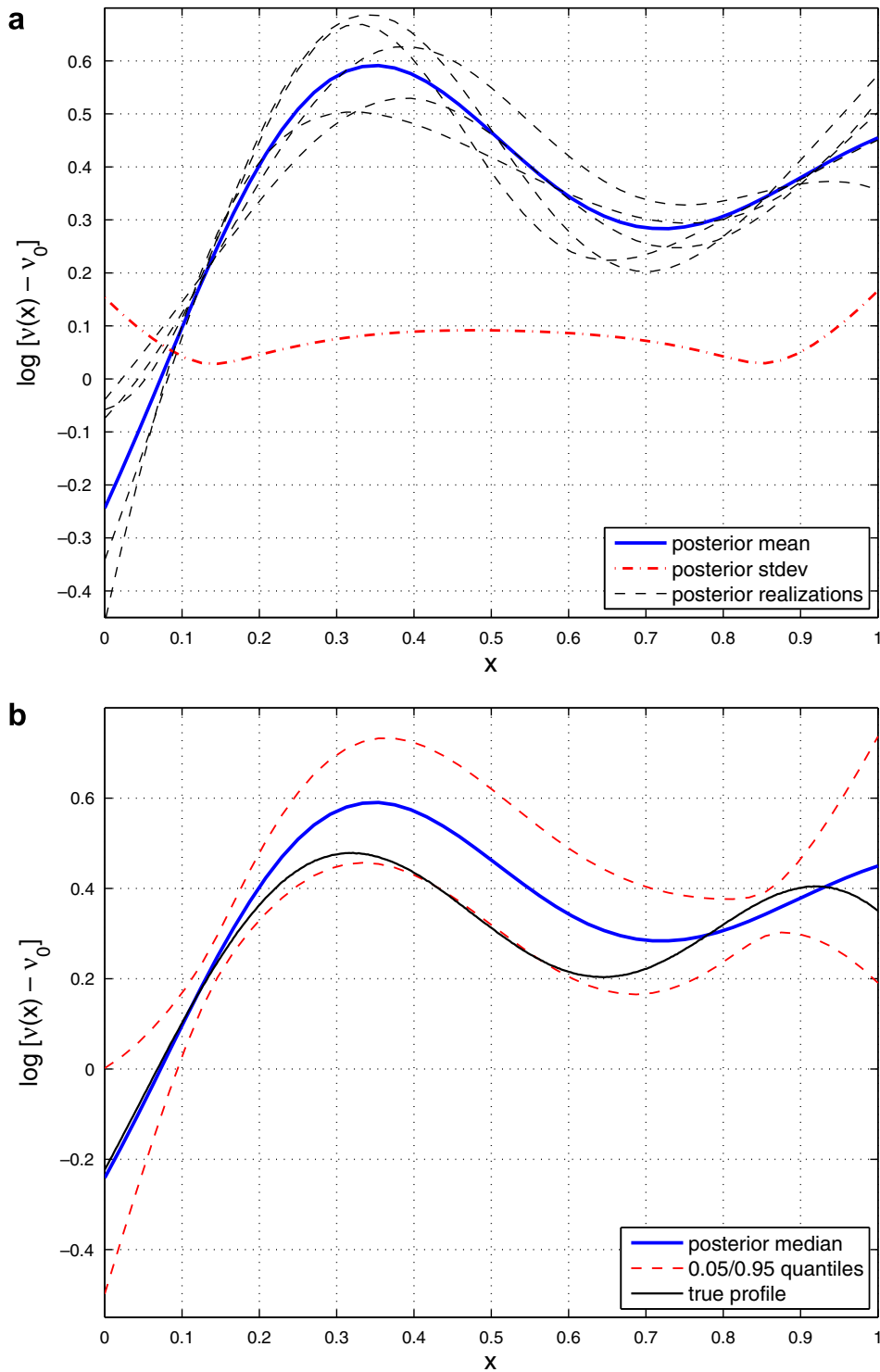
Fig. 2. Grid-based inversion of a log-diffusivity profile randomly drawn from a Gaussian process with $L = 0.3$ and scale pa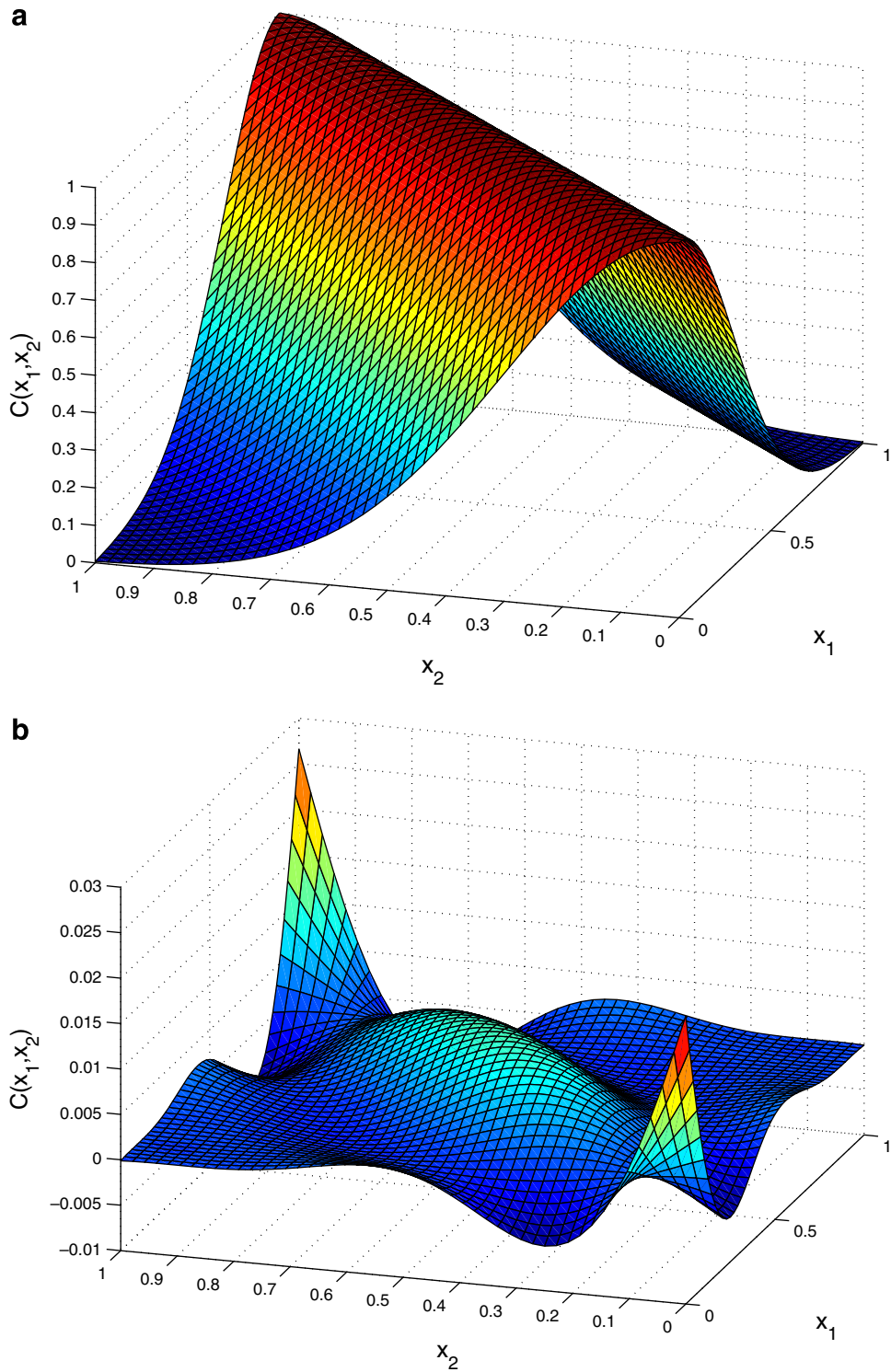rameter $\theta = 1.0$: (a) Mean, standard deviation, and five posterior realizations; (b) Median, 1-D credibility intervals, and true profile.

solution for the spectral decomposition of the corresponding integral operator. Instead, we solve the integral Eq. (8) numerically, using the Nystrom method [69] with a Gauss–Legendre quadrature rule and a LAPACK solver for the first K eigenvalues and eigenvectors of the resulting real symmetric matrix.
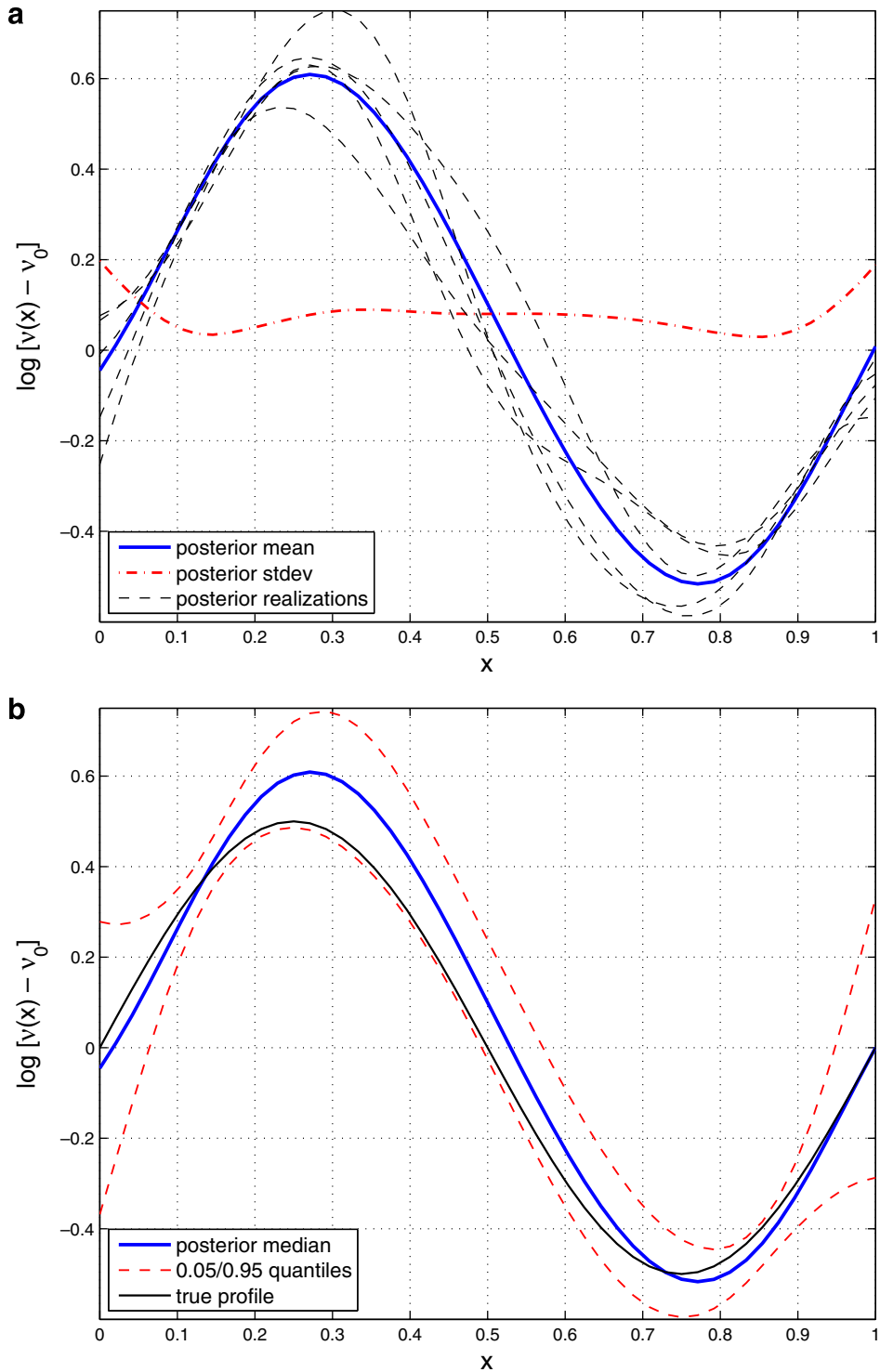
Fig. 3. Grid-based inversion of a sinusoidal log-diffusivity profile: (a) Mean, standard deviation, and five posterior realizations; (b) Median, 1-D credibility intervals, and true profile.

The hyperparameter θ is treated as in the previous section. The scale of the covariance kernel does not affect the eigenfunctions $\phi_k(x)$; it simply multiplies the eigenvalues λ_k. Thus, we can compute the K–L expansion of $M(x) \sim \mathcal{GP}(0, C_\theta)$ while fixing $\theta = 1$, and let the hyperparameter control the prior variance of the random variables c_k, $\mathbf{c} \sim N(\mathbf{0}, \theta\mathbf{I})$. The posterior density in (21) is re-written as follows:
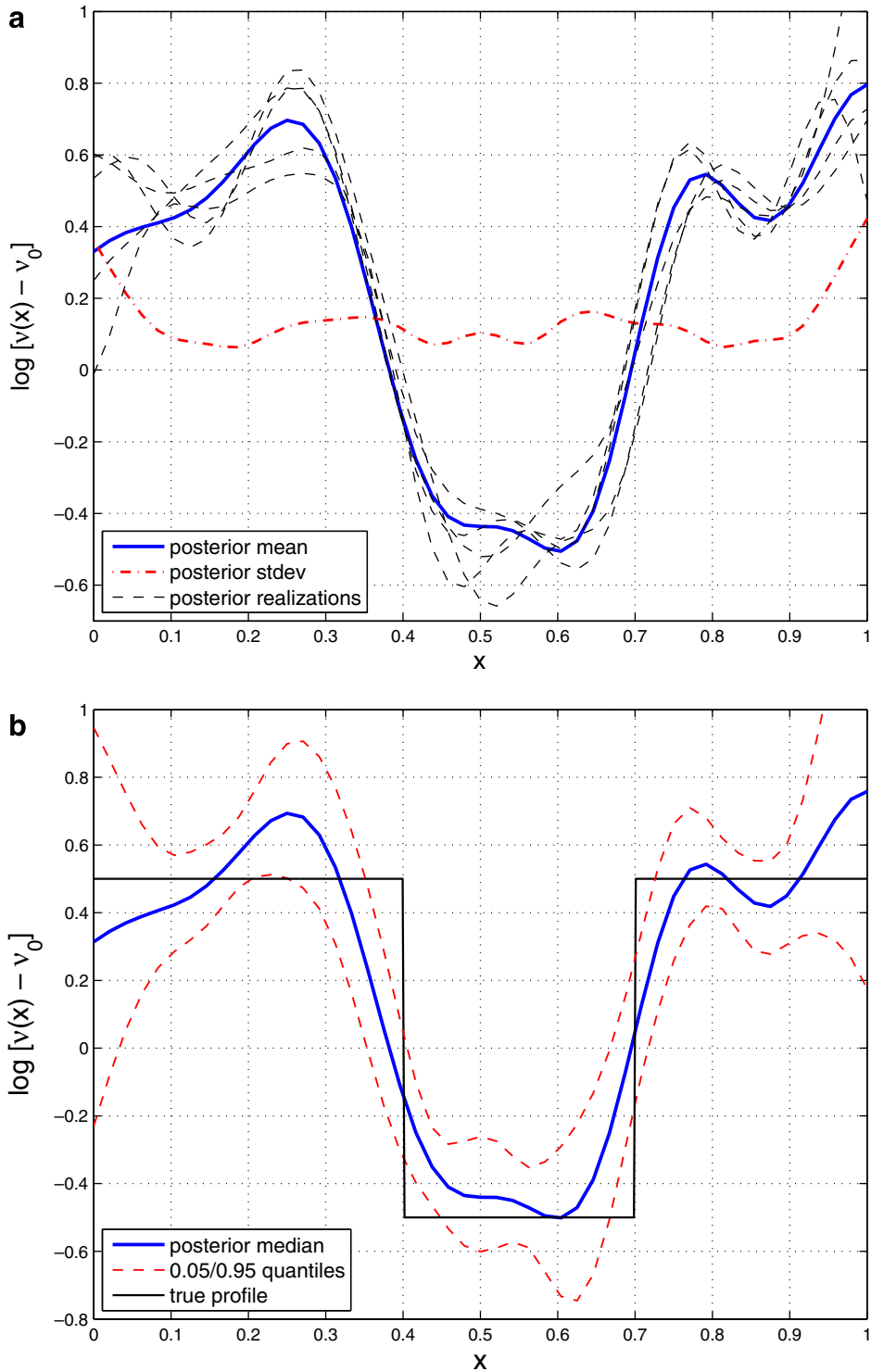
Fig. 4. Grid-based inversion of a well-shaped log-diffusivity profile: (a) Mean, standard deviation, and five posterior realizations; (b) Median, 1-D credibility intervals, and true profile.

$$p(\mathbf{c}, \theta|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{c})p(\mathbf{c}|\theta)p(\theta) \propto \exp\left(-\frac{(\mathbf{d} - \mathbf{G}(M_K(\mathbf{c})))^T(\mathbf{d} - \mathbf{G}(M_K(\mathbf{c})))}{2\varsigma^2}\right) \times \theta^{-\frac{K}{2}}\exp\left(-\frac{\mathbf{c}^T\mathbf{c}}{2\theta}\right) \times \theta^{-\alpha-1}\exp\left(-\frac{\beta}{\theta}\right), \qquad (43)$$

Fig. 5. Change in the covariance of $M(x)$ from the prior to the posterior, for inference of the random-draw target. Both (a) and (b) reflect marginalization over the scale parameter θ. In this example only, the hyperprior is $\theta \sim IG(3, 2)$: (a) Prior covariance, $L = 0.3$; (b) Posterior covariance with the random-draw profile.

