Surrogate and reduced-order modeling: a comparison of approaches for large-scale statistical inverse problems

M. Frangos, Y. Marzouk, K. Willcox, Massachusetts Institute of Technology; B. van Bloemen Waanders, Sandia National Laboratories

0.1 Introduction

Solution of statistical inverse problems via the frequentist or Bayesian approaches described in earlier chapters can be a computationally intensive endeavor, particularly when faced with large-scale forward models characteristic of many engineering and science applications. High computational cost arises in several ways. First, thousands or millions of forward simulations may be required to evaluate estimators of interest or to characterize a posterior distribution. In the large-scale setting, performing so many forward simulations is often computationally intractable. Second, sampling may be complicated by the large dimensionality of the input space—as when the inputs are fields represented with spatial discretizations of high dimension—and

Computational Methods for Large-Scale Inverse Problems and Quantification of Uncertainty. Edited by People on Earth

by nonlinear forward dynamics that lead to multimodal, skewed, and/or strongly correlated posteriors. In this chapter, we present an overview of surrogate and reduced-order modeling methods that address these computational challenges. For illustration, we consider a Bayesian formulation of the inverse problem. Though some of the methods we review exploit prior information, they largely focus on simplifying or accelerating evaluations of a stochastic model for the data, and thus are also applicable in a frequentist context.

Methods to reduce computational cost of solving of a statistical inverse problem can be classed broadly in three groups. First, there are many ways to reduce the cost of a posterior evaluation (or more specifically, the cost of a forward simulation) through surrogate models, reduced-order models, multigrid and multiscale approaches, and stochastic spectral approaches. Second, the dimension of the input space can be reduced, through truncated Karhunen-Loève expansions, coarse grids, and parameter-space reductions. The third set of approaches targets a reduction in the number of forward simulations required to compute estimators of interest, i.e., more efficient sampling. In the Bayesian setting, these methods include a wide range of adaptive and multi-stage Markov chain Monte Carlo (MCMC) schemes.

This chapter summarizes the state of the art in methods to reduce the computational complexity of statistical inverse problems, provides some introductory material to support the technical chapters that follow, and offers insight into the relative advantages of different methods through a simple illustrative example. Section 0.2 presents an overview of state-of-the-art approaches. Section 0.3 then describes our specific problem setup, with a focus on the structure of the large-scale forward model that underlies the inverse problem. We then present a detailed comparison of reduced-order modeling and stochastic spectral approximations in Sections 0.4 and 0.5, respectively. Section 0.6 presents an illustrative example and Section 0.7 provides some concluding remarks.

0.2 Reducing the computational cost of solving statistical inverse problems

As discussed in Section 0.1, we consider three general classes of approaches to reducing the computational cost of solving statistical inverse problems: reducing the cost of a forward simulation, reducing the dimension of the input parameter space, and reducing the number of samples (forward simulations) required.

0.2.1 Reducing the cost of forward simulations

Many attempts at accelerating inference in computationally intensive inverse problems have relied on surrogates for the forward model, typically constructed through repeated forward simulations that are performed in an offline phase. Eldred et al.

(2004) categorize surrogates into three different classes: data-fit models, reduced-order models, and hierarchical models, all of which have been employed in the statistical inverse problem setting.

Data-fit models are generated using interpolation or regression of simulation data from the input/output relationships in the high-fidelity model. In the statistical literature, Gaussian processes have been used extensively as surrogates for complex computational models (Kennedy and O'Hagan 2001). 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. Bliznyuk et al. (2008) use local experimental design combined with radial basis function approximations to approximate the posterior and demonstrate the effectiveness of their approach for a pollutant diffusion problem.

Reduced-order models are commonly derived using a projection framework; that is, the governing equations of the forward model are projected onto a subspace of reduced dimension. This reduced subspace is defined via a set of basis vectors, which, for general nonlinear problems, can be calculated via the proper orthogonal decomposition (POD) (Holmes et al. 1996; Sirovich 1987) or with reduced basis methods (Noor and Peters 1980). For both approaches, the empirical basis is pre-constructed using full forward problem simulations—or "snapshots." Wang and Zabaras (2005) use the POD to accelerate forward model calculations in a radiative source inversion problem. Galbally et al. (2010) combine POD with an empirical interpolation method (Barrault et al. 2004; Grepl et al. 2007) for Bayesian inference in a highly nonlinear combustion problem governed by an advection-diffusion-reaction partial differential equation. In both of these applications, the choice of inputs to the simulations—in particular, how closely the inputs must resemble the inverse solution—can be important (Wang and Zabaras 2005).

Hierarchical surrogate models span a range of physics-based models of lower accuracy and reduced computational cost. Hierarchical surrogates are derived from higher-fidelity models using approaches such as simplifying physics assumptions, coarser grids, alternative basis expansions, and looser residual tolerances. Arridge et al. (2006) use mesh coarsening for solving a linear inverse problem, employing spatial discretizations that are coarser than those necessary for accurate solution of the forward problem. Their Bayesian formulation also includes the statistics of the error associated with the discretization, but unfortunately the level of mesh coarsening required for decreasing the computational cost of large-scale problems to acceptable levels may result in large errors that are difficult to quantify or that even yield unstable numerical schemes. Balakrishnan et al. (2003) introduce a polynomial chaos (PC) representation of the forward model in a groundwater transport parameter identification problem, and obtain the PC coefficients by linear regression; again, this process depends on a series of representative snapshots obtained from repeated forward simulations.

0.2.2 Reducing the dimension of the input space

When the object of inference is a spatially distributed parameter, dimensionality of the input space may naively be tied to dimensionality of the spatial discretization. Often, however, there is knowledge of smoothness or structure in the parameter field that can lead to a more efficient basis. In particular, one can introduce a Karhunen-Loève (K-L) expansion based on the prior covariance, transforming the inverse problem to inference on a truncated sequence of weights of the K-L modes (Marzouk and Najm 2009). Li and Cirpka (2006) emphasize the role of K-L expansions in enabling geostatistical inversion on unstructured grids. Efendiev et al. (2006) use K-L expansions to parameterize a log-permeability field, introducing constraints among the weights in order to match known values of the permeability at selected spatial locations.

Recent work has also proposed a reduced-basis approach to reducing the dimension of the input space. In the same way that a reduced-order model is formed by projecting the state onto a reduced subspace, Lieberman (2009) considers projection of a high-dimensional parameter field onto an empirical parameter basis. The effectiveness of this approach is demonstrated for a groundwater inverse problem where MCMC sampling is carried out in the reduced parameter space.

0.2.3 Reducing the number of samples

A different set of approaches retain the full forward model but use simplified or coarsened models to guide and improve the efficiency of MCMC sampling. Christen and Fox (2005) 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. (2003) 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 (Geyer 1991) in which "swap proposals" allow information from the coarse-scale formulation to influence the fine-scale chain. Efendiev et al. (2006) 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.

0.3 General formulation

Here we describe a general inverse problem formulation and then focus on the structure of the forward problem. We consider the task of inferring inputs $\boldsymbol{\theta}$ from limited and imperfect observations $\mathbf{d}$. For simplicity, we let both $\mathbf{d}$ and $\boldsymbol{\theta}$ be finite-dimensional. The relationship between observables $\mathbf{d}$ and inputs $\boldsymbol{\theta}$ is *indirect*, and is denoted by $\mathbf{d} = \mathcal{G}(\boldsymbol{\theta}, \boldsymbol{\eta})$; the additional input $\boldsymbol{\eta}$ is a random variable encompassing measurement noise and/or modeling errors. The likelihood function $\pi(\mathbf{d}|\boldsymbol{\theta})$

describes the probability of measurements $\mathbf{d}$ given a particular value of $\boldsymbol{\theta}$. The likelihood function thus incorporates the forward model and the measurement noise, and results from the particular form of $\mathcal{G}$. In a Bayesian formulation of the inverse problem, $\boldsymbol{\theta}$ is treated as a random variable, endowed with a prior probability density $\pi_{\text{pr}}(\boldsymbol{\theta})$ that encodes any available prior knowledge about the inputs. Bayes' rule then yields the posterior probability density of the inputs, π_{post}:

$$\pi_{\text{post}}(\boldsymbol{\theta}) \equiv \pi(\boldsymbol{\theta}|\mathbf{d}) \propto \pi(\mathbf{d}|\boldsymbol{\theta})\, \pi_{\text{pr}}(\boldsymbol{\theta}). \tag{1}$$

The expression (1) casts the solution of the inverse problem as a probability density for the model inputs $\boldsymbol{\theta}$.