where the forward model **G** now maps functions $M : D \rightarrow \mathbb{R}$, representing the log-diffusivity, to $\mathbb{R}^{mn}$. $M_K(\mathbf{c})$ denotes the K-term K–L expansion (13) evaluated at **c**:

$$M_K(x; \mathbf{c}) = \sum_{k=1}^{K} \sqrt{\lambda_k} c_k \phi_k(x). \tag{44}$$

MCMC simulation from the posterior distribution proceeds as in Section 3.2: random-walk Metropolis updates are used for the parameters representing the field (here, $\mathbf{c}$) and Gibbs updates are used for the hyperparameter θ. Figs. 7–9 show the results of K–L-based inversion with various target profiles. Results obtained with a sufficiently large number of K–L modes K become indistinguishable from the grid-based inverse solutions. As expected, shorter prior correlation lengths require a larger number of K–L modes for accurate inverse solutions. Consider the remainder of the total *prior* variance integrated over the domain D (15), i.e., $1 - \sum_{k=1}^{K} \lambda_k$, shown in Fig. 10. This quantity decays exponentially fast with increasing K, reflecting the decay of the eigenvalues of the Gaussian covariance kernel (38), but eigenvalues corresponding to large-L kernels decay more quickly than those corresponding to small-L. Since the distributions of c_k are altered by conditioning on $\mathbf{d}$, the relative importance of the K–L modes changes in the posterior, but still decays at larger index. Fig. 11 compares MCMC estimates of the posterior moments of $M(x)$, obtained via grid-based inversion, to MCMC-estimated posterior moments of $M_K(x)$ obtained via K–L-based inversion with varying numbers of modes. In particular, we compute the L^2 distance between estimates of the posterior mean, $\|\hat{\mu}(M_K(x)) - \hat{\mu}(M(x))\|_2 = \left(\int_D |\hat{\mu}(M_K(x)) - \hat{\mu}(M(x))|^2 dx \right)^{1/2}$, and the L^2 distance between estimates of the posterior standard deviation, $\|\hat{\sigma}(M_K(x)) - \hat{\sigma}(M(x))\|_2$. Differences between these posterior estimates at first fall rapidly with increasing K, but then plateau. The plateau region appears at smaller K for the large-L cases (e.g., the line profile) and at larger K for the small-L cases (e.g., the well profile), and reflects the fact that differences between moments of the grid and K–L-based inverse solutions eventually become comparable to the variability of the MCMC estimates themselves. Indeed, each realization of a Markov chain yields slightly different estimates of the posterior mean and standard deviation, and differences among these realizations account for continued "jitter" in the plateau regions. To illustrate the spread in these estimates, we have plotted results from additional realizations of the MCMC chain at $K = 6$, 8, and 10, for the random-draw target. (Each realization corresponds to a distinct choice of random seed.) Differences in the magnitude of the plateau region associated with each target profile reflect the fact that variance of an MCMC estimate is dependent on the variance of the actual quantity being estimated [70]; the posterior associated with the well-shaped target, for instance, shows much larger variances than the posterior associated with the linear target.

Further insight into the contribution of each K–L mode to the inverse solution is obtained by examining boxplots of the posterior marginals of the mode strengths. In particular, we consider marginal densities of $\sqrt{\lambda_k} c_k$, the *scaled* contribution of each K–L mode. The K–L eigenfunctions multiplied by these factors each have an L^2 norm of unity, and thus the relative importance of each eigenfunction—e.g., each mode's contribution to the mean and spread of the posterior—is captured by
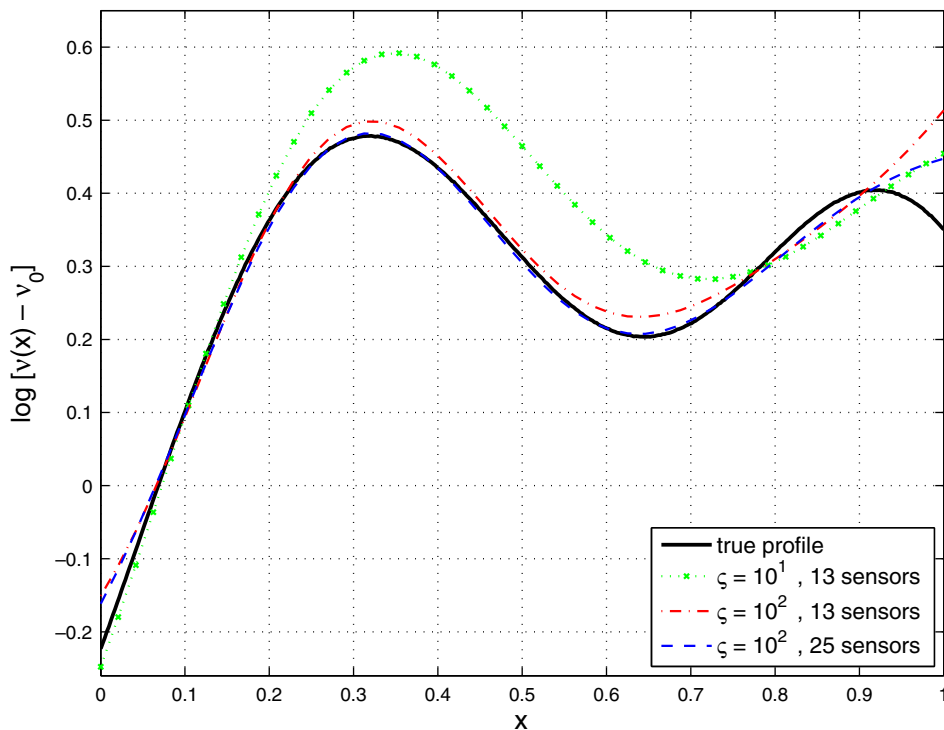

Fig. 6. Effect of measurement noise (variance ς^2) and more finely-spaced sensors, for inversion of the random-draw target. Posterior means (dotted and dashed lines) are compared to the true profile (solid line).
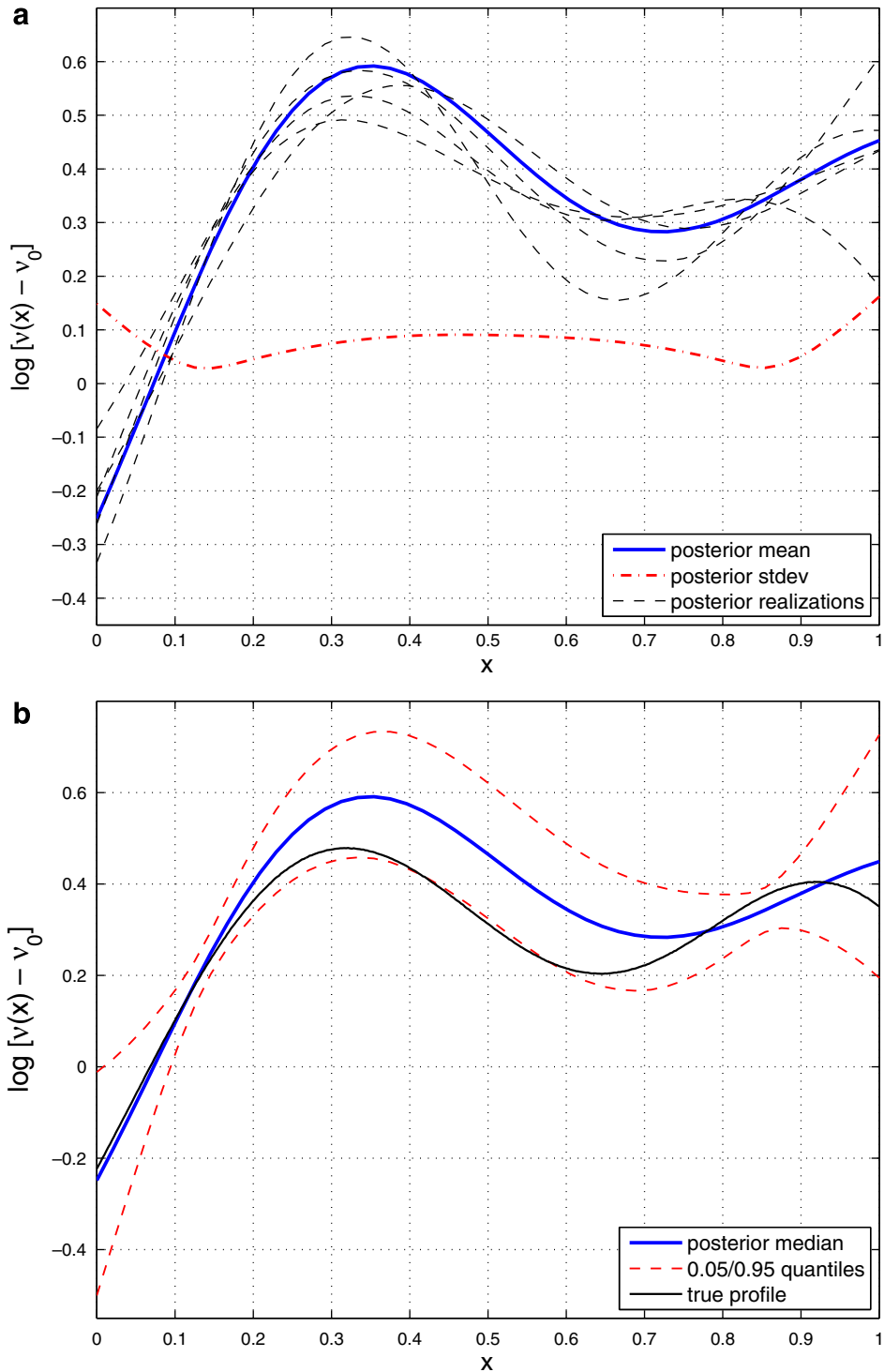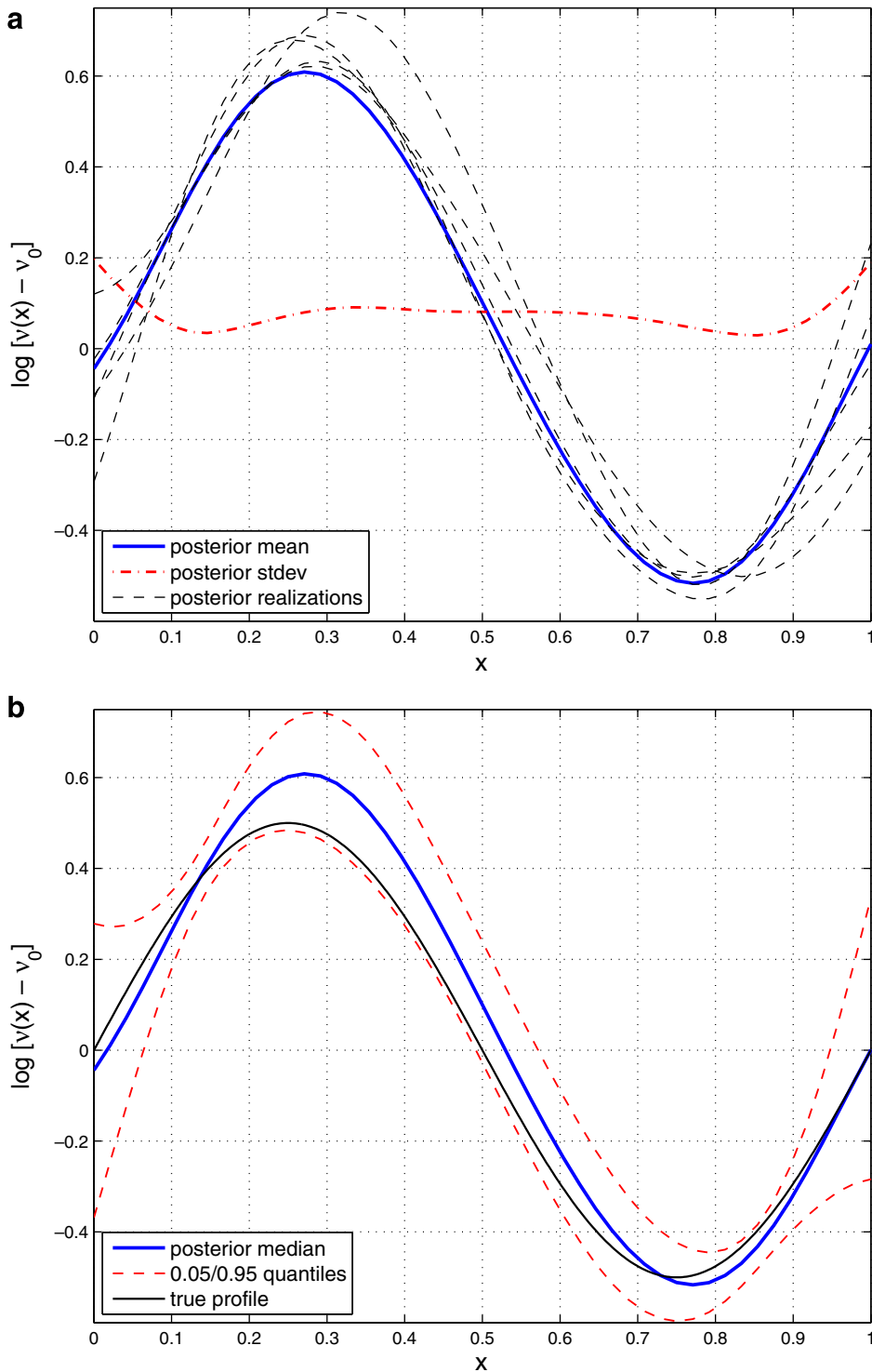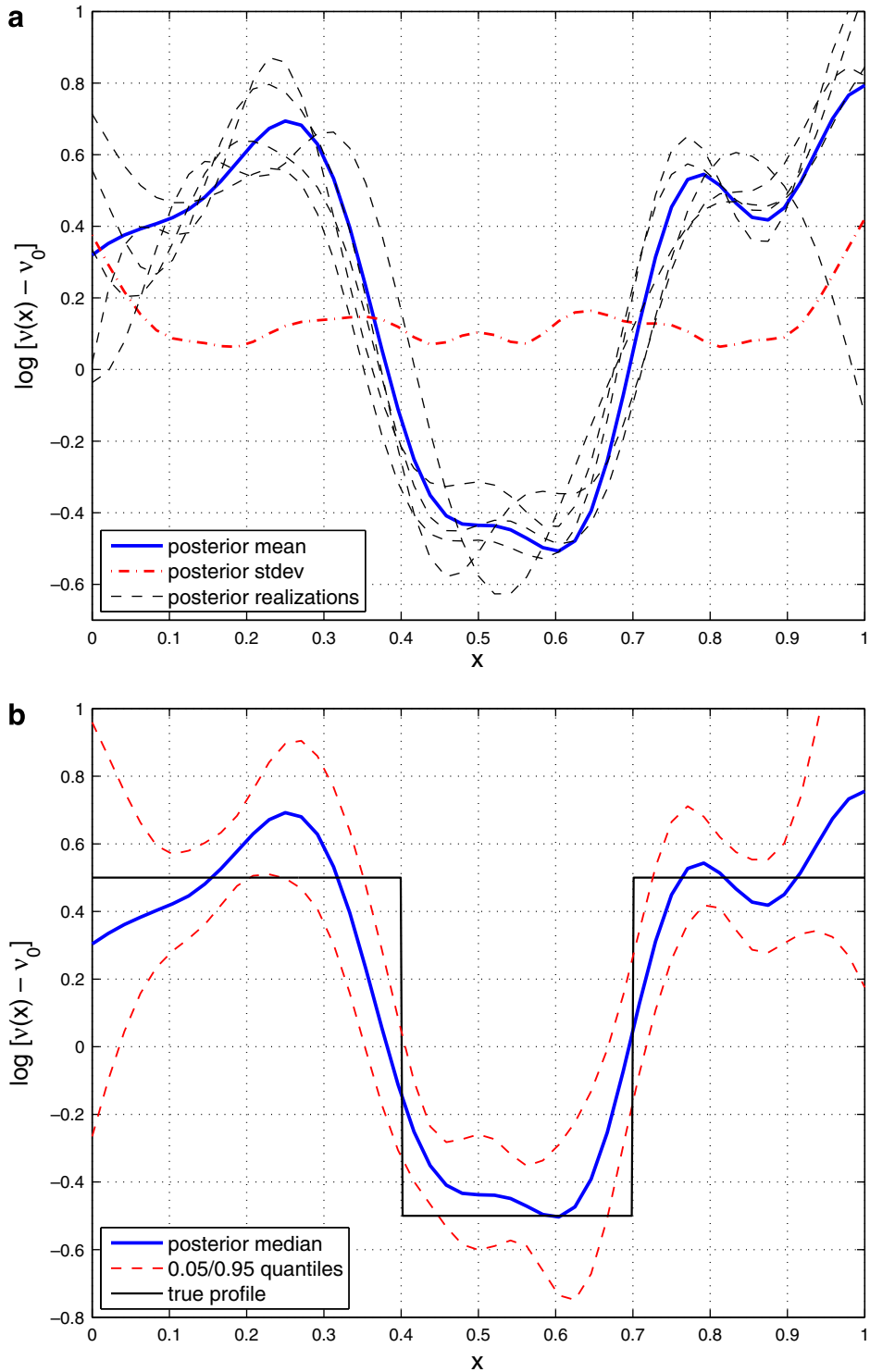
Fig. 7. K–L-based inversion of the random-draw log-diffusivity profile, $K = 6$: (a) Mean, standard deviation, and five posterior realizations; (b) Median, 1-D credibility intervals, and true profile.

the boxplots in Fig. 12. Results are reported for the random-draw and well-shaped targets. The horizontal line at the center of each box marks the median of the posterior marginal $p(\sqrt{\lambda_k}c_k|\mathbf{d})$; the extent of each box marks the 25% and 75% quantiles of the posterior marginal; and the vertical lines span the entire range of the MCMC samples. The importance of each mode does not decrease strictly with k. For instance, K–L mode $\phi_4^{L=0.3}(x)$ contributes more to the posterior of the random-draw target

Fig. 8. K–L-based inversion of the sinusoidal log-diffusivity profile, $K = 10$: (a) Mean, standard deviation, and five posterior realizations; (b) Median, 1-D credibility intervals, and true profile.

than $\phi_2^{L=0.3}(x)$ and $\phi_3^{L=0.3}(x)$; with the well-shaped target, mode $\phi_9^{L=0.1}(x)$ contributes more to the posterior than $\phi_7^{L=0.1}(x)$. At sufficiently large index, however, the (exponential) decrease of the λ_k takes over: variances of the mode strengths decrease and the medians tend towards zero.