This chapter focuses on methods to approximate the input–output map $\mathcal{G}(\cdot)$. In physical problems, $\mathcal{G}$ usually contains a deterministic map from $\boldsymbol{\theta}$ to some idealized observations $\mathbf{d}_0$; interaction of $\mathbf{d}_0$ and $\boldsymbol{\theta}$ with the random variable $\boldsymbol{\eta}$ then yields the actual data. We set aside this interaction and concentrate on the deterministic map, which can be broken into two elements: the state equations that describe the evolution of the state $\mathbf{u}$ in response to the input $\boldsymbol{\theta}$, and the output equations that define the relationship between outputs of interest $\mathbf{y}$—which in turn define the observations $\mathbf{d}$—and state $\mathbf{u}$. We present a general nonlinear model in the semi-discrete form that might result from spatial discretization of a partial differential equation (using, for example, a finite element or finite volume method), or that might represent a set of differential-algebraic equations (such as arises in circuit modeling). The semi-discrete equations describing such a system are

$$\dot{\mathbf{u}} = \mathbf{f}(\mathbf{u}, \boldsymbol{\theta}, t), \tag{2}$$

$$\mathbf{y} = \mathbf{h}(\mathbf{u}, \boldsymbol{\theta}, t), \tag{3}$$

with initial condition

$$\mathbf{u}(\boldsymbol{\theta}, 0) = \mathbf{u}^0(\boldsymbol{\theta}). \tag{4}$$

In this semi-discrete model, $\mathbf{u}(\boldsymbol{\theta}, t) \in \mathbb{R}^N$ is the discretized state vector of N unknowns, and $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ is the vector of p parametric inputs on some domain Θ. The vector $\mathbf{u}^0(\boldsymbol{\theta}) \in \mathbb{R}^N$ is the specified initial state, t is time, and the dot indicates a derivative with respect to time. The nonlinear discretized residual vector $\mathbf{f} : \mathbb{R}^N \times \mathbb{R}^p \times [0, \infty) \to \mathbb{R}^N$ is for generality written as a function of the state, parameter inputs, and time. Our observable outputs are represented in spatially discretized form with q components in the vector $\mathbf{y}(\boldsymbol{\theta}, t) \in \mathbb{R}^q$, and defined by the general nonlinear function $\mathbf{h} : \mathbb{R}^N \times \mathbb{R}^p \times [0, \infty) \to \mathbb{R}^q$.

If the governing equations are linear in the state (or if a linearized model of (2) and (3) is derived by linearizing about a steady state), then the system is written

$$\dot{\mathbf{u}} = \mathbf{A}(\boldsymbol{\theta})\mathbf{u} + \mathbf{g}(\boldsymbol{\theta}, t), \tag{5}$$

$$\mathbf{y} = \mathbf{H}(\boldsymbol{\theta})\mathbf{u}, \tag{6}$$

where $\mathbf{A} \in \mathbb{R}^{N \times N}$ is a matrix that possibly depends on the parameters but not on the state, and the general nonlinear function $\mathbf{g} \in \mathbb{R}^N$ represents the direct contributions

of the parameters to the governing equations as well as forcing due to boundary conditions and source terms. For the linearized output equation, $\mathbf{H} \in \mathbb{R}^{q \times N}$ is a matrix that maps states to outputs, and we have assumed no direct dependence of outputs on parameters (without loss of generality—such a term could easily be incorporated).

Note that in both (3) and (6), the output vector $\mathbf{y}$ is continuous in time. In most practical situations and in the Bayesian formulation described above, the observational data are finite-dimensional on some domain $\mathcal{D}$; thus $\mathbf{d} \in \mathcal{D} \subseteq \mathbb{R}^M$ consists of $\mathbf{y}$ evaluated at a finite set of times, $\mathbf{d} = (\mathbf{y}(t_1), \ldots, \mathbf{y}(t_{n_t}))$ with $M = n_t q$. We denote the deterministic forward model mapping inputs $\boldsymbol{\theta}$ to finite-time observations $\mathbf{d}$ by $\mathbf{G}(\boldsymbol{\theta}) : \Theta \to \mathcal{D}$.

With nonlinear forward models, the inverse solution is typically represented by samples simulated from the posterior distribution (1). From these samples, posterior moments, marginal distributions, and other summaries can be evaluated. Solution of the statistical inverse problem thus requires many evaluations of the forward model (typically many thousands or even millions)—a computationally prohibitive proposition for large-scale systems. In the following two sections, we describe approaches to accelerate solution of the forward model—model reduction via state projection onto a reduced basis, and a stochastic spectral approach based on generalized polynomial chaos.

0.4 Model reduction

Model reduction seeks to derive a low-complexity model of the system that is fast to solve but preserves accurately the relationship between input parameters $\boldsymbol{\theta}$ and outputs $\mathbf{y}$, represented by the input–output map $\mathcal{G}$. If the system forward solves required to evaluate samples drawn from the posterior distribution are performed using a reduced-order model as a surrogate for the large-scale system, then it becomes tractable to carry out the many thousands of iterations required to solve the statistical inverse problem. Most large-scale model reduction frameworks are based on a projection approach, which is described in general terms in this section. We briefly describe computation of the basis via POD and reduced basis methods, and then discuss various options that have been proposed for sampling the parameter space.

0.4.1 General projection framework

The first step in creating a projection-based reduced-order model is to approximate the N-dimensional state $\mathbf{u}(\boldsymbol{\theta}, t)$ by a linear combination of n basis vectors,

$$\mathbf{u} \approx \boldsymbol{\Phi} \mathbf{u}_r, \tag{7}$$

where $n \ll N$. The projection matrix $\boldsymbol{\Phi} \in \mathbb{R}^{N \times n}$ contains as columns the basis vectors $\boldsymbol{\phi}_i$, i.e., $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \ \cdots \ \boldsymbol{\phi}_n]$, and the vector $\mathbf{u}_r(\boldsymbol{\theta}, t) \in \mathbb{R}^n$ contains the corresponding modal amplitudes.

This approximation can be employed in the general nonlinear system (2) or the general linear system (5), in either case resulting in a residual (since in general the

N equations cannot all be satisfied with $n \ll N$ degrees of freedom). We define a left basis $\boldsymbol{\Psi} \in \mathbb{R}^{N \times n}$ so that $\boldsymbol{\Psi}^T \boldsymbol{\Phi} = \mathbf{I}$. Using a Petrov-Galerkin projection, we require the residual to be orthogonal to the space spanned by the columns of $\boldsymbol{\Psi}$. In the nonlinear case, this yields the reduced-order model of (2)–(4) as

$$\dot{\mathbf{u}}_r = \boldsymbol{\Psi}^T \mathbf{f}(\boldsymbol{\Phi}\mathbf{u}_r, \boldsymbol{\theta}, t), \tag{8}$$

$$\mathbf{y}_r = \mathbf{h}(\boldsymbol{\Phi}\mathbf{u}_r, \boldsymbol{\theta}, t), \tag{9}$$

$$\mathbf{u}_r(\boldsymbol{\theta}, 0) = \boldsymbol{\Psi}^T \mathbf{u}^0(\boldsymbol{\theta}), \tag{10}$$

where $\mathbf{y}_r(\boldsymbol{\theta}, t) \in \mathbb{R}^q$ is the reduced model approximation of the output $\mathbf{y}(\boldsymbol{\theta}, t)$. For the linear system (5), (6) with initial condition (4), the reduced-order model is

$$\dot{\mathbf{u}}_r = \mathbf{A}_r(\boldsymbol{\theta})\mathbf{u}_r + \boldsymbol{\Psi}^T \mathbf{g}(\boldsymbol{\theta}, t), \tag{11}$$

$$\mathbf{y}_r = \mathbf{H}_r(\boldsymbol{\theta})\mathbf{u}_r, \tag{12}$$

$$\mathbf{u}_r(\boldsymbol{\theta}, 0) = \boldsymbol{\Psi}^T \mathbf{u}^0(\boldsymbol{\theta}), \tag{13}$$

where $\mathbf{A}_r(\boldsymbol{\theta}) = \boldsymbol{\Psi}^T \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\Phi}$ and $\mathbf{H}_r(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\Phi}$.

As in the full problem described in Section 0.3, the reduced model output $\mathbf{y}_r$ is continuous in time. Following our notation for the full problem, we define the reduced model data vector $\mathbf{d}_r \in \mathcal{D} \subseteq \mathbb{R}^M$ to consist of $\mathbf{y}_r$ evaluated at a finite set of times, $\mathbf{d}_r = (\mathbf{y}_r(t_1), \ldots, \mathbf{y}_r(t_{n_t}))$ with $M = n_t q$. We denote the reduced forward model of state dimension n that maps $\boldsymbol{\theta}$ to $\mathbf{d}_r$ by $\mathbf{G}^n_{\mathrm{ROM}} : \Theta \to \mathcal{D}$.

Our approach to reducing the computational cost of solving the inverse problem is to use the reduced model $\mathbf{G}^n_{\mathrm{ROM}}$ in place of the full model $\mathbf{G}$ in the likelihood function $\pi(\mathbf{d}|\boldsymbol{\theta})$. This can lead to a dramatic reduction in the cost of an evaluation of the posterior distribution (1). However, care must be taken to ensure efficient construction and solution of the reduced-order models (8)–(10) or (11)–(13). In the case of general nonlinear parametric dependence, these models have low dimension but are not necessarily fast to solve, since for each new parameter $\boldsymbol{\theta}$, solution of the ROM requires evaluating the large-scale system matrices or residual, projecting those matrices/residual onto the reduced subspace, and then solving the resulting reduced-order model. Since many elements of these computations depend on N, the dimension of the large-scale system, in general this process will not be computationally efficient. One option is to employ linearization of the parametric dependence (Daniel et al. 2004; Grepl and Patera 2005; Veroy et al. 2003). A more general approach is to employ the missing point estimation approach (Astrid et al. 2008), which approximates nonlinear terms in the reduced-order model with selective spatial sampling, or the coefficient-function approximation (Barrault et al. 2004; Grepl et al. 2007), which replaces nonlinear parametric dependencies with a reduced-basis expansion and then uses interpolation to efficiently compute the coefficients of that expansion for new parameter values.

0.4.2 Computing the basis

The basis vectors can be calculated with several techniques. Methods to compute the basis in the large-scale setting include approximate balanced truncation (Gugercin and Antoulas 2004; Li and White 2002; Moore 1981; Penzl 2006; Sorensen and Antoulas 2002), Krylov-based methods (Feldmann and Freund 1995; Gallivan et al. 1994; Grimme 1997), proper orthogonal decomposition (POD) (Deane et al. 1991; Holmes et al. 1996; Sirovich 1987), and reduced basis methods (Fox and Miura 1971; Noor and Peters 1980). Balanced truncation and Krylov-based methods are largely restricted to linear systems, thus here we focus on a brief description of reduced basis methods and POD. While the development of these methods has been in the context of reduction of the forward problem for simulation and (to some extent) control, recent work has shown applicability to the inverse problem setting, by introducing strategies that use characteristics of the inverse problem (including Hessian and prior information) to inform computation of the basis (Galbally et al. 2010; Lieberman 2009; Nguyen 2005). We discuss some of these strategies in the next subsection.

In the reduced basis and POD methods, the basis is formed as the span of a set of state solutions, commonly referred to as snapshots. These snapshots are computed by solving the system (2) or (5) for selected values of the parameters $\boldsymbol{\theta}$. In the POD method of snapshots (Sirovich 1987), the resulting state solutions at selected times and/or parameter values are collected in the columns of the matrix $\mathbf{U} \in \mathbb{R}^{n \times n_s}$,

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}^1 & \mathbf{u}^2 & \dots & \mathbf{u}^{n_s} \end{bmatrix}, \tag{14}$$

where $\mathbf{u}^i$ is the ith snapshot and n_s is the total number of snapshots, which depends on both the number of parameter values considered and the number of timesteps sampled for each parameter value.

The POD basis is given by the left singular vectors of the matrix $\mathbf{U}$ that correspond to the largest singular values. A basis of dimension n is thus

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}^1 & \boldsymbol{\phi}^2 & \dots & \boldsymbol{\phi}^n \end{bmatrix}, \tag{15}$$

where $\boldsymbol{\phi}^i$ is the ith left singular vector of $\mathbf{U}$, which has corresponding singular value σ_i. It can be shown that the POD basis is optimal in the sense that, for a basis of dimension n, it minimizes the least squares error of the representation of the snapshots in the reduced basis. This error is given by the sum of the squares of the singular values corresponding to those modes not included in the basis:

$$\sum_{i=1}^{n_s} \|\mathbf{u}^i - \boldsymbol{\Phi}\boldsymbol{\Phi}^T\mathbf{u}^i\|_2^2 = \sum_{j=n+1}^{n_s} \sigma_j^2. \tag{16}$$

It is however important to note that this optimality of representation of the snapshot set in the general case provides no corresponding rigorous error bound on the resulting POD-based reduced-order model.

Since the POD basis is orthogonal, a common choice for the left basis is $\mathbf{\Psi} = \mathbf{\Phi}$. Other choices for $\mathbf{\Psi}$ are also possible, such as one that minimizes a least-squares weighted-residual (Bui-Thanh et al. 2008; Maday et al. 2002; Rovas 2003; Rozza and Veroy 2006), or one that includes output information through use of adjoint solutions (Lall et al. 2002; Willcox and Peraire 2002).

0.4.3 Computing a basis for inverse problem applications: sampling the parameter space

A critical issue in computing the reduced basis is sampling of the parameter space: the quality of the resulting reduced-order model is highly dependent on the choice of parameters for which snapshots are computed. This is particularly important in the statistical inverse problem setting, since solution of the inverse problem will likely require broad exploration of the parameter space. Furthermore, problems of interest may have high-dimensional parameter spaces and the range of parameters explored in solving the inverse problem may not be known a priori.

Sampling methods to build reduced-order models must thus address two challenges. First, a systematic strategy is needed to choose where and how many samples to generate. Second, the strategy must be scalable so that parameter spaces of high dimension can be effectively sampled with a small number of large-scale system solves. Standard sampling schemes such as uniform sampling (uniform gridding of the parameter space) or random sampling are one option for creating snapshots. However, if the dimension of the parameter space is large, uniform sampling will quickly become too computationally expensive due to the combinatorial explosion of samples needed to cover the parameter space. Random sampling, on the other hand, might fail to recognize important regions in the parameter space. One can use knowledge of the application at hand to determine representative parametric inputs, as has been done to sample the parameter space for the quasi-convex optimization relaxation method (Sou et al. 2005), and to generate a POD or Krylov basis for problems in which the number of input parameters is small. Some examples of applications of parameterized model reduction with a small number of parameters include structural dynamics (Allen et al. 2004), aeroelasticity (Amsallem et al. 2007), Rayleigh-Bénard convection (Ly and Tran 2001), design of interconnect circuits (Bond and Daniel 2005; Daniel et al. 2004), and parameters describing inhomogeneous boundary conditions for parabolic PDEs (Gunzburger et al. 2007). For optimal control applications, online adaptive sampling has been employed as a systematic way to generate snapshot information (Afanasiev and Hinze 2001; Fahl and Sachs 2003; Hinze and Volkwein 2005; Kunisch and Volkwein 1999). However, these methods have not been scaled to problems that contain more than a handful of parameters.

To address the challenge of sampling a high-dimensional parameter space to build a reduced basis, the greedy sampling method was introduced in Grepl (2005); Grepl and Patera (2005); Veroy and Patera (2005); Veroy et al. (2003) to adaptively choose samples by finding the location at which the estimate of the error in

the reduced model is maximum. The greedy sampling method was applied to find reduced models for the parameterized steady incompressible Navier-Stokes equations (Veroy and Patera 2005). It was also combined with *a posteriori* error estimators for parameterized parabolic PDEs, and applied to several optimal control and inverse problems (Grepl 2005; Grepl and Patera 2005). In Bui-Thanh et al. (2008), the greedy sampling approach was formulated as a sequence of adaptive model-constrained optimization problems that were solved to determine appropriate sample locations. Unlike other sampling methods, this model-constrained optimization sampling approach incorporates the underlying physics and scales well to systems with a large number of parameters. In Lieberman (2009), the optimization-based greedy sampling approach was extended to the inverse problem setting by formulating the sequence of optimization problems to also include the prior probability density. Lieberman (2009) also addressed directly the challenge of high-dimensional parameter spaces in MCMC sampling by performing reduction in both state and parameter, and demonstrated the approach on a subsurface model with a distributed parameter representing the hydraulic conductivity over the domain.

0.5 Stochastic spectral methods

Based on polynomial chaos (PC) representations of random variables and processes (Cameron and Martin 1947; Debusschere et al. 2004; Ghanem and Spanos 1991; Wan and Karniadakis 2005; Wiener 1938; Xiu and Karniadakis 2002), 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 exploit regularity in the dependence of an output or solution field on uncertain parameters. They constitute attractive alternatives to Monte Carlo simulation in numerous applications: transport in porous media (Ghanem 1998), structural mechanics, thermo-fluid systems (Le Maître et al. 2001, 2002), electrochemistry (Debusschere et al. 2003), and reacting flows (Najm et al. 2009; Reagan et al. 2004).

Stochastic spectral methods have more recently been applied in the *inverse* context (Marzouk et al. 2007), for both point and spatially distributed parameters (Marzouk and Najm 2009). Here, the essential idea is to construct a stochastic forward problem whose solution approximates the deterministic forward model over the support of the prior. This procedure—effectively propagating prior uncertainty through the forward model—yields a polynomial approximation of the forward model's dependence on uncertain parameters θ. The polynomial approximation then enters the likelihood function, resulting in a "surrogate" posterior density that is inexpensive to evaluate, often orders of magnitude less expensive than the original posterior.