Fig. 9. K–L-based inversion of the well-shaped log-diffusivity profile, $K = 15$: (a) Mean, standard deviation, and five posterior realizations; (b) Median, 1-D credibility intervals, and true profile.

Spatial correlations are also well-reproduced by the reduced-dimensionality inverse solution. Consider contours of the posterior covariance $\text{Cov}[M(x_1), M(x_2)]$, shown in Fig. 13. Solid lines are obtained via the grid-based inversion described in Section 3.3, while dashed lines represent the posterior covariance computed with K K–L modes. Very close agreement is ob-
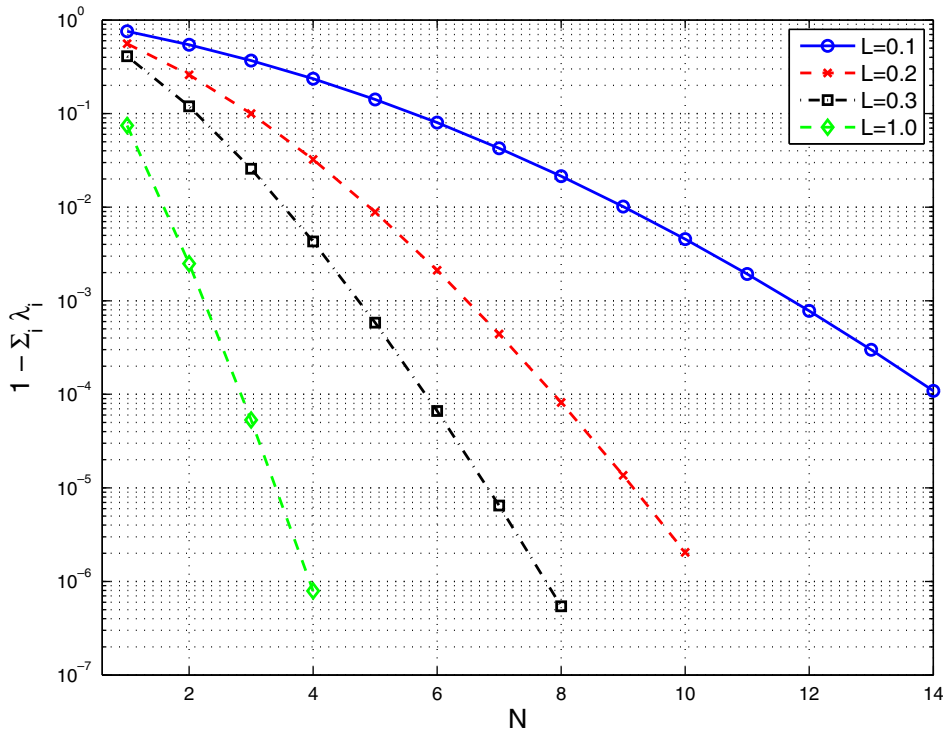
Fig. 10. Decay of K–L eigenvalues with different prior correlation lengths L. Vertical axis shows the missing fraction of the prior variance, $1 - \sum_i^K \lambda_i$, versus K.

served with increasing K. This result may be somewhat surprising, as $\phi_k(x)$ are not eigenfunctions of the posterior covariance and thus not an optimal basis for posterior in the sense of a K–L representation. Nonetheless, a modest number of these modes is able to capture the posterior covariance.

Eigenfunctions aside, the ability to reproduce the posterior covariance also depends on the emergence of correlations in the joint posterior of the K–L mode strengths c_k. Fig. 14 shows all of the one- and two-dimensional posterior marginals of $(\theta, \mathbf{c})$ conditioned on the random-draw target. Significant correlations are apparent among the lower-indexed modes, and between these modes and the hyperparameter θ. Higher-indexed modes, on the other hand, appear uncorrelated—and based on the shape of their 2-D marginals, mutually independent. We examine the 1-D marginals of these modes more closely in Fig. 15. The solid lines are conditional densities $p(c_k|\mathbf{d}, \theta)$ extracted from the posterior via kernel density estimation and plotted for different values of θ. The dashed lines are the corresponding conditional prior densities $p(c_k|\theta)$; recall that these are simply $c_k \sim N(0, \theta)$. The posterior densities of c_6, shown in Fig. 15(a), are shifted and somewhat narrower than their priors. (Much more dramatic departures from the prior are observed for c_1 through c_5.) Conditional posteriors of c_8 in Fig. 15(b), on the other hand, match the prior conditionals quite closely. Similarly close matching is observed for modes c_9, c_{10}, and so on. This pattern leads us to conjecture that, at sufficiently large m, the posterior distribution of K–L modes $c_{k \geqslant m}$ approaches the prior:

$$p(c_m, c_{m+1}, \ldots |\mathbf{d}, \theta) \to \prod_{k \geqslant m} p(c_k|\theta). \tag{45}$$

This conjecture is consistent with the decay of λ_k at large k; since higher-index modes have smaller λ_k, these modes should have less impact on the predicted $u(x, t)$ and on the data $\mathbf{d}$. Absent conditioning on the data, these modes revert to being independent and conditionally Gaussian. Section 3.5 revisits this issue in the case of coarser data.

A practical benefit of using K–L modes to compute the inverse solution is more efficient MCMC sampling. Fig. 16(a) plots $\gamma(s)/\gamma(0)$, the empirical autocorrelation at lag s, for several components of the Markov chain used to explore $p(\mathbf{z}, \theta|\mathbf{d})$ (i.e., grid-based inversion) and several components of the chain exploring $p(\mathbf{c}, \theta|\mathbf{d})$ (K–L-based inversion). In both cases, the noisy data vector $\mathbf{d}$ is obtained from the random-draw target. With the grid-based posterior, lower-index z_i multiply columns nearer the left side of the Cholesky factor $\mathbf{L}$, which contain a larger number of non-zero entries (specifically, $n - i + 1$). These modes mix less efficiently than their larger-i counterparts, even though we have individually tuned the proposal width of each single-component Metropolis update. Mode strengths c_k and the hyperparameter θ of the K–L-based posterior show much more rapid decay of their autocorrelations, reflecting improved mixing. In Fig. 16(b), we transform the c_k and z_i into realizations of $M(x)$ and compare the chain autocorrelations at particular spatial locations. With the grid-based posterior,
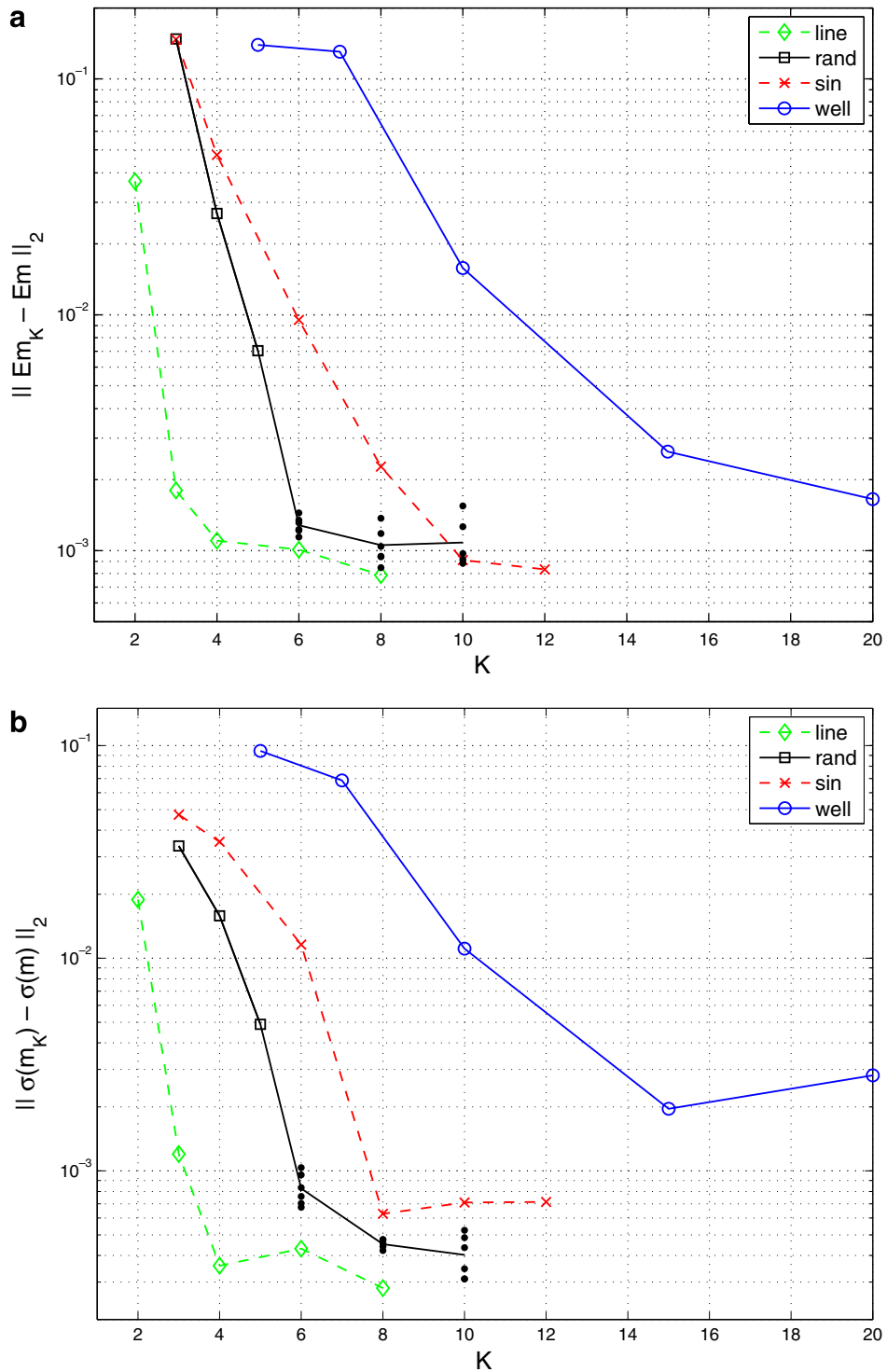
Fig. 11. Differences in posterior moments computed via grid-based inversion and K–L based inversion, versus K. At $K = 6$, 8, and 10, repeated symbols ($\bullet$) correspond to multiple MCMC simulations from the posterior of the random-draw target: (a) $L^2(D)$-norm of differences in the posterior mean; (b) $L^2(D)$-norm of differences in the posterior standard deviation.

mixing improves toward the right side of the physical domain D, as the value of the solution at larger x is influenced by a larger number of modes z_i—and in particular, better-mixing modes. The autocorrelations of K–L-based $M(x_i)$ still decay more rapidly, however.
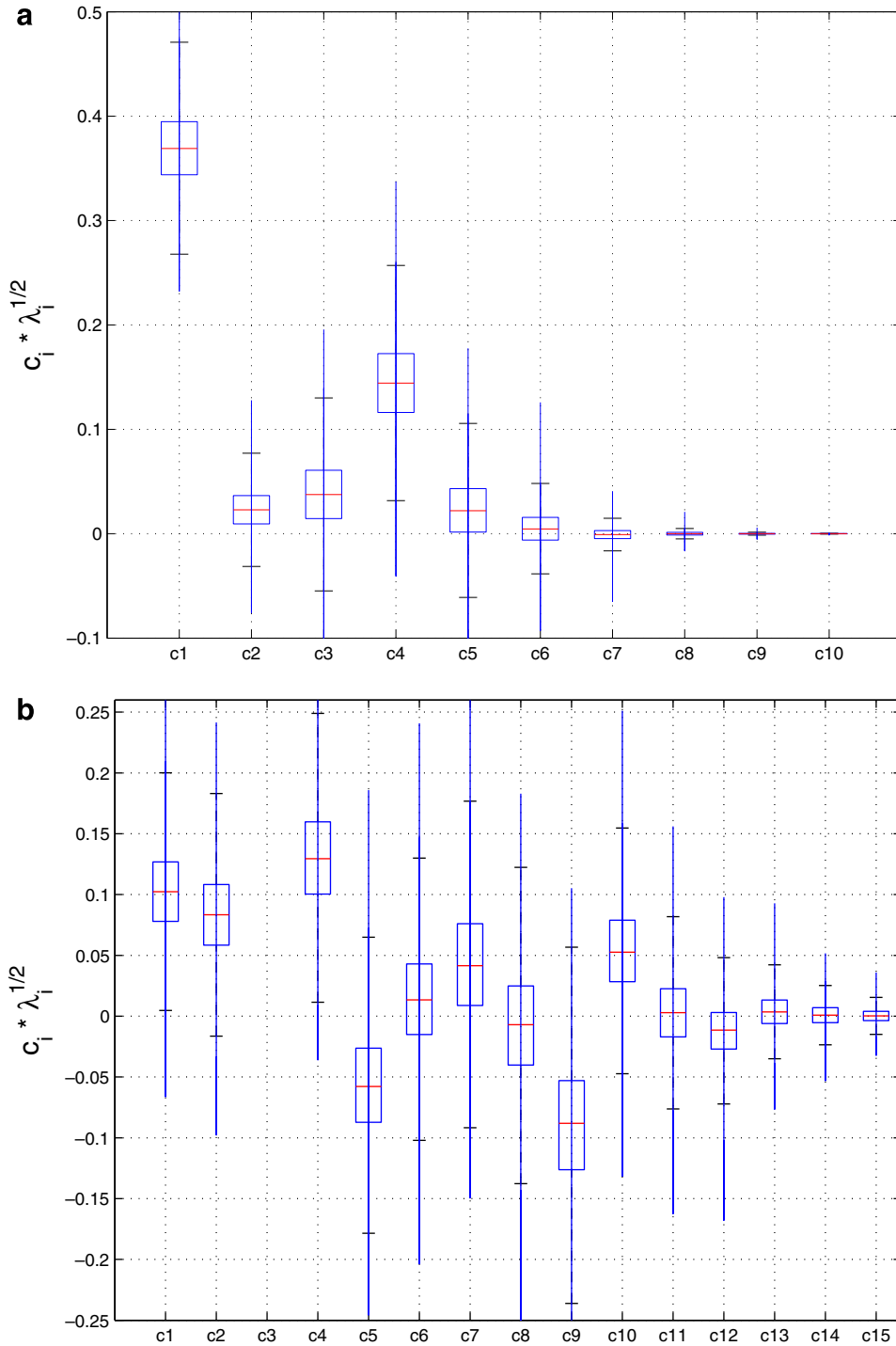
Fig. 12. Boxplot of the posterior marginals of the K–L mode strengths c_i, scaled by $\sqrt{\lambda_i}$: (a) Random-draw target; (b) Well-shaped target.