0.5.1 Surrogate posterior distribution

Here we describe the construction of a polynomial-chaos based surrogate posterior in greater detail. 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 $\boldsymbol{\theta}$ of this representation. For instance, if the unknown quantity is a field endowed with a Gaussian process prior, the finite representation may be a truncated K-L expansion with mode strengths $\boldsymbol{\theta}$ and priors $\theta_i \sim N(0,1)$. Let $\Theta \subseteq \mathbb{R}^p$ denote the support of the prior. The Bayesian formulation in Section 0.3 describes the inverse solution in terms of the posterior density of $\boldsymbol{\theta}$, which entails evaluations of the forward model $\mathbf{G}(\boldsymbol{\theta}) : \Theta \to \mathcal{D}$, with $\mathcal{D} \subseteq \mathbb{R}^M$.

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

$$\check{\theta}_i = c_i(\check{\boldsymbol{\xi}}) = \sum_{|\mathbf{k}| \leq P} c_{i\mathbf{k}} \Psi_{\mathbf{k}}(\check{\boldsymbol{\xi}}), \qquad i = 1, \ldots, p. \tag{17}$$

Here $\check{\boldsymbol{\xi}}$ is a vector of p independent and identically distributed (i.i.d.) random variables, $\mathbf{k} = (k_1, \ldots, k_p)$ is a multi-index with magnitude $|\mathbf{k}| \equiv k_1 + \ldots + k_p$, and $\Psi_{\mathbf{k}}$ are multivariate polynomials (of degree k_i in coordinate $\check{\xi}_i$) orthogonal with respect to the measure on $\check{\boldsymbol{\xi}}$ (Ghanem and Spanos 1991). The total polynomial order of the PC basis is truncated at P. The vector $\check{\boldsymbol{\theta}}$ will serve as an input to $\mathbf{G}$, thus specifying a *stochastic forward problem*.

Note that the distribution of $\check{\boldsymbol{\xi}}$ (e.g., standard normal, Beta, etc.) and the corresponding polynomial form of Ψ (e.g., Hermite, Jacobi, etc.) are intrinsic properties of the PC basis (Xiu and Karniadakis 2002). In the simplest construction, PC coefficients in (17) are then chosen such that $\check{\boldsymbol{\theta}}$ is distributed according to the prior on $\boldsymbol{\theta}$. This is not a strict requirement, however. A necessary condition on $\mathbf{c}$ is that $\Xi_\theta = \mathbf{c}^{-1}[\Theta]$, the inverse image of the support of the prior, be contained within the range of $\check{\boldsymbol{\xi}}$. This condition ensures that there is a realization of $\check{\boldsymbol{\xi}}$ corresponding to every feasible value of $\boldsymbol{\theta}$.

Having defined the stochastic forward problem, we can solve it with a Galerkin or collocation procedure (see Section 0.5.2 below), thus obtaining a PC representation for each component of the model output. Here G_i is the i-th component of $\mathbf{G}$ and G_i^P is its Pth-order PC approximation:

$$G_i^P(\check{\boldsymbol{\xi}}) = \sum_{|\mathbf{k}| \leq P} g_{i\mathbf{k}} \Psi_{\mathbf{k}}(\check{\boldsymbol{\xi}}), \qquad i = 1, \ldots, M. \tag{18}$$

The forward solution $\mathbf{G}^P$ obtained in this fashion is a polynomial function of $\check{\boldsymbol{\xi}}$. Evaluating $\mathbf{G}^P$ with a deterministic argument,[1] it can be viewed simply as a polynomial approximation of $\mathbf{G} \circ \mathbf{c}$, where $\circ$ denotes composition. We will use this approximation to replace $\mathbf{G}$ in the likelihood function $L(\boldsymbol{\theta})$.

Consider the simple case of additive noise, $\mathbf{d} = \mathcal{G}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbf{G}(\boldsymbol{\theta}) + \boldsymbol{\eta}$, such that $L(\boldsymbol{\theta}) = \pi_\eta(\mathbf{d} - \mathbf{G}(\boldsymbol{\theta}))$, with π_η being the probability density of $\boldsymbol{\eta}$. The likelihood

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

function can be rewritten as a function of $\boldsymbol{\xi}$:

$$L\left(\mathbf{c}(\boldsymbol{\xi})\right) = \pi_\eta\left(\mathbf{d} - \mathbf{G}\left(\mathbf{c}(\boldsymbol{\xi})\right)\right) \approx \pi_\eta\left(\mathbf{d} - \mathbf{G}^P(\boldsymbol{\xi})\right). \tag{19}$$

This change of variables from $\boldsymbol{\theta}$ to $\boldsymbol{\xi}$ lets us define a posterior density for $\boldsymbol{\xi}$:

$$\pi_\xi(\boldsymbol{\xi}) \propto L\left(\mathbf{c}(\boldsymbol{\xi})\right) \pi_{\text{pr}}\left(\mathbf{c}(\boldsymbol{\xi})\right) \det D\mathbf{c}(\boldsymbol{\xi}). \tag{20}$$

In this expression, $D\mathbf{c}$ is the Jacobian of $\mathbf{c}$, $\det$ denotes the determinant, and π_{pr} is the prior density of $\boldsymbol{\theta}$. The last two factors on the right side, $\pi_{\text{pr}}\left(\mathbf{c}(\boldsymbol{\xi})\right) \det D\mathbf{c}(\boldsymbol{\xi})$, are the probability density on $\boldsymbol{\xi}$ that corresponds to the prior on θ. Replacing the forward model in the likelihood function via (19) then yields the *surrogate* posterior density π_ξ^P:

$$\pi_\xi(\boldsymbol{\xi}) \approx \pi_\xi^P(\boldsymbol{\xi}) \propto \pi_\eta\left(\mathbf{d} - \mathbf{G}^P(\boldsymbol{\xi})\right) \pi_{\text{pr}}\left(\mathbf{c}(\boldsymbol{\xi})\right) \left|\det D\mathbf{c}(\boldsymbol{\xi})\right|. \tag{21}$$

Despite the change of variables, it is straightforward to recover the posterior expectation of an arbitrary function f:

$$\mathbb{E}_{\pi_{\text{post}}} f = \mathbb{E}_{\pi_\xi}(f \circ \mathbf{c}) \tag{22}$$

where $\pi_{\text{post}} \equiv \pi(\boldsymbol{\theta}|\mathbf{d})$ is the posterior density on Θ, and π_ξ is the corresponding posterior density of $\boldsymbol{\xi}$.

The surrogate posterior 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 solving the stochastic forward problem, i.e., obtaining the PC expansions in (18). 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. Moreover, as it requires only the forward model and the prior, the stochastic forward solution may be obtained "offline," independently of the data. Accuracy of the surrogate posterior depends on the order and family of the PC basis, as well as on the choice of transformation $\mathbf{c}$—for instance, whether the distribution of $\check{\boldsymbol{\theta}}$ assigns sufficient probability to regions of Θ favored by the posterior. A detailed discussion of these issues can be found in Marzouk et al. (2007). Some convergence results are summarized below.

0.5.2 Forward solution methodologies and convergence results

Solution of the stochastic forward problem (18) is an essential step in the inversion procedure outlined above. While a survey of polynomial chaos methods for solving ODEs and PDEs with random inputs is beyond the scope of this chapter (see for instance Najm (2009); Xiu (2009)), we highlight two broad classes of approaches. Stochastic Galerkin methods (Ghanem and Spanos 1991; Le Maître et al. 2001; Matthies and Keese 2005) involve a reformulation of the governing equations, essentially creating a larger system of equations for the PC coefficients $g_{i\mathbf{k}}$; these equations

are generally coupled, though one may take advantage of problem-specific structure in devising efficient solution schemes (Xiu and Shen 2009). Stochastic collocation methods (Babuška et al. 2007; Xiu and Hesthaven 2005), on the other hand, are "non-intrusive"; these require only a finite number of uncoupled deterministic simulations, with no reformulation of the governing equations of the forward model. Collocation methods using sparse grids (Bieri et al. 2009; Ganapathysubramanian and Zabaras 2007; Ma and Zabaras 2009; Nobile et al. 2008; Smolyak 1963; Xiu and Hesthaven 2005), offer great efficiency and ease of implementation for higher-dimensional problems.

For systems with more complex dynamics—discontinuities or bifurcations with respect to uncertain parameters, or even limit cycles (Beran et al. 2006)—global bases may be unsuitable. Instead, piecewise polynomial (Wan and Karniadakis 2005, 2009) or multi-wavelet (Le Maître et al. 2004) generalizations of polynomial chaos enable efficient propagation of uncertainty; such bases can also be used to construct surrogate posteriors. Indeed, the overall Bayesian inference scheme is quite flexible with regard to *how* one chooses to solve a stochastic forward problem. Error analysis of the Bayesian stochastic spectral framework (Marzouk and Xiu 2009) has reinforced this flexibility. The relevant convergence results can be summarized as follows. Consider the mean-square error, with respect to π_{pr}, in the forward solution: $e(P) \equiv \|\mathbf{G}^P(\boldsymbol{\xi}) - \mathbf{G}(\boldsymbol{\xi})\|_{L^2_{\pi_{\text{pr}}}}$. Suppose that observational errors η_i are additive and i.i.d. Gaussian. If $e(P)$ converges at a particular rate, $e(P) \leq CP^{-\alpha}$, then at sufficiently large P, the Kullback-Leibler (KL) divergence of the true posterior from the surrogate posterior maintains at least the same rate of convergence, $D\left(\pi_\xi^P \| \pi_\xi\right) \lesssim P^{-\alpha}$. In particular, exponential convergence of the forward solution implies exponential convergence of the surrogate posterior to the true posterior. (Recall that the Kullback-Leibler divergence quantifies the difference between probability distributions in information theoretic terms (Gibbs and Su 2002).) These results provide a guideline relevant to *any* approximation of the forward model.

0.6 Illustrative example

We explore the relative advantages and disadvantages of model reduction and stochastic spectral approaches in the context of a simple transient source inversion problem. Consider a dimensionless diffusion equation on a square domain $\Omega = [0,1] \times [0,1]$ with adiabatic boundaries:

$$\frac{\partial u}{\partial t} = \nabla^2 u + \frac{s}{2\pi\gamma^2} \exp\left(-\frac{|\boldsymbol{\theta} - \mathbf{x}|^2}{2\gamma^2}\right)[1 - H(t-T)], \tag{23}$$

$$\nabla u \cdot \mathbf{n} = 0 \quad \text{on } \partial\Omega, \tag{24}$$

$$u(\mathbf{x}, 0) = 0 \quad \text{in } \Omega. \tag{25}$$

The solution field $u(\mathbf{x}, t)$ can represent temperature or the concentration of some contaminant species, with $\mathbf{x} \equiv (x_1, x_2) \in \Omega$ and time $t \geq 0$. $H(t)$ denotes the unit

step function. Thus, the source term in (23) comprises a single localized source, active on the interval $t \in \mathcal{T} = [0, T]$ and centered at location $\boldsymbol{\theta} \in \Theta = \Omega$ with strength s and characteristic width γ.

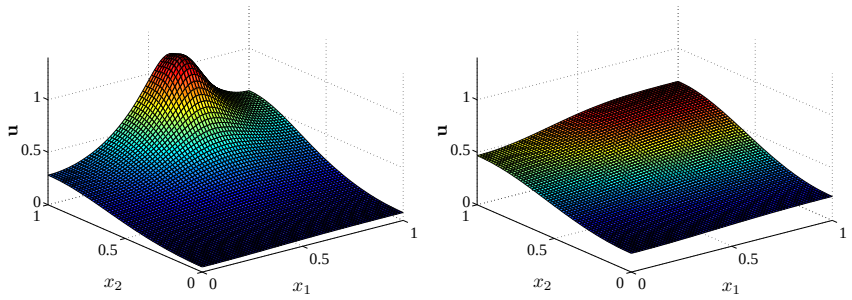
The governing equations (23)–(25) are discretized on a uniform spatial grid using a second-order-accurate finite difference scheme. This spatial discretization and the application of the boundary conditions lead to a semi-discrete system of the form (5)–(6), linear in the state but nonlinear in the parameters $\boldsymbol{\theta}$. The state vector $\mathbf{u}(t)$ contains $u(\mathbf{x}, t)$ evaluated at the N grid points; the sparse matrix $A \in \mathbb{R}^{N \times N}$ reflects the spatial discretization and application of the Neumann boundary conditions; and $\mathbf{g}(\boldsymbol{\theta})$ is a nonlinear function representing the source term in Eq. (23). Note that $\mathbf{g}$ is here just a function of the parameters (the source location) and not the state.

In the inverse problem, we are given noisy observations of the solution $u(\mathbf{x}, t)$ at a few locations in space and a few instants in time. From these data, we wish to infer the source location $\boldsymbol{\theta} = (\theta_1, \theta_2)$. For simplicity, we assume that the shutoff time T, strength s, and source width γ are known. We assume that observations of u are available for $n_t = 3$ time instants, $t \in \{0.1, 0.2, 0.3\}$, at $q = 9$ locations on a uniform 3×3 grid covering the domain Ω. The forward model $\mathbf{G}(\boldsymbol{\theta})$ is thus a map from the source location $\boldsymbol{\theta}$ to noise-free observations $\mathbf{d}_0 \in \mathbb{R}^{qn_t}$. These observations are perturbed with additive Gaussian noise $\boldsymbol{\eta}$ to yield the data vector $\mathbf{d} = \mathbf{d}_0 + \boldsymbol{\eta}$. Components of $\boldsymbol{\eta}$ are i.i.d., $\boldsymbol{\eta} \sim N\left(0, \sigma^2 I\right)$. The likelihood function is therefore given by $L(\boldsymbol{\theta}) = \pi_\eta\left(\mathbf{d} - \mathbf{G}(\boldsymbol{\theta})\right)$. The prior on $\boldsymbol{\theta}$ reflects a uniform probability assignment over the entire domain of possible source locations, $\theta_i \sim U(0, 1)$. The posterior density is then

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{d}) &\propto \pi_\eta\left(\mathbf{d} - \mathbf{G}(\boldsymbol{\theta})\right) \mathbf{1}_\Omega(\boldsymbol{\theta}) \\ &= \begin{cases} \exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{d} - \mathbf{G}(\boldsymbol{\theta})\right)^T \left(\mathbf{d} - \mathbf{G}(\boldsymbol{\theta})\right)\right) & \text{if} \quad \boldsymbol{\theta} \in \Omega, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{26}$$

Figure 1 shows an example forward solution, obtained on a 69×69 uniform grid. The plots show the solution field $u(\mathbf{x}, t)$ before and after the source shutoff time of $T = 0.2$. The source is located at $\boldsymbol{\theta} = (0.6, 0.9)$, with strength $s = 2$ and width $\gamma = 0.05$. The solution field at the earlier time is peaked around the source location and contains useful information for the inverse problem. After the shutoff time, however, the field tends to flatten out due to diffusion. Eventually observations of the u-field at times well after the shutoff will provide no useful information for inference of the source location.

We now consider solution of the inverse problem using two approximation approaches—a POD-based reduced order model and a polynomial chaos surrogate obtained with pseudospectral stochastic collocation. Constructing the approximate posterior distribution in *either* case requires evaluating the forward model at a number of parameter values $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_Q\}$. With POD, these forward model evaluations are used to construct the snapshot matrix, while for stochastic collocation they are used to evaluate



(a) $t = 0.15$ (b) $t = 0.25$

Figure 1 Solution field $u(\mathbf{x}, t)$ of the full forward model at two different times for source parameters $\boldsymbol{\theta} = (0.6, 0.9)$, $T = 0.20$, $s = 2$, and $\gamma = 0.05$.

integrals for the PC coefficients:

$$g_{i\mathbf{k}}\langle\Psi_{\mathbf{k}}^2\rangle = \int_{\Theta} \mathbf{G}_i(\boldsymbol{\theta})\,\Psi_{\mathbf{k}}(\boldsymbol{\theta})\,\pi_{\mathrm{pr}}(\boldsymbol{\theta})\,d\boldsymbol{\theta} \tag{27}$$

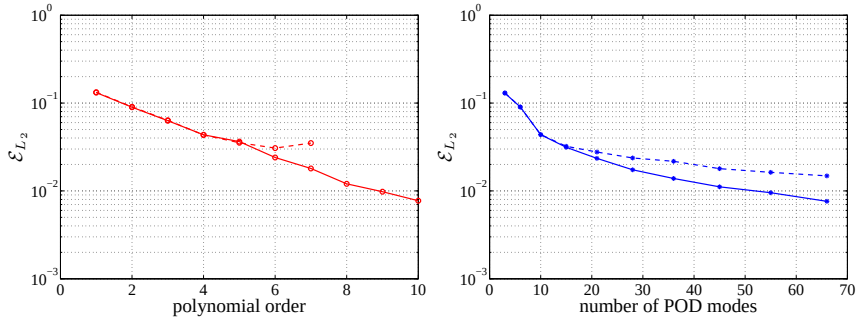
$$\approx \sum_{j=1}^{Q} \mathbf{G}_i(\boldsymbol{\theta}_j)\,\Psi_{\mathbf{k}}(\boldsymbol{\theta}_j)\,w_j. \tag{28}$$

Since the prior π_{pr} is uniform on Θ, the polynomials $\Psi_{\mathbf{k}}$ are taken to be bivariate Legendre polynomials, while the nodes and weights $\{\boldsymbol{\theta}_j, w_j\}_{j=1}^{Q}$ are chosen according to a Gaussian quadrature rule. In particular, we use a tensor product of l-point Gauss-Legendre rules on $[0, 1]$, such that $Q = l^2$. In the comparison below, we use the same nodal set of Q parameter values to construct the snapshot matrix (14). This is certainly not the only choice (or even the best choice) of parameter values to employ for POD. Our selection is motivated mostly by simplicity, so that identical forward simulations support both approximation techniques. We revisit this choice in later remarks.