3.4. Reduced-dimensionality inversion with polynomial chaos acceleration

We now construct a stochastic forward problem whose solution captures the output of the deterministic forward model over the support of the prior distribution, and use this solution to formulate a surrogate posterior distribution. The resulting scheme is intended to achieve an accelerated, reduced-dimensionality Bayesian solution of the inverse problem, as described in Section 2.5.2.
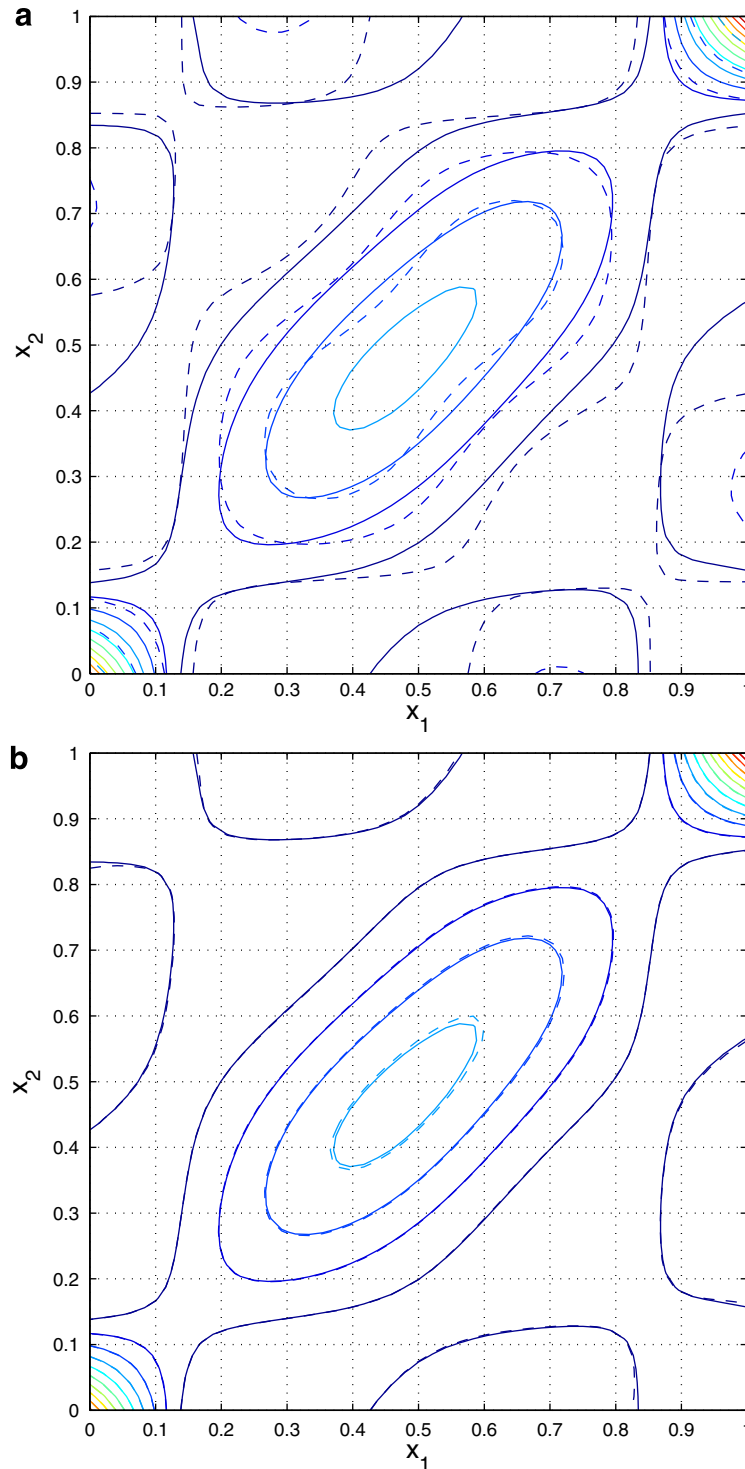
Fig. 13. Contours of the posterior covariance, $\text{Cov}[M(x_1), M(x_2)]$. Solid lines are obtained via grid-based inversion; dashed lines are obtained via reduced-dimensionality inversion with K K–L modes. All are for inference of the random-draw target: (a) $K = 4$; (b) $K = 8$.

We begin with the K–L representation of the log-diffusivity field and the hierarchical priors on c_k derived in the previous section. Following the procedure in Section 2.5.2, we must introduce an invertible/differentiable transformation $\mathbf{c} = \mathbf{g}(\xi)$, where ξ is a vector of i.i.d. random variables underlying a PC basis. Here, we let $\mathbf{g}$ be a simple scaling transformation

$$\mathbf{c} = \mathbf{g}(\xi) = \varpi\xi, \tag{46}$$
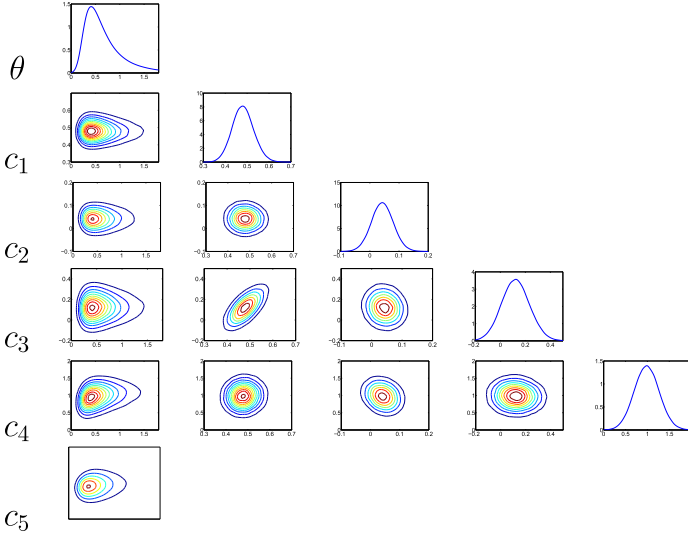
where ξ is a K-vector of independent standard normal random variables and ϖ is a free parameter to be specified. This transformation is just a first-order Gauss–Hermite PC expansion, defining the uncertainty that we propagate through the forward model. Since the prior distribution of each c_k has support over the entire real line, as does the $N(0,1)$ distribution of each ξ_k, we have considerable freedom in choosing $\varpi > 0$; any choice of ϖ will map the range of ξ onto the range of $\mathbf{c}$ [6]. The choice is particularly open since θ, the prior variance of c_k, is itself unknown. Here we will fix $\varpi^2 = 0.5$, which is the mode of the hyperprior $p(\theta)$.

Together, (37), (44), and (46) define a stochastic forward problem. The input is the Gaussian random field $M(x) \equiv \log[v(x, \xi(\omega)) - v_0]$ represented with a truncated K–L expansion, and the outputs are random variables $u(x_i, t_j, \omega)$ giving the value of the scalar field at each measurement location and time. We write these random variables in terms of their PC expansions $u(x_i, t_j, \omega) = \sum_l u_l^{ij} \Psi_l(\xi(\omega))$, and collect them in a vector $\widetilde{\mathbf{G}}(\xi) \equiv (u(x_1, t_1; \xi), \ldots, u(x_m, t_n; \xi))$. Solving the stochastic forward problem—i.e., using Galerkin projection to compute the coefficients u_l^{ij}—requires transforming the input log-diffusivity into an actual diffusivity $v(x, \omega)$. The Gauss–Hermite PC representation of this log-normal process may be evaluated analytically [71,72]. Recall that the multivariate polynomial functionals comprising the PC basis are given by the tensor product of one-dimensional polynomials $\psi_i(\xi)$, here Hermite polynomials of order i. Each multivariate polynomial Ψ_l is associated with a multi-index $\alpha^l = (\alpha_1^l, \ldots, \alpha_K^l) \in \mathbb{N}^K$, where $\sum_{k=1}^K \alpha_k \leqslant p$:

$$\Psi_l(\xi) = \prod_{k=1}^K \psi_{\alpha_k^l}(\xi_k). \tag{47}$$

The diffusivity $v(x, \omega) = v_0 + \exp(M_K(x, \xi(\omega)))$ then has a PC expansion $v(x, \omega) = \sum_l^P \tilde{v}_l(x) \Psi_l(\xi)$ with coefficients

$$\tilde{v}_{l=0}(x) = v_0 + e^{\sigma^2(x)/2}, \tag{48}$$

$$\tilde{v}_{l \geqslant 1}(x) = e^{\sigma^2(x)/2} \prod_{k=1}^K \frac{\left[\varpi \sqrt{\lambda_k} \phi_k(x)\right]^{\alpha_k^l}}{\sqrt{\alpha_k^l!}}, \tag{49}$$

where[4]

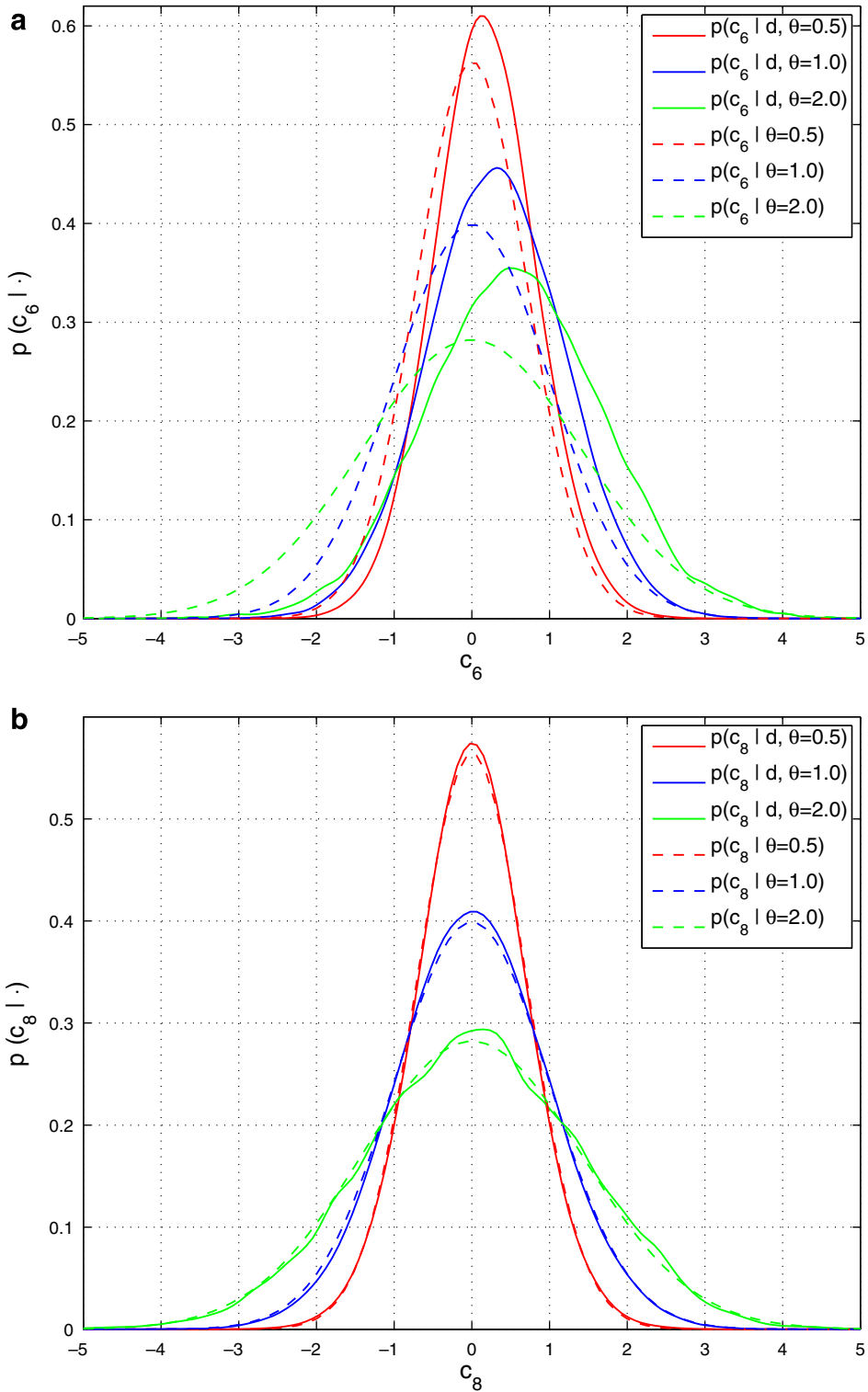[4] Note that as $K \to \infty$, we have $\sigma^2(x)/\varpi^2 \to 1$.

Fig. 15. Limiting distributions of K–L modes, 13 sensors; inference of the random-draw target: (a) c_6 prior and posterior conditional distributions; (b) c_8 prior and posterior conditional distributions.

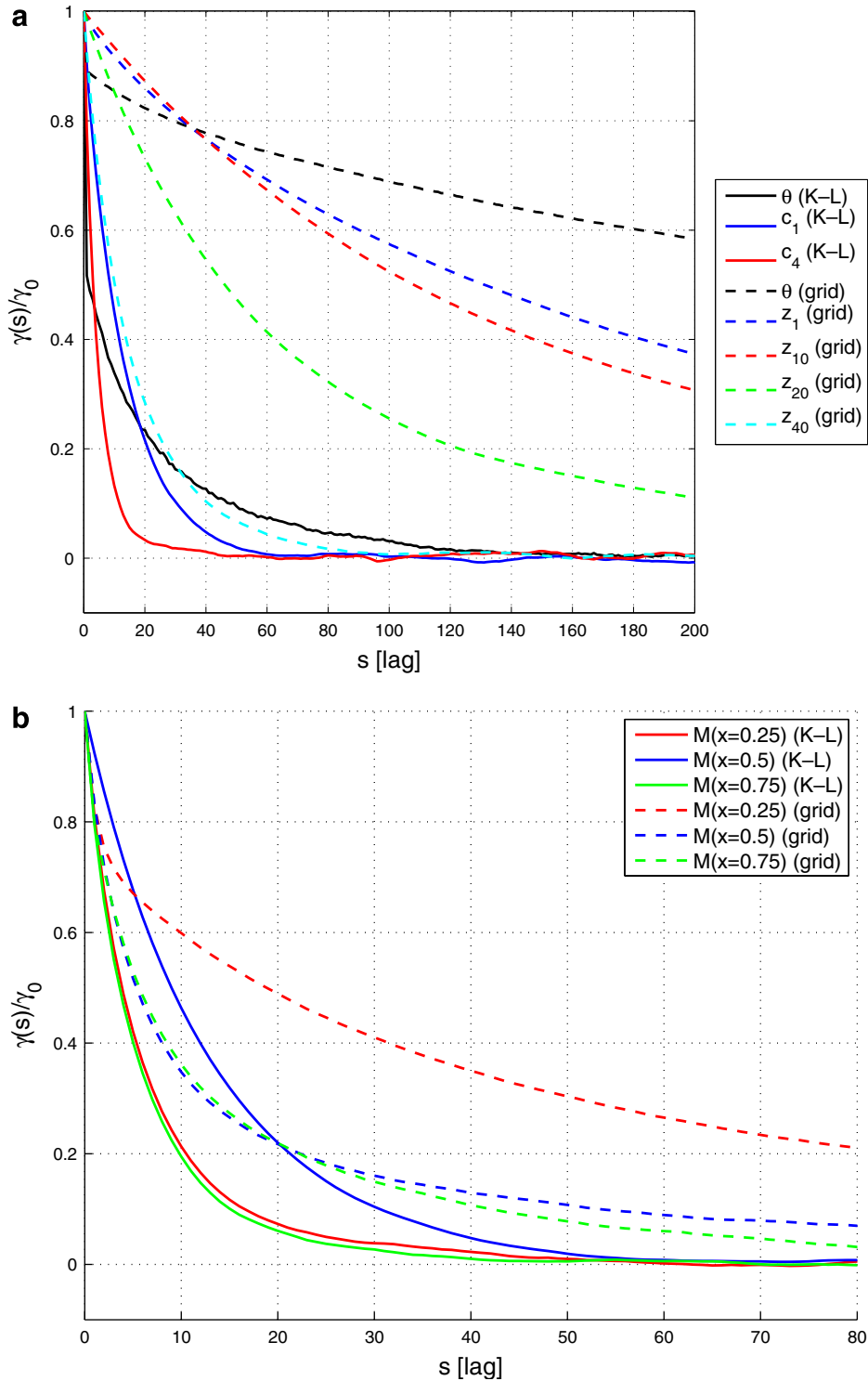$$\sigma^2(x) = \varpi^2 \sum_{k=1}^{K} \lambda_k \phi_k^2(x).$$

(50)

Fig. 16. Autocorrelation at lag s of components of the MCMC chain, and of field values $M(x)$ parameterized by the MCMC chain variables: (a) Chain variables; (b) $M(x_i)$.

This PC expansion is introduced into the transient diffusion equation (37). Using a pseudospectral stochastic Galerkin scheme coupled with the same finite-difference spatial discretization and RKC time integrator as in the deterministic problem, we obtain PC expansions for the outputs of interest $\widetilde{\mathbf{G}}(\xi)$.

The surrogate posterior density may then be written in terms of ξ:

$$p(\xi, \theta|\mathbf{d}) \propto \widetilde{L}(\xi) p_{\mathbf{c}|\theta}(\mathbf{g}(\xi)|\theta)| \det(D\mathbf{g}(\xi))|p(\theta) \propto p(\mathbf{d}|\xi) p_{\xi|\theta}(\xi|\theta) p(\theta)$$

$$\propto \exp\left(-\frac{(\mathbf{d} - \widetilde{\mathbf{G}}(\xi))^T(\mathbf{d} - \widetilde{\mathbf{G}}(\xi))}{2\varsigma^2}\right) \times \left(\frac{\theta}{\varpi^2}\right)^{-\frac{K}{2}} \exp\left(-\frac{\xi^T\xi}{2\theta/\varpi^2}\right) \times \theta^{-\alpha-1}\exp\left(-\frac{\beta}{\theta}\right). \tag{51}$$

MCMC sampling from this posterior proceeds as in the previous two sections, except that the full conditional $p(\theta|\xi, \mathbf{d})$ used for Gibbs updates is now $IG(\alpha + K/2, \beta + \varpi^2\left(\sum_{k=1}^{K}\xi_k^2\right)/2)$.

A useful diagnostic of the stochastic forward solution's fidelity is the probability density of the forward model outputs $u(x_i, t_j, \xi)$. Fixing the number of terms K in the K–L expansion of $M(x)$, these densities may be estimated in one of two ways. A direct (and computationally expensive) method is to sample $\mathbf{c}$ and solve the forward problem for each sample, forming a histogram or kernel density estimate from the resulting collection of forward model outputs. Alternatively, one can sample ξ and substitute it into the PC expansion $\widetilde{G}(\xi)$, again forming a histogram of the resulting values. Fig. 17 shows such estimates at two measurement locations and times. While a lower-order PC basis ($p = 2$) produces a somewhat poor density estimate, the probability density quickly converges to its true shape—obtained by the direct method—as p increases. Reasonable agreement is obtained even at $p = 4$. Also, note that these densities are not log-normal; their shapes reflect the nonlinearity of the forward maps from $v(x)$ to $u(x_i, t_j)$.