We first evaluate the accuracy of the different forward models as a function of the "order" of approximation. The L^2 error of an approximate forward model $\widetilde{\mathbf{G}}$ is defined as

$$e = \int_{\Theta} \|\mathbf{G}(\boldsymbol{\theta}) - \widetilde{\mathbf{G}}(\boldsymbol{\theta})\|_2\,\pi_{\mathrm{pr}}(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \tag{29}$$

In other words, this is the prior-weighted error in model predictions integrated over the parameter space. The precise meaning of "order" depends on context, of course. With stochastic collocation, we take order to be the maximal polynomial degree P. In the projection approach, order is the dimension of the reduced model, i.e., the number of POD basis vectors n retained. Figure 2 shows the L^2 error for both



(a) Stochastic collocation, $\mathbf{G}^P_{\mathrm{SC}}$. (b) POD, $\mathbf{G}^n_{\mathrm{ROM}}$.

Figure 2 L^2 error in the approximate forward models, versus polynomial degree P for stochastic collocation and versus the number of modes n for POD. Dotted line represents $Q = 36$.

methods of approximation. In Figure 2(a), the error decreases more or less exponentially with polynomial degree, provided a sufficiently high order quadrature rule ($Q = 100$) is applied. Results with a 6-point quadrature rule in each direction ($Q = 36$) diverge for $P > 5$; this is understandable, as the degree of the integrand in (27) increases with $|\mathbf{k}|$ and aliasing errors corrupt the higher-degree polynomial coefficients. In Figure 2(b), error decreases with the number of POD modes n, but the accuracy of the reduced model at larger n depends on Q. For smaller Q—i.e., less dense coverage of the parameter space—the error begins to plateau at a larger value than for $Q = 100$. These results suggest that sufficiently large values of both Q and n are needed for an accurate reduced-order model.

Turning from the forward model approximation to the posterior distribution, Figure 3 shows the posterior density of the source location, $\pi(\boldsymbol{\theta}|\mathbf{d})$ (26), for various forward model approximations. The data $\mathbf{d}$ reflect the same source parameters used in Figure 1, i.e., $\boldsymbol{\theta} = (0.6, 0.9)$. Observations of the exact solution field are perturbed with Gaussian noise $\boldsymbol{\eta} \sim N\left(0, \sigma^2 I\right)$, with $\sigma = 0.2$. The noise magnitude is therefore roughly 20–40% of the nominal values of $\mathbf{d}_0$. Figure 3(a) shows the baseline case: contours of the exact posterior density, obtained via evaluations of the full forward model $\mathbf{G}$. Plots (b) and (c) show contours of the approximate posterior density π^P_{SC} obtained by evaluation of the stochastic collocation model $\mathbf{G}^P_{\mathrm{SC}}$, at polynomial orders $P = 2$ and $P = 10$, respectively. Plots (d) and (e) show the contours of the approximate posterior density π^n_{ROM} obtained with POD models $\mathbf{G}^n_{\mathrm{ROM}}$ of dimension $n = 6$ and $n = 66$, respectively. Both the stochastic collocation and POD models were constructed with $Q = 100$ forward evaluations, at parameter values $\boldsymbol{\theta}$ chosen with a 10-point quadrature rule in each direction. At sufficiently high P or n, both types of models yield close agreement with the true posterior density. Note that the true posterior is multi-modal; all three modes are well captured by the surrogate

posterior densities in Figure 3(d)–(e).

A more quantitative measurement of posterior error is the Kullback-Leibler (KL) divergence from the true posterior to the approximate posterior. Letting $\widetilde{\pi}$ denote the approximate posterior density, the KL divergence of $\widetilde{\pi}(\boldsymbol{\theta})$ from $\pi(\boldsymbol{\theta})$ is:

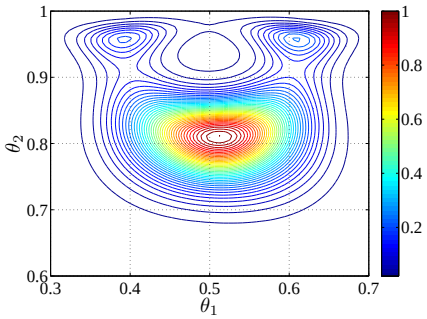
$$D(\pi\|\widetilde{\pi}) = \int_{\Theta} \pi(\boldsymbol{\theta}) \log \frac{\pi(\boldsymbol{\theta})}{\widetilde{\pi}(\boldsymbol{\theta})} \, \mathrm{d}\boldsymbol{\theta}. \tag{30}$$

Figure 4 shows $D\left(\pi\|\widetilde{\pi}\right)$ versus the order of approximation for the POD and stochastic collocation approaches. The true value of $\boldsymbol{\theta}$ and all of the other source parameters are identical to Figure 3; the same data vector $\mathbf{d}$ is used throughout. We contrast POD and collocation models constructed using either $Q = 36$ and $Q = 100$ nodes in the parameter space Θ. The integral in (30) was evaluated with the trapezoidal rule, using a uniform grid of dimension 69×69 on the set $\Theta' = [0.3, 0.7] \times [0.6, 1] \subset \Theta$. (Note that ignoring the remaining area of the parameter domain contributes negligible error, since the posterior density in these regions is nearly zero, as can be seen in Figure 3.)

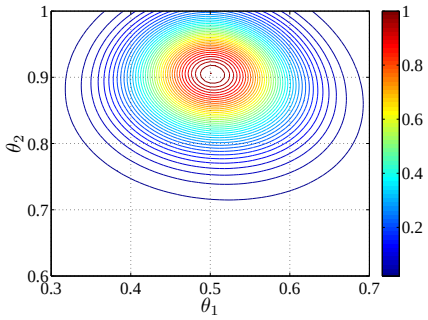
Figures 4(a) and 4(c) show that, provided the value of Q is sufficiently high, both approaches achieve similar posterior accuracy. As in Figure 2, the accuracy of the polynomial chaos-based surrogate posterior degrades due to aliasing errors when Q is too small; Q effectively limits the maximum polynomial degree that is meaningful to employ. When Q is larger, however, rapid convergence of $D\left(\pi\|\widetilde{\pi}\right)$ with respect to polynomial degree is observed. Error in the POD-based surrogate posterior also decays rapidly with increasing model order in the $Q = 100$ case. When fewer $\boldsymbol{\theta}$ values are used to train the POD model ($Q = 36$), errors tend to be larger. Interestingly, though, these errors start to decay anew for $n > 40$. With $n = 66$ modes, the errors associated with the $Q = 36$ and $Q = 100$ surrogate posteriors π^n_{ROM} differ by less than a factor of two.

Computational speedups over the full model, which takes approximately six seconds per forward simulation, are significant in both cases. But at similar levels of accuracy, the POD models are several times more costly in forward simulation time than the stochastic collocation models. Indeed, in the present case it is faster to evaluate a polynomial expansion (18) than to integrate a reduced-order model (11)–(13). On the other hand, if Q is small, then the stochastic collocation approach cannot achieve high accuracy, but error in the POD-based posterior continues to decline as more modes are added. For the current problem, therefore, if only a limited number of full model evaluations can be performed (offline), the POD approach is a better choice; while if Q can be chosen sufficiently high, the stochastic collocation approach yields equivalent accuracy with much smaller online computational cost.

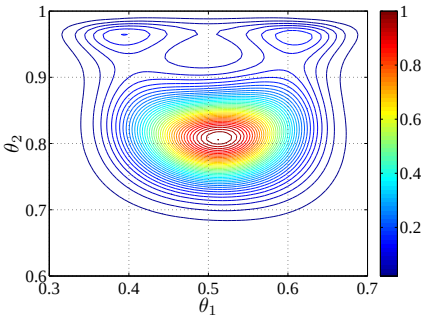
The computational times reported here and shown in Figure 4 were obtained on a desktop PC with an Intel Core 2 Duo processor at 3.16 GHz and 4 GB of RAM.



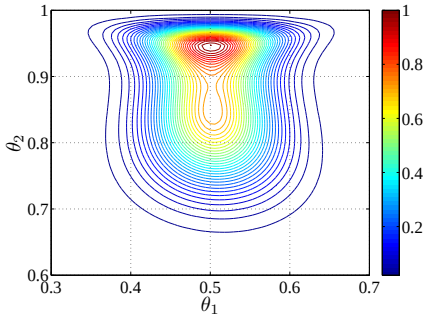
(a) Full forward model.



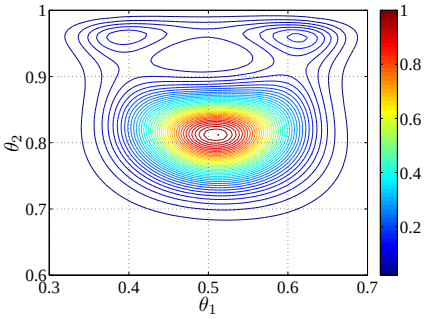
(b) Stochastic collocation at degree $P = 2$.



(c) Stochastic collocation at degree $P = 10$.