Figs. 18 and 19 are based on inversion of the random-draw target profile. We fix the number of K–L modes to $K = 6$, since this value provided accurate results in Section 3.3, and vary the order p of the PC basis used to solve the stochastic forward problem. MCMC samples from the surrogate posterior (51) are then transformed into realizations of $M(x)$. Even at low order, the posterior mean, standard deviation, and quantiles thus obtained are not far from their direct counterparts; at $p = 4$ and $p = 6$, we find that these summaries of the posterior distribution are visually indistinguishable from the profiles in Fig. 7. Fig. 18 quantifies differences between the posterior means/standard deviations obtained with six K–L modes and direct forward problem solutions, and those obtained with six K–L modes and sampling of the surrogate posterior. Again, we plot the L^2 norm of the differences between MCMC estimates of these quantities: $\|\hat{\mu}(M_K^p(x)) - \hat{\mu}(M_K(x))\|_2$ and $\|\hat{\sigma}(M_K^p(x)) - \hat{\sigma}(M_K(x))\|_2$. The difference in posterior mean estimates drops more than an order of magnitude from $p = 2$ to $p = 4$ and continues to fall towards $p = 6$. At both $p = 4$ and $p = 6$, we plot results from four separate realizations of the MCMC chain on $\tilde{\pi}_\xi$ (corresponding to different random seeds) in order to illustrate the variability of the MCMC estimates. In this regime, as in the case of K-convergence in Section 3.3, the distance between estimated means of the direct and surrogate posteriors becomes comparable to the standard deviations of the MCMC estimates themselves. (Both Figs. 18 and 11(a) show plateaus around 10^{-3} for the random-draw target.) Differences between the estimates of posterior standard deviation show similar dependence on p.

The surrogate posterior accurately captures spatial correlations among possible values of the inverse solution. Fig. 19 shows contours of the posterior covariance with varying p: solid lines correspond to the direct posterior, while dashed lines represent the surrogate posterior. Very close agreement is observed at $p = 4$, and this agreement improves further at $p = 6$.

Of course, the ultimate goal of introducing the stochastic forward problem and surrogate posterior is greater computational efficiency. Significant speedups were obtained with a similar approach in [6] for inverse estimation of parameters in PDEs. In the present context, even though the inverse solution is a spatial field, the pattern of computational costs is the same as in [6]. The initial cost of the scheme is offset by the computation of stochastic forward solutions, but then grows very slowly, because the cost per MCMC iteration is orders of magnitude smaller for the surrogate posterior (51) than for direct solutions of the transient diffusion equation (43). Table 1 summarizes the cost at each stage of three representative calculations: inferring the random-draw target by the methods of Sections 3.2, 3.3, and 3.4. For a fixed number of MCMC iterations, K–L based inversion with $K = 6$ is approximately one order of magnitude faster than grid-based inversion, because fewer posterior evaluations are required per MCMC iteration in the former case (using the single-component MCMC sampler). It may be possible to design more efficient MCMC methods for both cases, perhaps updating blocks of components at a time with suitably shaped proposals, but reducing the chain dimension will inevitably enable greater efficiency [73] and reduce the number of posterior evaluations required. We also note that comparing the grid-based and K–L based solutions at a fixed number of MCMC iterations is not truly fair to the K–L parameterization, since the $(\mathbf{c}, \theta)$ chain mixes more rapidly than the $(\mathbf{z}, \theta)$ chain (see Fig. 16). The Monte Carlo error obtained with 2×10^5 MCMC samples from $p(\mathbf{c}, \theta|\mathbf{d})$ is thus matched at a *larger* number of MCMC samples from $p(\mathbf{m}, \theta|\mathbf{d})$.

The third row of Table 1 shows even greater speedups, entirely independent of the MCMC implementation. The majority of the computational time in this case is spent on the stochastic forward solve. Yet we emphasize that, because it depends only on the prior and forward model, the stochastic forward solve may be performed "offline" before introducing any data. Afterwards, sampling is inexpensive. Here, the cost per posterior evaluation and per MCMC iteration is 1.8 orders of magnitude smaller than for the direct K–L based posterior. Including the time for the stochastic forward solve, inference via exploration of the surrogate reduced-dimensionality posterior is 2.3 orders of magnitude faster than exploration of the direct full-dimensional posterior, with negligible loss of accuracy.
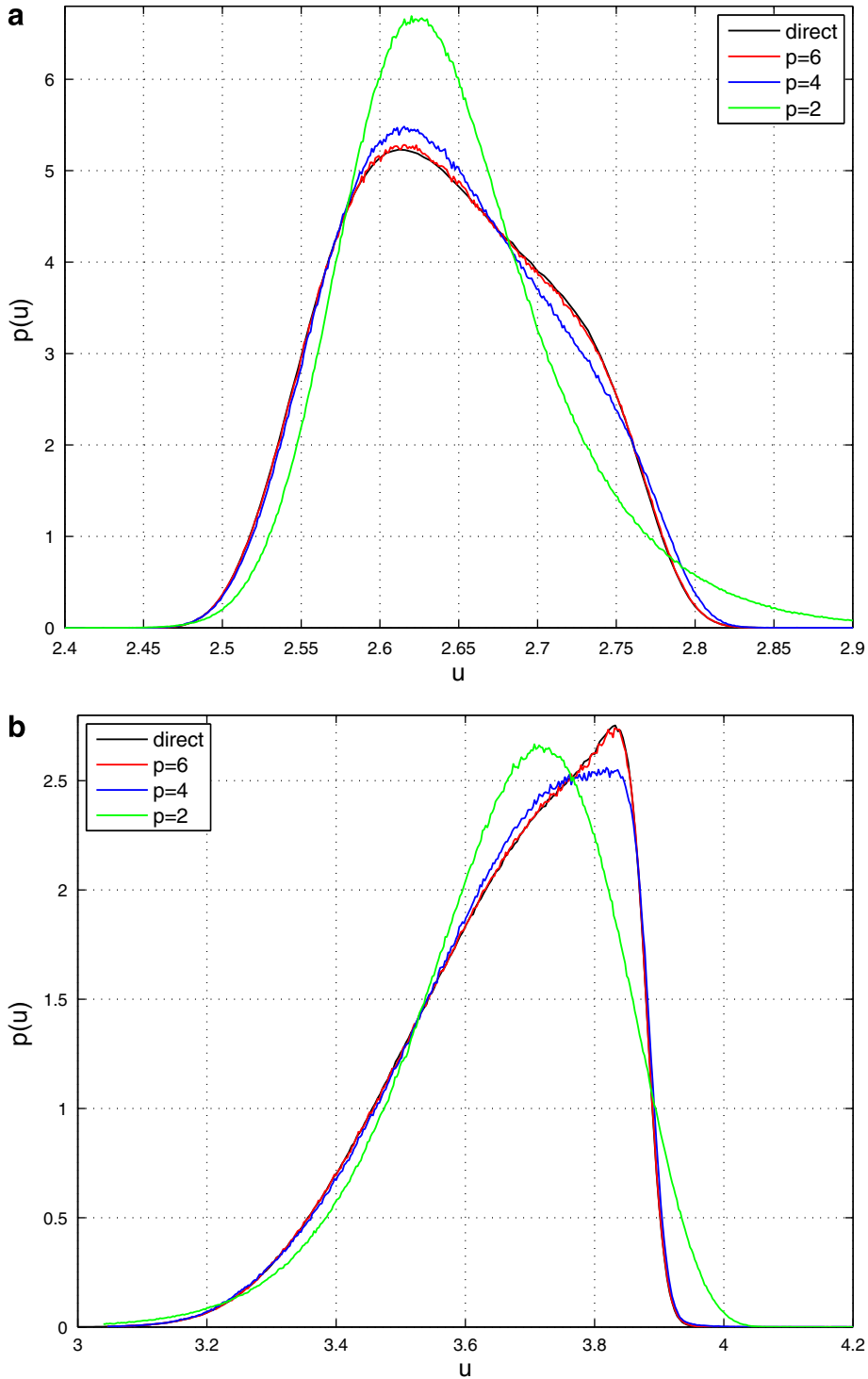
Fig. 17. Probability density of the scalar $u(x, t)$ at two measurement locations and times, given input uncertainty in the diffusivity field. Here, $v(x, \omega)$ is the log-normal random field defined in Section 3.4: (a) $u(x = 1/6, t = 0.0125)$; (b) $u(x = 5/12, t = 0.0225)$.

3.5. Data length scales

It is useful to consider the behavior of the inverse solution as one coarsens the length scale on which data is collected. For instance, what if observations of the scalar field $u(x, t)$ were limited to the boundaries of the domain, $x = 0.0$ and $x = 1.0$? We still take noisy measurements at nine successive times, spaced uniformly over the interval $t \in [0.01, 0.03]$. Fig. 20 shows the
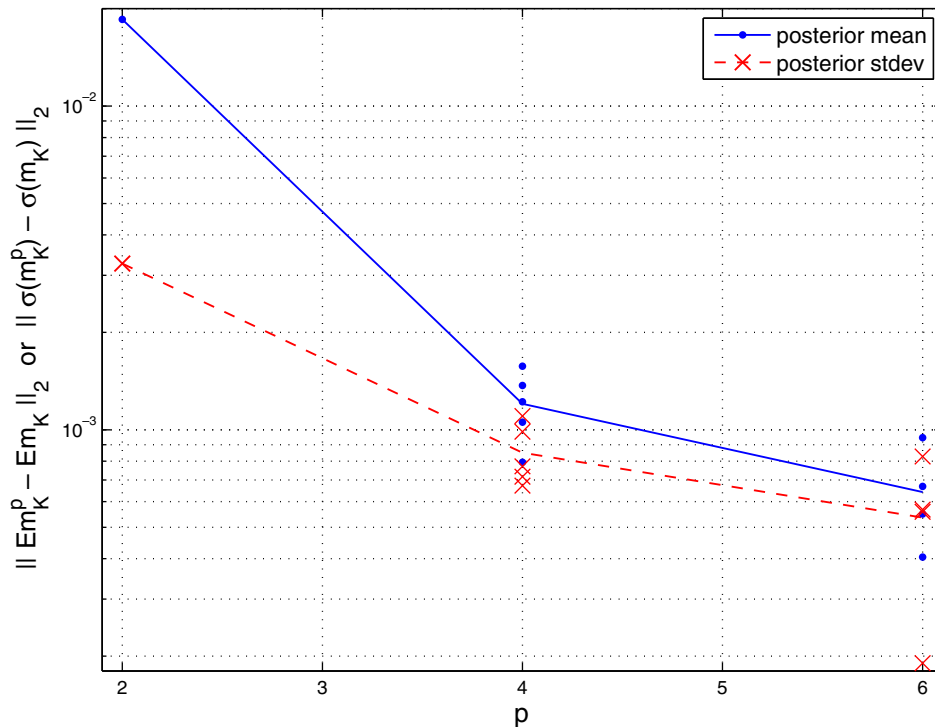
Fig. 18. Differences between means/standard deviations of $M(x)$ computed via the "direct" posterior (43) and via the surrogate posterior (51), versus p. All results are for inversion of the random-draw target with six K–L modes. At $p = 4$ and $p = 6$, repeated symbols correspond to multiple MCMC simulations from the surrogate posterior.

inverse solution corresponding to the sinusoidal target, obtained with $K = 10$ K–L modes and direct forward problem solutions as described in Section 3.3. Contrast these results with those in Fig. 8 or Fig. 3. The two-sensor results show much greater variability, particularly in the center of the domain, compared to their 13-sensor counterparts. The posterior mean and median still appear sinusoidal, along with the majority of the posterior realizations. All of these profiles are closest to the true solution near the domain boundaries. Asymmetry in the standard deviation profile may be ascribed to asymmetry in the realizations of the sensor noise perturbing observations of $u(x,t)$.

Examining the posterior distribution of $M(x)$ does not complete the story, however. In particular, the posterior distributions of the K–L mode strengths c_k exhibit interesting features when conditioned on coarser data. Fig. 21 shows boxplots of the marginal posteriors of c_k and θ, contrasting 13-sensor and 2-sensor inference of the sinusoidal target. First, we note that the hyperparameter θ and the mode strengths c_k—particularly the lower-index c_k—have narrower posterior distributions in the data-rich case. As discussed in Section 3.3, higher-index modes approach a zero-mean limiting distribution, but crucially, this tendency is observed at much lower k in the 2-sensor case. To further elucidate this point, consider the matrix of one- and two-dimensional posterior marginals, shown in Fig. 22 for two sensors. The marginal distributions of $c_6, c_7, \ldots$ do indeed appear quite similar in shape and range, and moreover, correlations among the modes weaken at larger k, becoming quite negligible for $k > 6$. We examine the limiting distributions quantitatively as in Section 3.3, plotting posterior conditionals $p(c_6|\mathbf{d}, \theta)$ in Fig. 23. The posterior conditionals of c_6 in the 13-sensor case are far from the corresponding Gaussian prior conditionals $p(c|\theta)$, but in the the 2-sensor case $p(c_6|\mathbf{d}, \theta)$ matches the Gaussian prior $N(0, \theta)$ quite closely for various values of θ. Even closer matching is observed at higher k. These observations lead us to expand upon the conjecture of (45), by suggesting that the posterior distribution of K–L modes $c_{k \geqslant m}$ approaches the prior at *smaller* index m as the length scale of the data is coarsened.

Implications of this statement are numerous. One possibility is that when the data is scarce relative to the complexity of the model (understood here as the spatial complexity of $\log \nu(x)$ as constrained by the prior), dimensionality reduction based on the prior may be improved upon. For instance, in the present case of 2-sensor inversion of the sinusoidal profile, the fraction of the posterior standard deviation contained in modes 6 and higher (integrated over D) is 14.1%. These modes are thus important to the posterior distribution, but if their joint distribution is essentially unchanged by conditioning on the data, the corresponding c_k could be removed from the inference process. Adding unconditioned realizations of $\sum_{k=6}^{\infty} \sqrt{\lambda_k} c_k \phi_k(x)$ to posterior realizations of $M_{K=5}(x)$ would then yield samples from the full posterior.
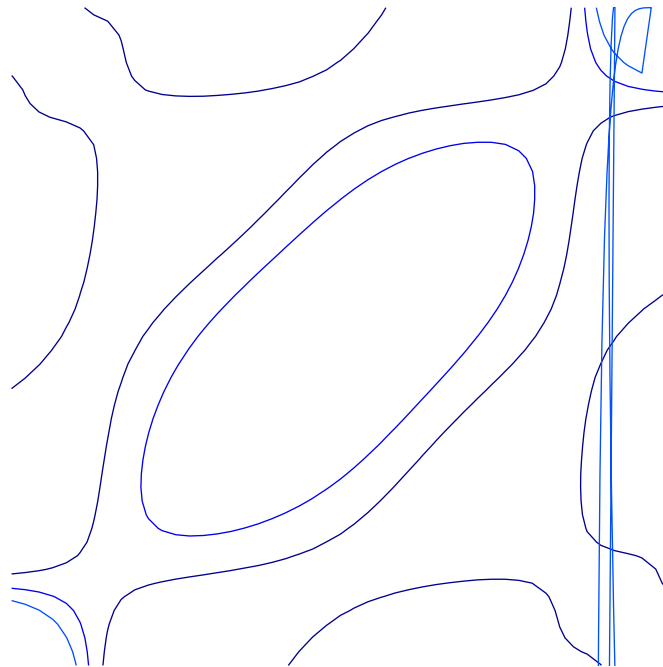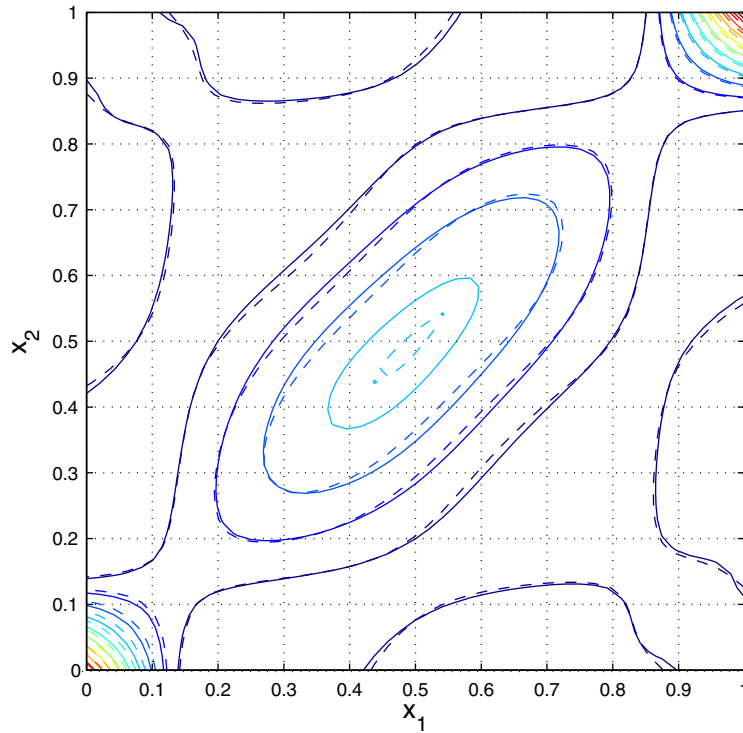
3.6. Prior and posterior hyperparameters

As a further numerical demonstration, we use the combined PC/K–L formulation to infer a log-diffusivity profile that, un-like the random-draw target, has no particular relationship to the Gaussian process prior. At the same time, we treat the

Table 1
Computational times, in seconds, for inference of the "random-draw" target profile (e.g., Fig. 2) by three different methods.

	Time for stochastic forward solve	Time per posterior evaluation	Time per MCMC iteration	Total time for inversion (2×10^5 samples)
Grid-based; $n = 49$	–	5.0×10^{-3}	2.5×10^{-1}	49,500
K–L; $K = 6$	–	5.2×10^{-3}	3.4×10^{-2}	6860
K–L+PC; $K = 6$, $p = 4$	137	9.2×10^{-5}	5.5×10^{-4}	248

prior mean μ_0, heretofore fixed at zero, as a hyperparameter and more closely analyze the posterior distributions of the hyperparameters μ_0 and θ.

The new target log-diffusivity is obtained by integrating the three-dimensional Lorenz system [74] at parameter values ($\bar{\sigma} = 10$, $\bar{\beta} = 8/3$, $\bar{\rho} = 28$) that produce chaotic behavior.[5] A portion of the trajectory of the second state variable y is extracted, rescaled to the $D = [0, 1]$ domain, and translated so that it has non-zero mean on D. We emphasize that the this procedure is simply a device for producing a "generic" continuous target profile without any particular symmetries; the dynamics of the Lorenz system in no way enter the inference procedure.

Treating the prior mean μ_0 as a hyperparameter entails only small modifications to the inference procedure described in Section 3.4. First, μ_0 must be assigned a hyperprior; we use a standard normal, $\mu_0 \sim N(0, 1)$. The mean then appears as the first term in the K–L expansion of $M(x, \omega)$: $M_K(x; \mu_0, \mathbf{c}) = \mu_0 + \sum_{k=1}^{K} \sqrt{\lambda_k} c_k \phi_k(x)$. Since this mean is now uncertain, we must add an additional stochastic dimension to the PC basis in order to solve the stochastic forward problem. A K-term K–L expansion now implies $K + 1$ stochastic dimensions. The scaling transformation in (46) can be partitioned as follows:

$$\mu_0 = \varpi_0 \xi_0 \tag{52}$$
$$c_i = \varpi_1 \xi_i, \quad i = 1 \ldots K \tag{53}$$

with $\boldsymbol{\xi} \equiv (\xi_0, \xi_1, \ldots, \xi_K)$. The PC coefficients of the log-normal diffusivity $v(x, \omega)$ are still obtained analytically with a simple modification to (49) and (50). To these equations, we add a $k = 0$ term with $\phi_0(x) = 1$, $\lambda_0 = 1$, and $\varpi = \varpi_0$. The joint posterior density may then be written as

$$p(\xi_0, \xi_1, \ldots, \xi_K, \theta | \mathbf{d}) \propto \exp\left(-\frac{(\mathbf{d} - \widetilde{\mathbf{G}}(\boldsymbol{\xi}))^T (\mathbf{d} - \widetilde{\mathbf{G}}(\boldsymbol{\xi}))}{2\varsigma^2}\right) \times \left(\frac{\theta}{\varpi_1^2}\right)^{-\frac{K}{2}} \exp\left(-\frac{\sum_{k=1}^{K} \xi_k^2}{2\theta/\varpi_1^2}\right) \times \theta^{-\alpha-1} \exp\left(-\frac{\beta}{\theta}\right)$$
$$\times \exp\left(-\frac{\xi_0^2}{2/\varpi_0^2}\right). \tag{54}$$

MCMC sampling from the posterior proceeds as in Section 3.4, with Gibbs updates for θ and Metropolis updates for ξ. Since the hyperprior on ξ_0 is independent of θ, the full conditional $p(\theta | \xi, \mathbf{d})$ is unchanged.

Results from inversion of the "Lorenz target" profile are shown in Fig. 24. As before, data are obtained from 13 sensors with noise standard deviation $\varsigma = 10^{-1}$, and with the prior correlation length fixed at $L = 0.2$. For simplicity, we put $\varpi_0 = \varpi_1$. The true profile, shown in black, is generally contained within the marginal quantiles. The posterior standard deviation is, as usual, largest near the edges of the domain. These results are obtained with 10 K–L modes (the same number previously used for $L = 0.2$) and a PC basis of order $p = 4$.

To verify the accuracy of this inversion, Fig. 25 directly compares the posterior mean obtained with the combined K–L/PC approach to that obtained with the full-dimensional grid-based approach of Section 3.2. The grid-based approach is applied to identical data from the Lorenz target, retaining the hyperpriors on μ_0 and θ. Mean profiles obtained with the two approaches are nearly indistinguishable. We perform this comparison not only for the 13 sensor, $\varsigma = 10^{-1}$ case of Fig. 24, but for a "finer data" case in which 25 sensors are evenly distributed throughout the domain, each with noise magnitude $\varsigma = 10^{-2}$. Results obtained with finer data preserve the extremely close agreement between the K–L/PC and grid-based inversions. Moreover, finer data yields much closer reconstruction of the true Lorenz target profile.

The impact of the prior-mean hyperparameter can be understood by examining the posterior marginals of μ_0 and the first few K–L mode strengths c_i (or equivalently, ξ_i). These one- and two-parameter marginal densities are shown in Fig. 26, for

---

[5] The Lorenz system is governed by the following ODEs:

$$dx/dt = \bar{\sigma}(y - x),$$
$$dy/dt = x(\bar{\rho} - z) - y,$$
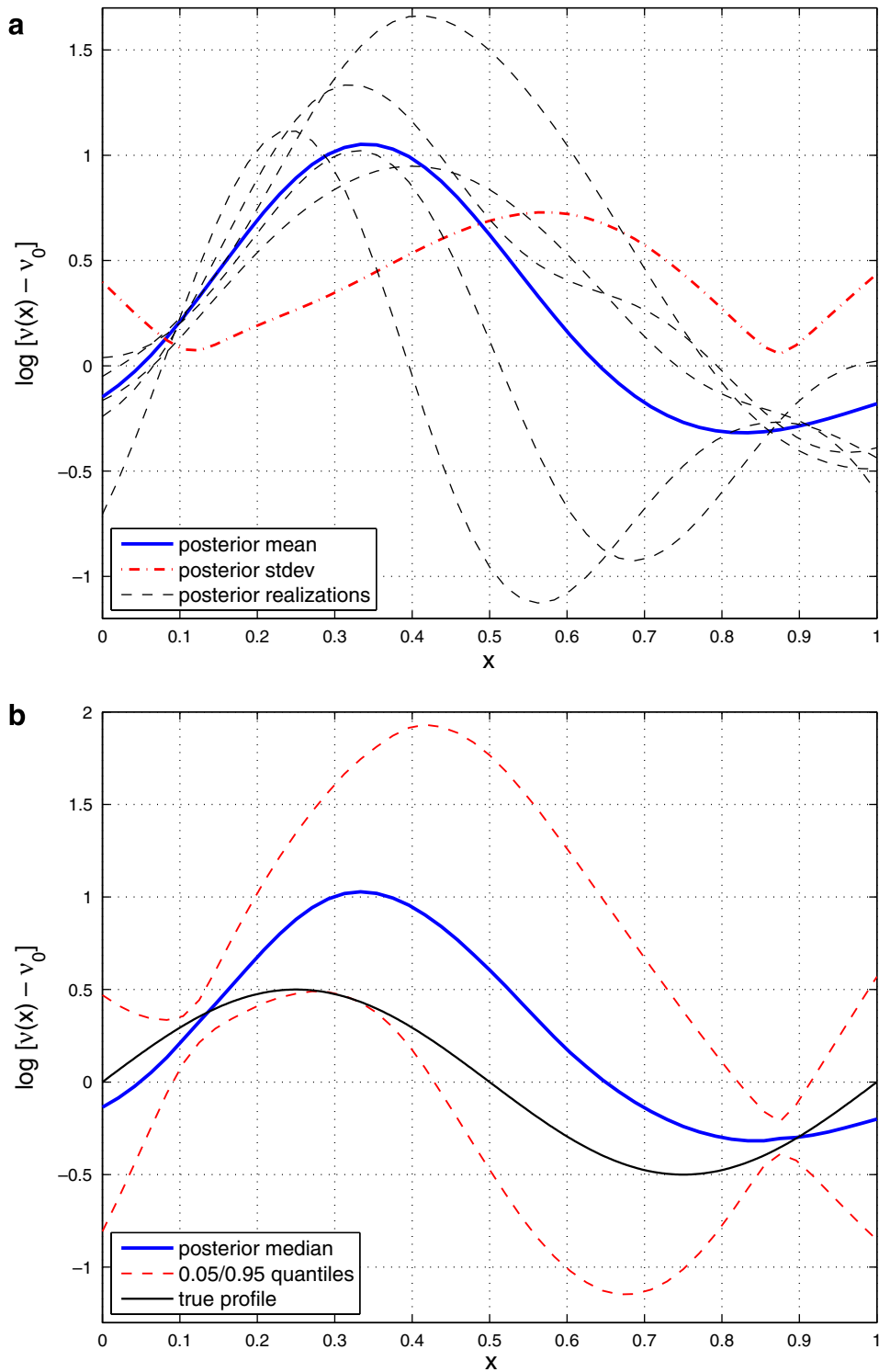$$dz/dt = xy - \bar{\beta}z.$$

Fig. 20. Inversion of the sinusoidal log-diffusivity profile with only two sensors (at $x = 0.0$ and $x = 1.0$): (a) Mean, standard deviation, and five posterior realizations; (b) Median, 1-D credibility intervals, and true profile.

13-sensor inversion of the Lorenz target with the K–L/PC approach. Very tight negative correlation between μ_0 and ξ_1 is apparent. This relationship is not surprising, as the first K–L mode $\phi_1(x)$ is the flattest; it contributes nearly the same shape to the solution profile as does the mean. Changing the weight on this mode shifts the log-diffusivity up and down in much
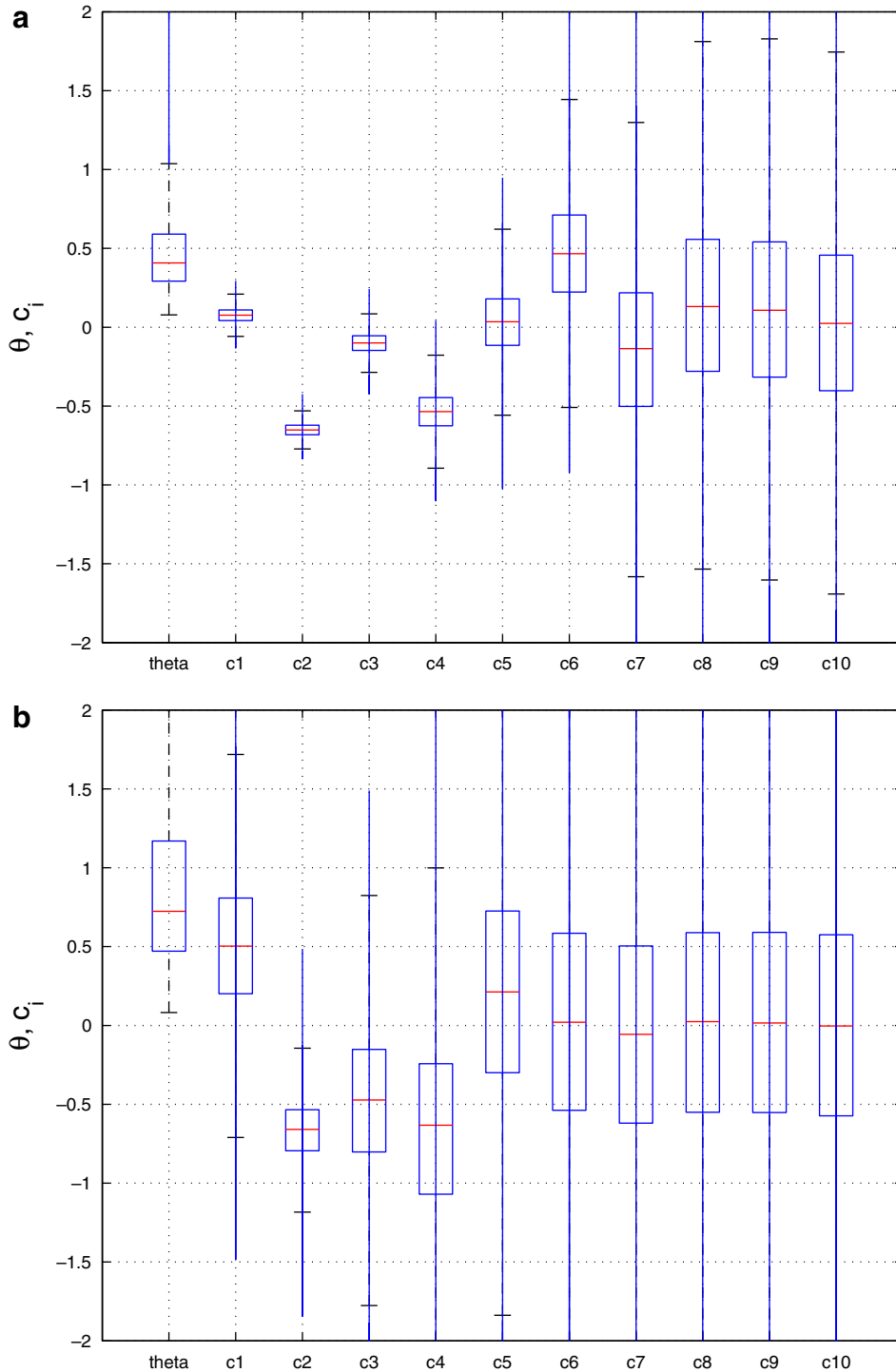
Fig. 21. Boxplot of the posterior marginals of the K–L mode strengths c_i, sinusoidal target: (a) 13 sensors; (b) 2 sensors.

the same way as changing the prior mean. Additional correlation is observed between μ_0 and ξ_3; $\phi_3(x)$ is the next even eigenfunction, and this dependence accounts for deviation of $\phi_1(x)$ from the flat mean near the edges of the domain. Similar posterior correlations are observed in the 25-sensor, small sensor noise case. Because of these strong correlations, inclusion of the prior mean as a hyperparameter does not have a significant impact on posterior profiles of $M(x)$ in these cases.
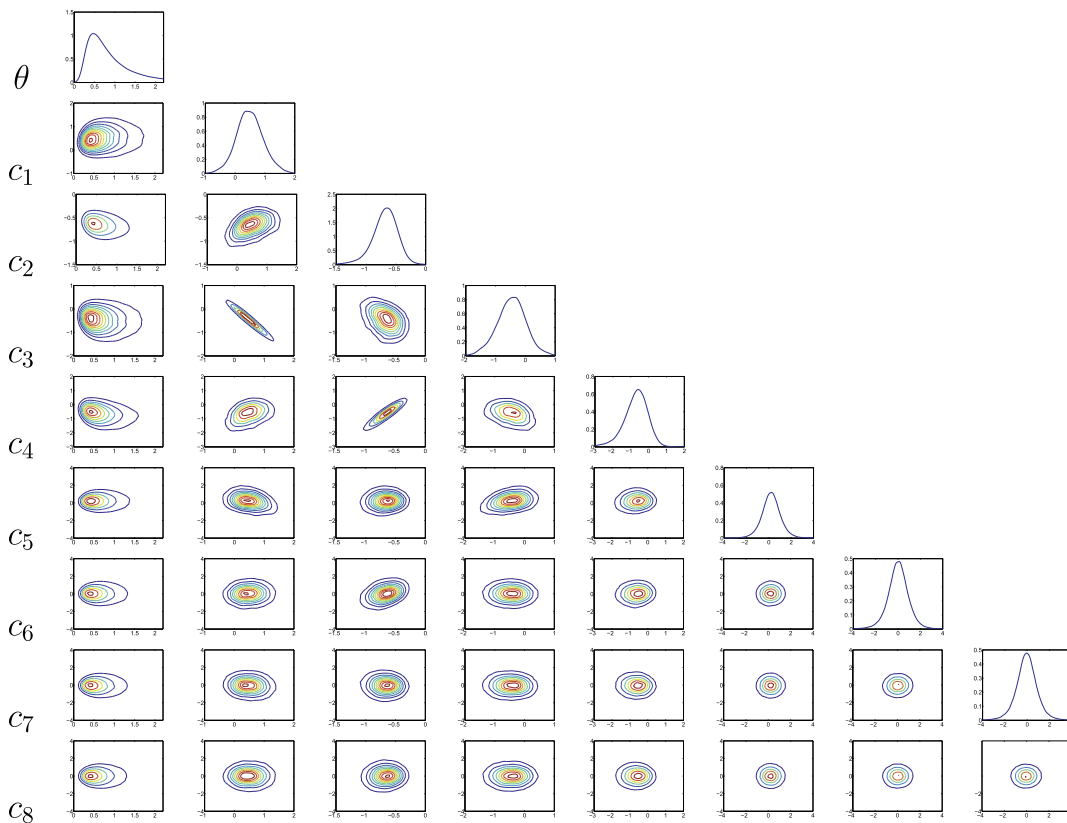
The posterior marginal of the prior-variance hyperparameter θ shows interesting changes with the data. In Fig. 27 we contrast the posterior density $p(\theta|\mathbf{d})$, in both the 13-sensor/larger-noise and 25-sensor/smaller-noise cases, with the inverse gamma hyperprior $p(\theta)$ (39). Both posteriors differ significantly from the hyperprior. The 13-sensor case favors lower values of θ, while the finer-data case favors higher values of θ. This result is borne out in the posterior distribution of the K–L mode strengths ξ_i (equivalently c_i). Finer data, and thus closer reconstruction of the Lorenz target, move the posterior distributions of the higher-index K–L modes further away from zero. Recall that the full conditional $p(\theta|\boldsymbol{\xi}, \mathbf{d})$ is an inverse gamma distribution with scale parameter equal to $\beta + \varpi^2 \left(\sum_{k=1}^{K} \xi_k^2 \right)/2$. As the magnitude of $\sum_k \xi_k^2$ increases, the scale parameter grows, pushing the posterior distribution of θ towards higher values. The strength of the regularization, inversely proportional to θ, therefore decreases as the quality of the data increases. Notably, this result emerges automatically from the hierarchical Bayesian formulation.