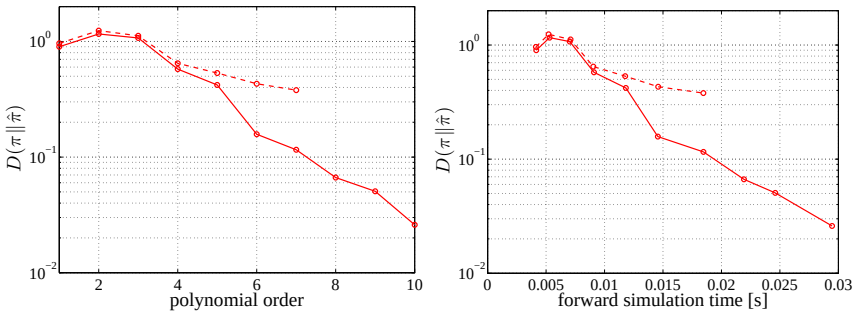


(d) POD with $n = 6$ modes.



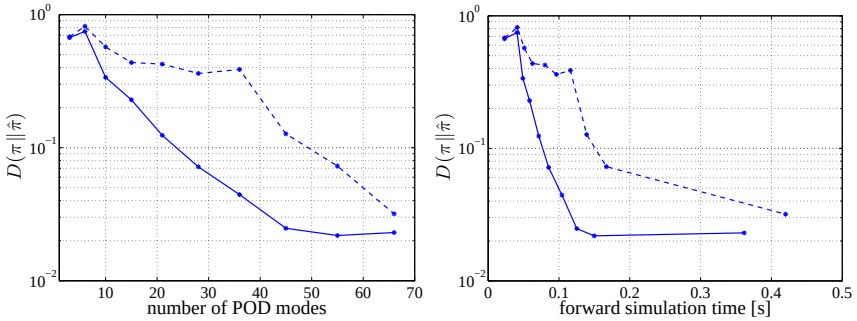
(e) POD with $n = 66$ modes.

Figure 3 Contours of the posterior density $\pi(\boldsymbol{\theta}|\mathbf{d})$ using the full forward model and various approximations. Contour lines are plotted at 40 equally spaced contour levels in the range $[0.02, 1]$.



(a) Posterior error vs. order, $\mathbf{G}^P_{\mathrm{SC}}$.

(b) Posterior error vs. simulation time, $\mathbf{G}^P_{\mathrm{SC}}$.



(c) Posterior error vs. order, $\mathbf{G}^n_{\mathrm{ROM}}$.

(d) Posterior error vs. simulation time, $\mathbf{G}^n_{\mathrm{ROM}}$.

Figure 4 Kullback-Leibler divergence from the exact posterior to the approximate posterior. Plots (a) and (b) show the results for stochastic collocation models of different polynomial degree. Plots (c) and (d) show the results for POD models of different order. In both cases, models are constructed with forward simulations at either $Q = 36$ (dotted) or $Q = 100$ (solid) nodes in parameter space.

0.7 Conclusions

The simple numerical example above suggests that, while model reduction may require fewer offline forward simulations to achieve a certain accuracy, polynomial chaos-based surrogate posteriors may be significantly cheaper to evaluate in the online phase. These conclusions are necessarily quite problem-specific, however. Our illustrative example was strongly nonlinear in the parameters, but these parameters were limited to two dimensions. The governing equations were also linear in the state $\mathbf{u}$. It is important to consider how the relative computational cost of these methods scales with dimensionality of the parameter space and with the dynamical complexity of the forward model, among other factors.

These questions underscore many challenges and open problems in surrogate and reduced-order modeling for statistical inverse problems. Techniques are needed to rigorously inform construction of the surrogate model with components of the inverse formulation, specifically incorporating both prior information and data. While some success has been demonstrated in this regard for simple problems, challenges remain in incorporating prior models for more complex fields (e.g., with discontinuities or other geometric structure) and in conditioning on data collected at multiple scales. High dimensionality also raises several open issues; successful surrogate modeling in this context should exploit the spectrum of the forward operator and any smoothing or structure provided by the prior in order to reduce the number of input parameters. Rigorous error bounds on the posterior computed using a surrogate or reduced-order model remain another outstanding challenge. Without some way of estimating the effects of using a surrogate in place of the full model, we cannot quantitatively answer questions such as how many samples are required to compute the basis.

Answers to these questions again may depend on the details of the algorithmic approaches. In the previous example, we focused on simplicity and did not apply more sophisticated approaches for either model reduction or polynomial chaos approximation. With model reduction, for example, we did not employ greedy sampling in the parameter space; with stochastic collocation, we did not employ sparse grids (particularly anisotropic and adaptive sparse grids for high-dimensional problems), nor did we explore partitioning of the prior support and/or alternate polynomial bases. In tackling more complicated problems, one should certainly draw from the literature in all of these areas in developing forward model approximation schemes suited to inverse problems.

Beyond reducing the complexity of forward simulations, what is needed even more are *combinations* of the approaches discussed in Section 0.2—simultaneously approximating the forward model, reducing the size of the input space, and reducing the requisite number of samples. Many of these methods will be described in the ensuing chapters.

Finally, we close by noting that this volume focuses on solution of the inverse problem—success is measured by our ability to accurately estimate some parameter of interest. However, in many scientific and engineering applications, the inverse

problem is merely one step on the path to solving an ultimate design or control problem; that is, the inference task is often followed by a decision problem downstream. Potentially significant opportunities exist for an integrated consideration of inference and decision problems. For example, the downstream decision problem can inform solution of the inverse problem by defining the level of accuracy to which specific features of the parameter should be resolved. The decision problem might also guide the construction of more parsimonious surrogate or reduced-order models.

Bibliography

Afanasiev K and Hinze M 2001 Adaptive control of a wake flow using proper orthogonal decomposition *Lecture Notes in Pure and Applied Mathematics* vol. 216 Marcel Dekker pp. 317–332.

Allen M, Weickum G and Maute K 2004 Application of reduced order models for the stochastic design optimization of dynamic systems *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, pp. 1–19, Albany, NY, USA. AIAA 2004-4614.

Amsallem D, Farhat C and Lieu T 2007 Aeroelastic analysis of F-16 and F-18/A configurations using adapted CFD-based reduced-order models *48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, vol. AIAA 2007-2364, pp. 1–20, Honolulu, Hawaii.

Arridge S, Kaipio J, Kolehmainen V, Schweiger M, Somersalo E, Tarvainen T and Vauhkonen M 2006 Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems* **22**, 175–195.

Astrid P, Weiland S, Willcox K and Backx T 2008 Missing point estimation in models described by proper orthogonal decomposition. *IEEE Transactions on Automatic Control, to appear*.

Babuška I, Nobile F and Tempone R 2007 A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis* **45**(3), 1005–1034.

Balakrishnan S, Roy A, Ierapetritou MG, Flach GP and Georgopoulos PG 2003 Uncertainty reduction and characterization for complex environmental fate and transport models: an empirical Bayesian framework incorporating the stochastic response surface method. *Water Resources Research* **39**(12), 1350.

Barrault M, Maday Y, Nguyen N and Patera A 2004 An "empirical interpolation" method: Application to efficient reduced-basis discretization of partial differential equations. *CR Acad Sci Paris Series.* **I**, 339–667.

Beran PS, Pettit CL and Millman DR 2006 Uncertainty quantification of limit-cycle oscillations.. *J. Comput. Phys.* **217**(1), 217 – 47.

Bieri M, Andreev R and Schwab C 2009 Sparse tensor discretizations of elliptic sPDEs. Technical report, ETH. Research Report No. 2009–07.

Bliznyuk N, Ruppert D, Shoemaker C, Regis R, Wild S and Mugunthan P 2008 Bayesian calibration of computationally expensive models using optimization and radial basis function approximation. *Journal of Computational and Graphical Statistics* **17**(2), 1–25.

Bond B and Daniel L 2005 Parameterized model order reduction of nonlinear dynamical systems *Proceedings of the IEEE Conference on Computer-Aided Design*. IEEE, San Jose, CA.

Bui-Thanh T, Willcox K and Ghattas O 2008 Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing* **30**(6), 3270–3288.

Cameron R and Martin W 1947 The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals. *Annals of Mathematics* **48**, 385–392.

Christen JA and Fox C 2005 MCMC using an approximation. *Journal of Computational and Graphical Statistics* **14**(4), 795–810.

Daniel L, Siong O, Chay L, Lee K and White J 2004 Multiparameter moment matching model reduction approach for generating geometrically parameterized interconnect performance models. *Transactions on Computer Aided Design of Integrated Circuits* **23**(5), 678–693.

Deane A, Kevrekidis I, Karniadakis G and Orszag S 1991 Low-dimensional models for complex geometry flows: Application to grooved channels and circular cylinders. *Phys. Fluids* **3**(10), 2337–2354.

Debusschere B, Najm H, Matta A, Knio O, Ghanem R and Le Maître O 2003 Protein labeling reactions in electrochemical microchannel flow: Numerical simulation and uncertainty propagation. *Physics of Fluids* **15**(8), 2238–2250.