4. Conclusions

This paper pursues a Bayesian approach to inverse problems in which the unknown quantity is a spatiotemporal field. Regularization enters the formulation through a hierarchical Gaussian process prior, and the posterior distribution provides a quantitative assessment of uncertainty in the inverse solution. We address computational challenges associated with repeated evaluations of a complex forward model, and with high dimensionality of the posterior.

We introduce a Karhunen–Loève expansion of the prior Gaussian process in order to parameterize the unknown field with fewer degrees of freedom. This representation also defines a stochastic forward problem whose solution approximates the deterministic forward model over the support of the prior. We use Galerkin projection to solve the stochastic forward problem on a polynomial chaos basis. The polynomial representation of the forward model then replaces the full forward model in the likelihood function, yielding a reduced-dimensionality surrogate posterior density that may be evaluated orders of magnitude more quickly than the original posterior.

This approach is demonstrated on a transient diffusion equation with prescribed source terms, where our goal is to recover an inhomogeneous diffusivity field from sparse and noisy observations of the solution. We compare three Bayesian solution methods corresponding to three different posterior distributions: a full-dimensional, grid-based posterior with
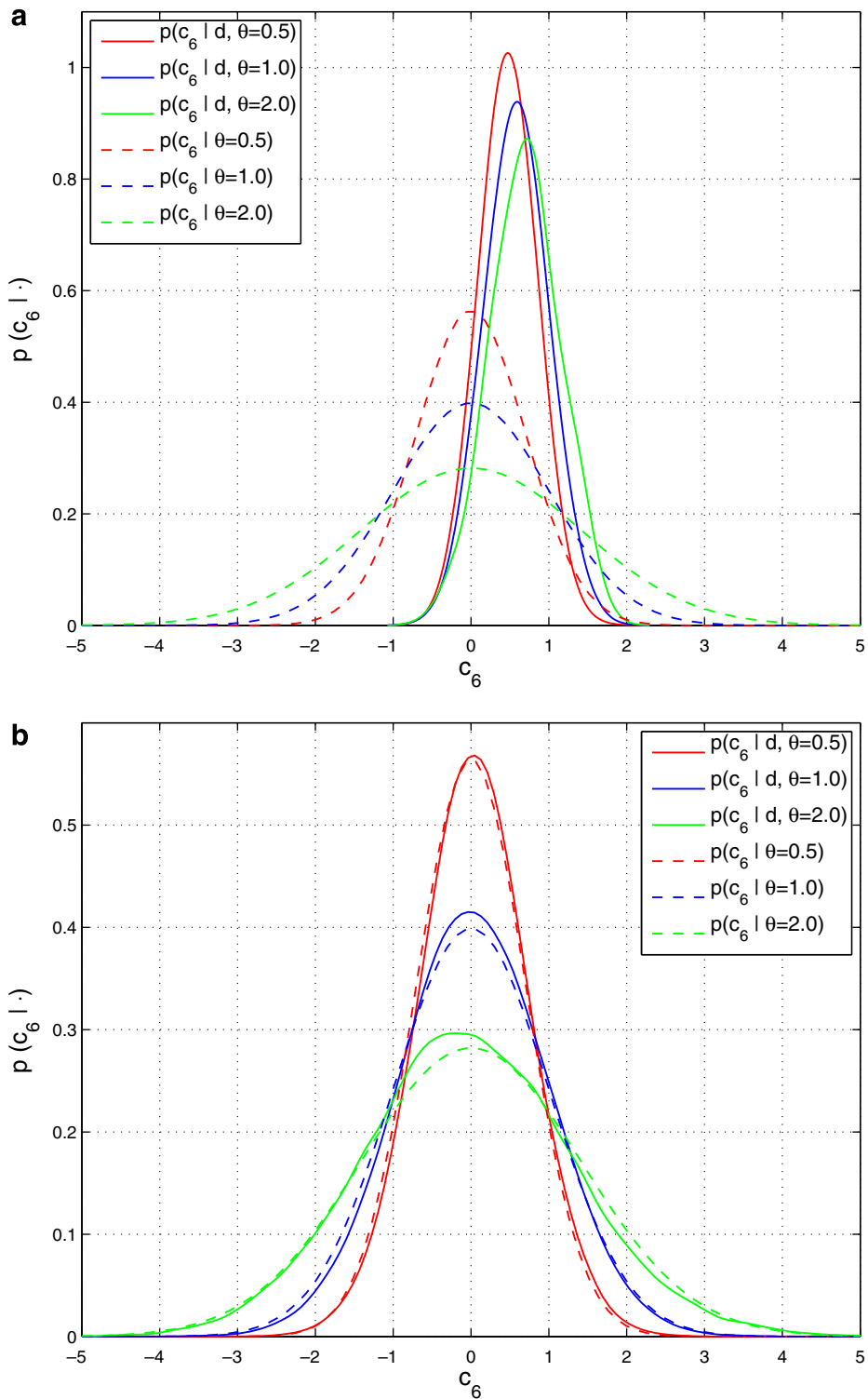
Fig. 23. Prior and posterior conditional distributions of K–L mode c_6; inference of the sinusoidal target: (a) 13 sensors; (b) 2 sensors.

direct forward model evaluations (42); a posterior based on the K–L mode strengths, with direct forward model evaluations (43); and a posterior based on transformation of the K–L mode strengths to the variables underlying a PC representation of the forward model (51). In each case, we apply a single-component MCMC sampler with Metropolis updates for the field
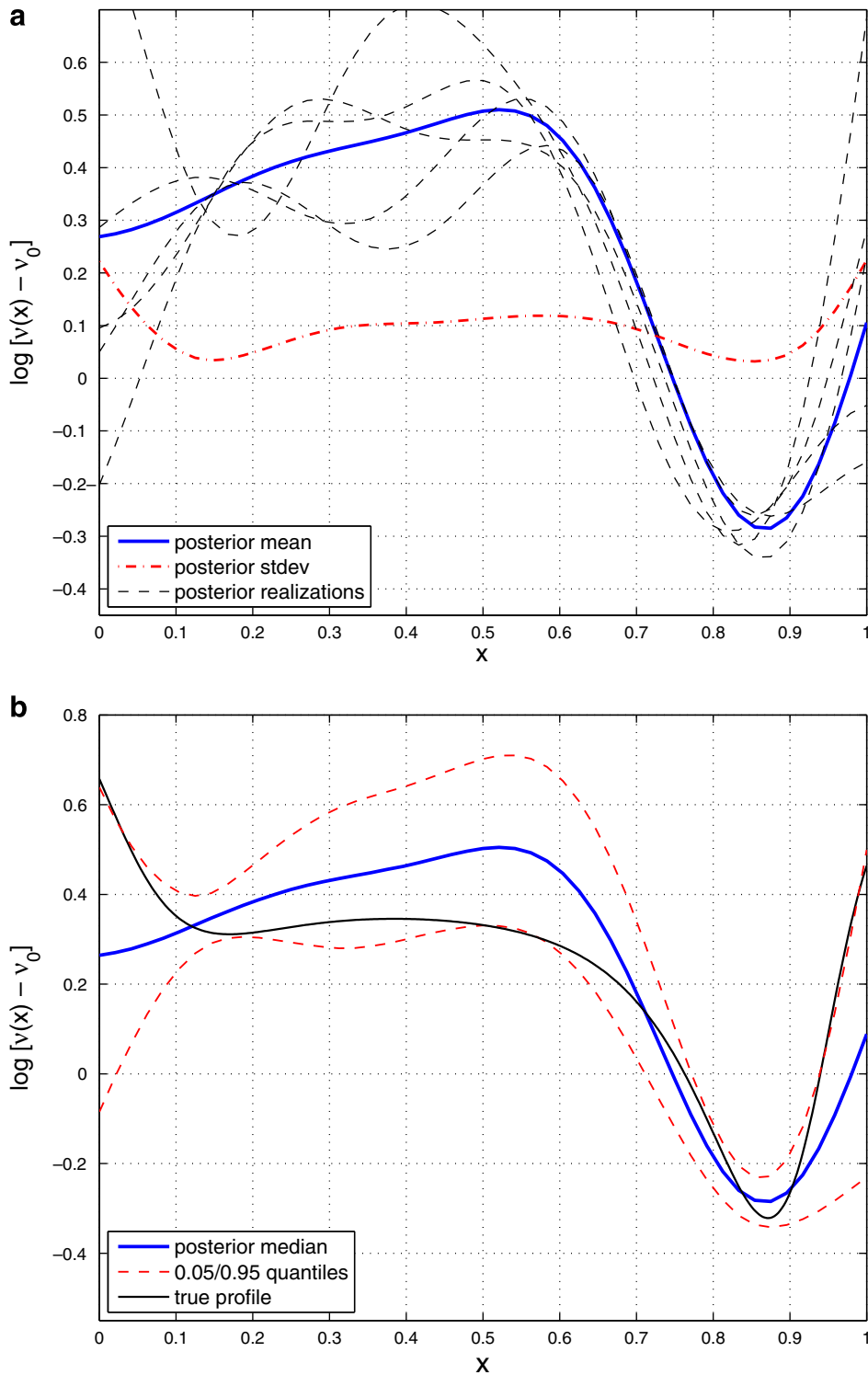
Fig. 24. Inversion of the "Lorenz target" log-diffusivity profile using dimensionality reduction and polynomial chaos, with both the prior mean and prior variance treated as hyperparameters; $K = 10$, $p = 4$: (a) Mean, standard deviation, and five posterior realizations; (b) Median, 1-D credibility intervals, and true profile.

parameters and Gibbs updates for the scale of the prior covariance, treated as a hyperparameter. The prior mean may also be treated as a hyperparameter. We invert for several "target" profiles of the log-diffusivity, representing different scales of spatial variation and different degrees of smoothness.
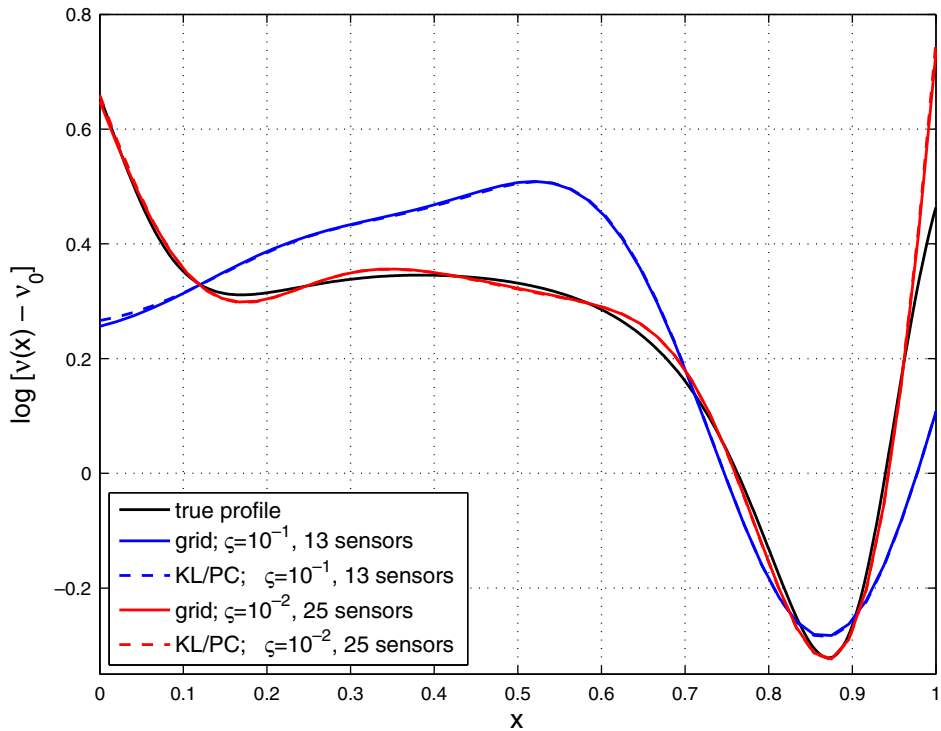
Fig. 25. Effect of measurement noise ς and more finely-spaced sensors on inversion of the Lorenz target with the K–L/PC formulation, compared to the unaccelerated grid-based formulation. Solid or dashed colored lines are posterior means, while the true profile is shown in solid black.
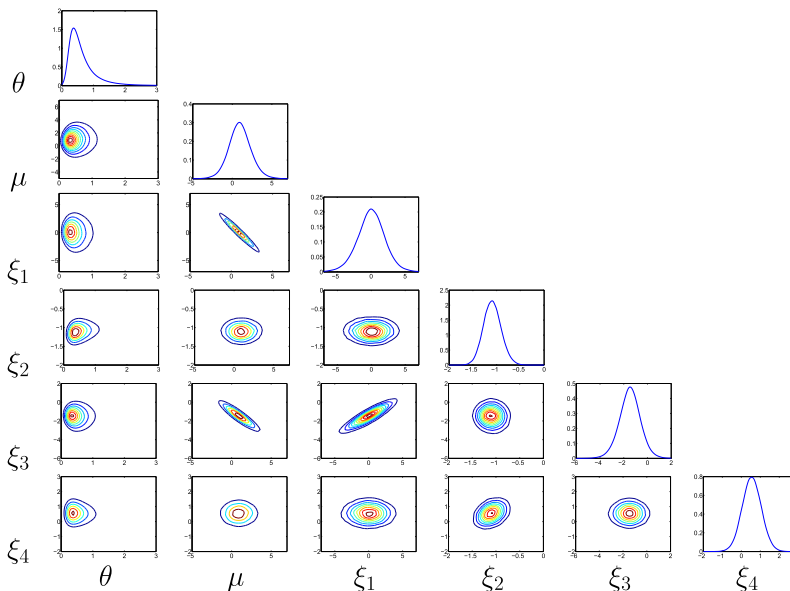

Fig. 26. 1-D and 2-D posterior marginals of the hyperparameters θ and μ_0 and the first few K–L mode strengths ξ_i; inference of the Lorenz target with the K–L/PC formulation.

Our findings are summarized as follows:

- Moments and quantiles of the inverse solution calculated from the latter two posterior distributions converge rapidly to those of the baseline case (grid-based, direct forward problem evaluations) as we increase the number of K–L modes K and the PC basis order p. Shorter prior correlation lengths require a larger number of K–L modes to be retained. Convergence of the PC-based surrogate posterior to the baseline case may depend on the nonlinearity of the forward operator, but very close agreement is obtained here at order $p = 4$.
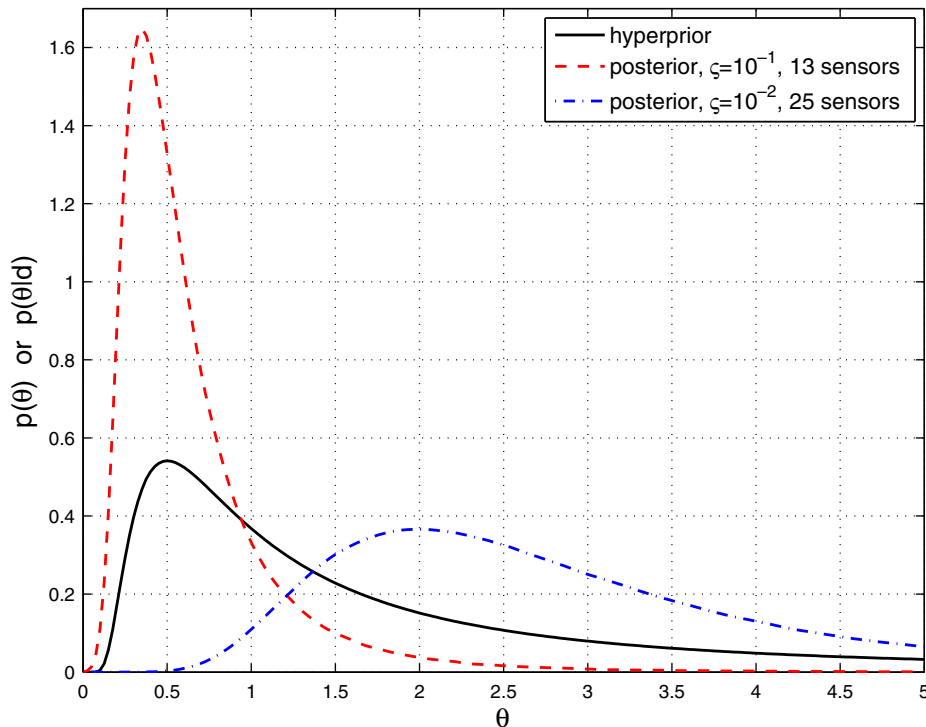
Fig. 27. Posterior marginal density of the variance hyperparameter θ, in both the "regular data" (13 sensors, $\varsigma = 10^{-1}$) and "finer data" (25 sensors, $\varsigma = 10^{-2}$) cases, contrasted with the hyperprior density on θ; inference of the Lorenz target with K–L/PC formulation.

- The relative importance of the K–L modes changes from the prior to the posterior, but still decays at large index, consistent with the smoothing effect of the prior.
- Both the lower-dimensional posterior and the PC-based surrogate posterior accurately capture spatial correlations among possible values of the inverse solution.
- The K–L parameterization allows for more efficient MCMC sampling than the full-dimensional parameterization.
- At sufficiently high index, we find that the posterior distribution of each K–L mode reverts to the prior. With less data—in particular, with sensors placed only at the boundaries—we observe greater posterior variability. The weights of the K–L modes then revert to the prior distribution at lower index. In these cases, certain modes may be relatively unaffected by the data, yet make significant contributions to the posterior variance.
- Measures of the speedup associated with the new scheme are quite promising. We observe nearly one order of magnitude in speedup, along with more efficient MCMC sampling, in moving from a grid-based parameterization of the unknown field to a truncated K–L parameterization, while retaining direct forward solutions. An additional 1–1.5 orders of magnitude of speedup are then obtained by using the reduced-dimensionality surrogate posterior. These figures *include* the time needed to solve the stochastic forward problem. In practical settings, this step may be performed offline before observing any data.

Extensions to the basic methodology proposed here are numerous. For instance, more advanced Bayesian modeling may take the form of estimating the prior correlation length L, here presumed known. If sufficient data are available, one could infer L by treating it as an additional hyperparameter; unlike the prior variance or prior mean, this hyperparameter would affect the K–L eigenfunctions. Also, the K–L parameterization used here will apply equally well to non-Gaussian priors, though in such a case, additional effort may be required to evaluate or sample from the non-Gaussian joint prior density on the weights of the K–L modes.

Extension to multi-dimensional spatial domains or prior covariance kernels with slower eigenvalue decay is straightforward in principle, but the resulting stochastic forward problems may be larger and thus more computationally challenging. Recent algorithmic developments enabling more efficient solution of stochastic PDEs with random-field coefficients would be useful to apply within the present Bayesian inverse framework. These developments include adaptive polynomial degree and sparse truncations of the PC basis [34,75], and other methods of stochastic model reduction [76]. It is also crucial to note that solution of the stochastic forward problem need not be restricted to Galerkin methods: stochastic collocation [77], particularly on sparse grids [78], could be used as well. One may also explore partitioning the prior support, as described in [6], and solving several smaller stochastic forward problems in parallel.

Finally, the basis on which we represent the unknown field in this work has been determined from the prior. An important implication is that the present dimensionality reduction scheme requires some degree of correlation or structure in the prior; without the corresponding decay of eigenvalues in the K–L expansion, the scheme would have no opportunity for truncation. On the other hand, results in Section 3.5 suggest that features prominent in the prior may not always have a significant effect on the measurements, and that the forward model and the data could therefore also be taken into consideration when pursuing dimensionality reduction. In this sense, one would ultimately like to find a basis emphasizing features of the unknown field that are most affected by the data [79]; such a basis may be yet more efficient than the bases considered here. Effective means of computing such a basis, for nonlinear forward models in a Bayesian framework, are a promising direction for further work.

Acknowledgments

Appendix. Reproducing kernel Hilbert space

Here we briefly review properties of reproducing kernel Hilbert spaces (RKHS). Useful connections among RKHS, Gaussian process priors, regularization, and the Karhunen–Loève expansion are discussed in Section 2.6. Much more detailed expositions of RKHS may be found in [80,59,38,44].

A RKHS is a Hilbert space H of functions for which pointwise evaluation is a bounded linear functional. Let $L_t : H \to \mathbb{R}$ denote such a functional, where elements of H are real-valued functions defined on some domain D; for $f \in H$, $L_t(f) = f(t)$. By the Riesz representation theorem, there exists a unique $K_t \in H$ called the representer of evaluation at t, such that

$$L_t f = \langle f, K_t \rangle_H = f(t) \tag{55}$$

for all $f \in H$, where $\langle \cdot, \cdot \rangle_H$ is the inner product in H. The function $K(t,s) \equiv K_t(s)$ is called the *reproducing kernel*, since the reproducing property

$$\langle K(t, \cdot), K(t', \cdot) \rangle_H = K(t, t') \tag{56}$$

follows directly from (55). Now consider two elements f and g of H, with $f \equiv \sum_i \alpha_i K(t_i, \cdot)$ and $g \equiv \sum_j \beta_j K(t_j, \cdot)$. Their inner product $\langle f, g \rangle$ is

$$\left\langle \sum_i \alpha_i K_{t_i}, \sum_j \beta_j K_{t_j} \right\rangle = \sum_{ij} \alpha_i \beta_j \left\langle K_{t_i}, K_{t_j} \right\rangle = \sum_{ij} \alpha_i \beta_j K(t_i, t_j). \tag{57}$$

From the properties of the inner product, it is straightforward to show that $K(t,s) : D \times D \to \mathbb{R}$ is symmetric and positive definite.

Conversely, it can be shown that for every positive definite function K on $D \times D$ there exists a unique RKHS $H(K)$ with K as its reproducing kernel [80]. In fact, $H(K)$ is the completion of the space spanned by $\{f(\cdot) = K(x, \cdot)\}_{x \in D}$ with respect to the inner product (57).

An intuitive appreciation of RKHS may come by contrasting (55) with the situation in $L^2(D)$. In $L^2(D)$, the representer of evaluation is a delta function:

$$\int_D f(s)\delta(s-t)ds = f(t) \tag{58}$$

but the delta function is not in $L^2(D)$, and thus L^2 is not a RKHS. (Indeed, $L^2(D)$ has no bounded linear evaluation functional.) In a RKHS, the existence of the representer $K_t \in H(K)$ implies a relationship between a function value at t and its value at *other* points $t' \in D$ [38]. The Gaussian kernel, for instance, captures the notion that nearby values are relevant. Other kernels may capture periodicity in f, polynomial trends, and other features.

If the reproducing kernel K is also a Mercer kernel, we can alternatively construct the RKHS $H(K) \subset L^2(D)$ from eigenfunctions of the integral operator defined by the kernel. The eigenfunctions ψ_i and eigenvalues λ_i are given by

$$\int_D K(s,t)\psi_i(t)dt = \lambda_i \psi_i(s). \tag{59}$$

The kernel K can then be expanded in terms of the orthonormal sequence of eigenfunctions $\{\psi_i\}$:

$$K(s,t) = \sum_{i=1}^{\infty} \lambda_i \psi_i(s)\psi_i(t). \tag{60}$$

Now consider two functions $f, g : D \to \mathbb{R}$, with projections

$$f_i = \int_D f(s)\psi_i(s)ds \tag{61}$$

and similarly for g. We define their inner product as

$$\langle f, g \rangle_K = \left\langle \sum_i f_i \psi_i, \sum_j g_j \psi_j \right\rangle \equiv \sum_i \frac{f_i g_i}{\lambda_i}, \tag{62}$$

which induces the norm

$$\|f\|_K^2 = \left\langle \sum_i f_i \psi_i, \sum_j f_i \psi_i \right\rangle \equiv \sum_i \frac{f_i^2}{\lambda_i}. \tag{63}$$

It can be shown [59] that $f \in H(K)$ if and only if $\|f\|_K^2 < \infty$, that is, iff:

$$\sum_i \frac{f_i^2}{\lambda} < \infty. \tag{64}$$

Requiring the RKHS norm to be finite in effect enforces some "smoothness" on members of $H(K)$, in contrast to $L^2(D)$; since we divide by λ_i in (63), the Fourier coefficients f_i must decay quickly enough for the sum to be finite. The representer property (55) is, of course, preserved under the definition of inner product in (62). In fact the inner products (57) and (62) are equivalent [59].

References

[1] C.R. Vogel, Computational Methods for Inverse Problems, SIAM, 2002.
[2] S.N. Evans, P.B. Stark, Inverse problems as statistics, Inverse Problems 18 (2002) R55–R97.
[3] L. Tenorio, Statistical regularization of inverse problems, SIAM Rev. 43 (2) (2001) 347–366.
[4] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer, 2005.
[5] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, 2005.
[6] Y.M. Marzouk, H.N. Najm, L.A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, J. Comput. Phys. 224 (2) (2007) 560–586.
[7] N. Wiener, The homogeneous chaos, Am. J. Math. 60 (1938) 897–936.
[8] R. Cameron, W. Martin, The orthogonal development of nonlinear functionals in series of Fourier–Hermite functionals, Ann. Math. 48 (1947) 385–392.
[9] R. Ghanem, P. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer-Verlag, New York, 1991.
[10] D. Xiu, G. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.
[11] 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 J. Sci. Comput. 26 (2) (2004) 698–719.
[12] X. Wan, G.E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, J. Comput. Phys. 209 (2005) 617–642.
[13] R. Ghanem, Probabilistic characterization of transport in heterogeneous media, Comput. Methods Appl. Mech. Eng. 158 (1998) 199–220.
[14] R. Ghanem, J. Red-Horse, A. Sarkar, Modal properties of a space-frame with localized system uncertainties, in: A. Kareem, A. Haldar, B.S., Jr., E. Johnson (Eds.), 8th ASCE Specialty Conference of Probabilistic Mechanics and Structural Reliability, PMC200-269, ASCE, 2000.
[15] O. Le Maître, O. Knio, H. Najm, R. Ghanem, A stochastic projection method for fluid flow I. Basic formulation, J. Comput. Phys. 173 (2001) 481–511.
[16] O. Le Maître, M. Reagan, H. Najm, R. Ghanem, O. Knio, A stochastic projection method for fluid flow II. Random process, J. Comput. Phys. 181 (2002) 9–44.
[17] D. Xiu, D. Lucor, C.-H. Su, G. Karniadakis, Stochastic modeling of flow-structure interactions using generalized polynomial chaos, ASME J. Fluids Eng. 124 (2002) 51–59.
[18] 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, Phys. Fluids 15 (8) (2003) 2238–2250.
[19] M. Reagan, H. Najm, B. Debusschere, O. Le Maître, O. Knio, R. Ghanem, Spectral stochastic uncertainty quantification in chemical systems, Comb. Theory Model. 8 (2004) 607–632.
[20] J. Wang, N. Zabaras, Using Bayesian statistics in the estimation of heat source in radiation, Int. J. Heat Mass Transfer 48 (2005) 15–29.
[21] G. Berkooz, P. Holmes, J.L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, Annu. Rev. Fluid Mech. 25 (1993) 539–575.
[22] 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 Resour. Res. 39 (12) (2003) 1350.
[23] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. Royal Stat. Soc.: Series B 63 (3) (2001) 425–464.
[24] J.A. Christen, C. Fox, MCMC using an approximation, J. Comput. Graph. Stat. 14 (4) (2005) 795–810.
[25] D. Higdon, H. Lee, C. Holloman, Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems, Bayesian Stat. 7 (2003) 181–197.
[26] 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.
[27] Y. Efendiev, T.Y. Hou, W. Luo, Preconditioning Markov chain Monte Carlo simulations using coarse-scale models, SIAM J. Sci. Comput. 28 (2006) 776–803.
[28] H. Lee, D. Higdon, Z. Bi, M. Ferreira, M. West, Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media, Technometrics 44 (3) (2002) 230–241.
[29] J. Wang, N. Zabaras, A Markov random field model of contamination source identification in porous media flow, Int. J. Heat Mass Transfer 49 (2006) 939–950.
[30] J. Kaipio, E. Somersalo, Statistical inverse problems: discretization, model reduction, and inverse crimes, J. Comput. Appl. Math. 198 (2) (2007) 493–504.
[31] W. Li, O.A. Cirpka, Efficient geostatistical inverse methods for structured and unstructured grids, Water Resour. Res. 42 (2006) W06402.

[32] D. Xiu, G. Karniadakis, Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos, Comput. Methods Appl. Mech. Eng. 191 (2002) 4927–4948.
[33] O. Le Maître, M. Reagan, B. Debusschere, H. Najm, R. Ghanem, O. Knio, Natural convection in a closed cavity under stochastic, non-Boussinesq conditions, SIAM J. Sci. Comput. 26 (2) (2004) 375–394.
[34] P. Frauenfelder, C. Schwab, R. Todor, Finite elements for elliptic problems with stochastic coefficients, Comput. Methods Appl. Mech. Eng. 194 (2005) 205–228.
[35] H.G. Matthies, A. Keese, Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations, Comput. Methods Appl. Mech. Eng. 194 (2005) 1295–1331.
[36] M. Grigoriu, Stochastic Calculus, Birkhäuser, 2002.
[37] P.G. Hoel, S.C. Port, C.J. Stone, Introduction to Stochastic Processes, Waveland Press, 1987.
[38] M. Seeger, Gaussian processes for machine learning, Int. J. Neural Syst. 14 (2) (2004) 1–38.
[39] C.K.I. Williams, C.E. Rasmussen, Gaussian processes for regression, in: D.S. Teuretzky, M.C. Mozer, M.E. Hasselmo (Eds.), Advances in Neural Information Processing Systems, vol. 8, MIT Press, 1996.
[40] N.A.C. Cressie, Statistics for Spatial Data, Wiley, 1993.
[41] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.
[42] M.G. Genton, Classes of kernels for machine learning: a statistics perspective, J. Mach. Learning Res. 2 (2001) 299–312.
[43] M. Loève, Probability Theory II, fourth ed., Springer, 1978.
[44] R.J. Adler, J.E. Taylor, Random Fields and Geometry, Springer, 2007.
[45] P. Abrahamsen, A review of Gaussian random fields and correlation functions, Technical Report 917, Norwegian Computing Center, Oslo, Norway, 1997.
[46] R. Courant, D. Hilbert, Methods of Mathematical Physics, vol. 1, Wiley-Interscience, 1953.
[47] L. Sirovich, Turbulence and the dynamics of coherent structures. Part 1: coherent structures, Quart. Appl. Math. 45 (3) (1987) 561–571.
[48] 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.
[49] R. Aster, B. Borchers, C. Thurber, Parameter Estimation and Inverse Problems, Academic Press, 2004.
[50] W.P. Gouveia, J.A. Scales, Resolution of seismic waveform inversion: Bayes versus Occam, Inverse Problems 13 (1997) 323–349.
[51] A. Malinverno, Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, Geophys. J. Int. 151 (2002) 675–688.
[52] C. Jackson, M.K. Sen, P.L. Stoffa, An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions, J. Clim. 17 (2004) 2828–2841.
[53] J. Wang, N. Zabaras, Hierarchical Bayesian models for inverse problems in heat conduction, Inverse Prob. 21 (2005) 183–206.
[54] D.J.C. MacKay, Comparison of approximate methods for handling hyperparameters, Neural Comput. 11 (5) (1999) 1035–1068.
[55] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 1996.
[56] 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.
[57] W. Schoutens, Stochastic Processes and Orthogonal Polynomials, Springer, 2000.
[58] 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.
[59] G. Wahba, Spline Models for Observational Data, SIAM, 1990.
[60] G.S. Kimeldorf, G. Wahba, A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, Ann. Math. Stat. 41 (2) (1970) 495–502.
[61] D.S. Oliver, L. Cunha, A.C. Reynolds, Markov chain Monte Carlo methods for conditioning a permeability field to pressure data, Math. Geol. 29 (1) (1997) 61–91.
[62] B. Sommeijer, L. Shampine, J. Verwer, RKC: an explicit solver for parabolic PDEs, J. Comput. Appl. Math. 88 (1997) 315–326.
[63] 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.
[64] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, second ed., Chapman & Hall, CRC, 2003.
[65] A. Gelman, Prior distributions for variance parameters in hierarchical models, Bayesian Anal. 1 (3) (2006) 515–534.
[66] H.K.H. Lee, C.H. Holloman, C.A. Calder, D.M. Higdon, Flexible Gaussian processes via convolution, Discussion paper 2002-09, Duke University, Department of Statistical Science, 2002.
[67] J.P. Kaipio, V. Kolehmainen, M. Vauhkonen, E. Somersalo, Inverse problems with structural prior information, Inverse Prob. 15 (1999) 713–729.
[68] D. Calvetti, E. Somersalo, A Gaussian hypermodel to recover blocky objects, Inverse Prob. 23 (2007) 733–754.
[69] W. Press, S.A. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes in C: The Art of Scientific Computing, second ed., Cambridge University Press, Cambridge, UK, 1992.
[70] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, second ed., Springer, 2004.
[71] R.G. Ghanem, The nonlinear Gaussian spectrum of log-normal stochastic processes and variables, ASME J. Appl. Mech. 66 (4) (1999) 964–973.
[72] W. Luo, Wiener chaos expansion and numerical solutions of stochastic partial differential equations, Ph.D. thesis, California Institute of Technology, May 2006.
[73] J. Dunkley, M. Bucher, P.G. Ferreira, K. Moodley, Fast and reliable Markov chain Monte Carlo technique for cosmological parameter estimation, Mon. Notices Roy. Astronom. Soc. 356 (3) (2005) 925–936.
[74] E.N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. 20 (2) (1963) 130–141.
[75] T.Y. Hou, W. Luo, B. Rozovskii, H.-M. Zhou, Wiener Chaos expansions and numerical solutions of randomly-forced equations of fluid mechanics, J. Comput. Phys. 216 (2006) 687–706.
[76] A. Doostan, R.G. Ghanem, J. Red-Horse, Stochastic model reduction for chaos representations, Comput. Methods Appl. Mech. Eng. 196 (2007) 3951–3966.
[77] I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM J. Numer. Anal. 45 (3) (2007) 1005–1034.
[78] D. Xiu, J.S. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM J. Sci. Comput. 27 (3) (2005) 1118–1139.
[79] S. Kraut, R.H. Anderson, J.L. Krolik, A generalized Karhunen–Loève basis for efficient estimation of tropospheric refractivity using radar clutter, IEEE Trans. Signal Process. 52 (1) (2004) 48–60.
[80] N. Aronszajn, Theory of reproducing kernels, Trans. Am. Math. Soc. 68 (3) (1950) 337–404.