Debusschere B, Najm H, Pébay P, Knio O, Ghanem R and Le Maître O 2004 Numerical challenges in the use of polynomial chaos representations for stochastic processes. *SIAM J. Sci. Comp.* **26**(2), 698–719.

Efendiev Y, Hou TY and Luo W 2006 Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM Journal on Scientific Computing* **28**, 776–803.

Eldred M, Giunta S and Collis S 2004 Second-order corrections for surrogate-based optimization with model hierarchies AIAA Paper 2004-4457, in Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference.

Fahl M and Sachs E 2003 Reduced order modelling approaches to PDE–constrained optimization based on proper orthogonal decompostion In *Large-Scale PDE-Constrained Optimization* (ed. Biegler L, Ghattas O, Heinkenschloss M and van Bloemen Waanders B) Lecture Notes in Computational Science and Engineering, Vol. 30. Springer-Verlag, Heidelberg.

Feldmann P and Freund R 1995 Efficient Linear Circuit Analysis by Padé Approximation via the Lanczos Process.. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **14**, 639–649.

Fox RL and Miura H 1971 An approximate analysis technique for design calculations. *AIAA Journal* **9**(1), 177–179.

Galbally D, Fidkowski K, Willcox K and Ghattas O 2010 Nonlinear model reduction for uncertainty quantification in large-scale inverse problems. *International Journal for Numerical Methods in Engineering* **81**(12), 1581–1608.

Gallivan K, Grimme E and Van Dooren P 1994 Padé Approximation of Large-Scale Dynamic Systems with Lanczos Methods Proceedings of the 33rd IEEE Conference on Decision and Control.

Ganapathysubramanian B and Zabaras N 2007 Sparse grid collocation schemes for stochastic natural convection problems. *J. Comput. Phys.* **225**(1), 652–685.

Geyer CJ 1991 Markov chain Monte Carlo maximum likelihood In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (ed. Keramidas EM), vol. 23, pp. 156–163. Interface Foundation of North America.

Ghanem R 1998 Probabilistic characterization of transport in heterogeneous media. *Comput. Methods Appl. Mech. Engrg.* **158**, 199–220.

Ghanem R and Spanos P 1991 *Stochastic Finite Elements: A Spectral Approach*. Springer Verlag, New York.

Gibbs AL and Su FE 2002 On choosing and bounding probability metrics. *International Statistical REview* **70**(3), 419–435.

Grepl M 2005 *Reduced-Basis Approximation and A Posteriori Error Estimation for Parabolic Partial Differential Equations* PhD thesis MIT Cambridge, MA.

Grepl M and Patera A 2005 *A posteriori* error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM-Mathematical Modelling and Numerical Analysis (M2AN)* **39**(1), 157–181.

Grepl M, Maday Y, Nguyen N and Patera A 2007 Efficient reduced-basis treatment of non-affine and nonlinear partial differential equations. *Mathematical Modelling and Numerical Analysis (M2AN)* **41**(3), 575–605.

Grimme E 1997 *Krylov Projection Methods for Model Reduction* PhD thesis Coordinated-Science Laboratory, University of Illinois at Urbana-Champaign.

Gugercin S and Antoulas A 2004 A survey of model reduction by balanced truncation and some new results. *International Journal of Control* **77**, 748–766.

Gunzburger M, Peterson J and Shadid J 2007 Reduced-order modeling of time-dependent PDEs with multiple parameters in the boundary data. *Computer Methods in Applied Mechanics and Engineering* **196**, 1030–1047.

Higdon D, Lee H and Holloman C 2003 Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems. *Bayesian Statistics* **7**, 181–197.

Hinze M and Volkwein S 2005 Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control In *Dimension Reduction of Large-Scale Systems* (ed. Benner P, Mehrmann V and Sorensen D), pp. 261–306 Lecture Notes in Computational and Applied Mathematics.

Holmes P, Lumley J and Berkooz G 1996 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, Cambridge, UK.

Kennedy MC and O'Hagan A 2001 Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B* **63**(3), 425–464.

Kunisch K and Volkwein S 1999 Control of Burgers' equation by reduced order approach using proper orthogonal decomposition. *Journal of Optimization Theory and Applications* **102**, 345–371.

Lall S, Marsden J and Glavaski S 2002 A subspace approach to balanced truncation for model reduction of nonlinear control systems. *International Journal on Robust and Nonlinear Control* **12**(5), 519–535.

Le Maître O, Knio O, Najm H and Ghanem R 2001 A stochastic projection method for fluid flow I. Basic formulation. *J. Comput. Phys.* **173**, 481–511.

Le Maître O, Najm H, Ghanem R and Knio O 2004 Multi-resolution analysis of Wiener-type uncertainty propagation schemes. *J. Comput. Phys.* **197**, 502–531.

Le Maître O, Reagan M, Najm H, Ghanem R and Knio O 2002 A stochastic projection method for fluid flow II. Random process. *J. Comput. Phys.* **181**, 9–44.

Li J and White J 2002 Low rank solution of Lyapunov equations. *SIAM Journal on Matrix Analysis and Applications* **24**(1), 260–280.

Li W and Cirpka OA 2006 Efficient geostatistical inverse methods for structured and unstructured grids. *Water Resources Research* **42**, W06402.

Lieberman C 2009 *Parameter and state model reduction for bayesian statistical inverse problems* Master's thesis Massachusetts Institute of Technology.

Ly H and Tran H 2001 Modeling and control of physical processes using proper orthogonal decomposition. *Mathematical and Computer Modeling* **33**(1–3), 223–236.

Ma X and Zabaras N 2009 An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *J. Comp. Phys.* **228**, 3084–3113.

Maday Y, Patera A and Rovas D 2002 A blackbox reduced-basis output bound method for noncoercive linear problems In *Studies in Mathematics and its Applications* (ed. Cioranescu D and Lions J), vol. 31.

Marzouk YM and Najm HN 2009 Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comp. Phys.* **228**(6), 1862–1902.

Marzouk YM and Xiu D 2009 A stochastic collocation approach to Bayesian inference in inverse problems. *Comm. Comput. Phys.* **6**(4), 826–847.

Marzouk YM, Najm HN and Rahn LA 2007 Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comp. Phys.* **224**, 560–586.

Matthies HG and Keese A 2005 Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering* **194**, 1295–1331.

Moore B 1981 Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Transactions on Automatic Control* **AC-26**(1), 17–31.

Najm HN 2009 Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Ann. Rev. Fluid Mech.* **41**, 35–52.

Najm HN, Debusschere BJ, Marzouk YM, Widmer S and Maître OL 2009 Uncertainty quantification in chemical systems. *Int. J. Numer. Meth. Eng.* In press.

Nguyen C 2005 *Reduced basis approximation and a posteriori error bounds for nonaffine and nonlinear partial differential equations: Application to inverse analysis.* PhD thesis Singapore-MIT Alliance.

Nobile F, Tempone R and Webster C 2008 An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Num. Anal.* **46**(5), 2411–2442.

Noor A and Peters J 1980 Reduced basis technique for nonlinear analysis of structures. *AIAA Journal* **18**(4), 455–462.

Penzl T 2006 Algorithms for model reduction of large dynamical systems. *Linear Algebra and its Applications* **415**(2–3), 322–343.

Reagan M, Najm H, Debusschere B, Maître OL, Knio O and Ghanem R 2004 Spectral stochastic uncertainty quantification in chemical systems. *Comb. Theo. Mod.* **8**, 607–632.

Rovas D 2003 *Reduced-basis Output Bound Methods for Parametrized Partial Differential Equations* PhD thesis Massachusetts Institute of Technology.

Rozza G and Veroy K 2006 On the stability of the reduced basis method for Stokes equations in parametrized domains.. *submitted to Elsevier Science*.

Sirovich L 1987 Turbulence and the dynamics of coherent structures. Part 1: Coherent structures. *Quarterly of Applied Mathematics* **45**(3), 561–571.

Smolyak S 1963 Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Math. Dokl.* **4**, 240–243.

Sorensen D and Antoulas A 2002 The Sylvester equation and approximate balanced reduction. *Linear Algebra and its Applications* **351–352**, 671–700.

Sou K, Megretski A and Daniel L 2005 A quasi-convex optimization approach to parameterized model-order reduction *IEEE/ACM Design Automation Conference*, Anaheim, CA.

Veroy K and Patera A 2005 Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: Rigorous reduced-basis *a posteriori* error bounds. *International Journal for Numerical Methods in Fluids* **47**, 773–788.

Veroy K, Prud'homme C, Rovas D and Patera A 2003 A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference*, Orlando, FL. AIAA Paper 2003-3847.

Wan X and Karniadakis GE 2005 An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *J. Comput. Phys.* **209**, 617–642.

Wan X and Karniadakis GE 2009 Error control in multi-element generalized polynomial chaos method for elliptic problems with random coefficients. *Comm. Comput. Phys.* **5**(2–4), 793–820.

Wang J and Zabaras N 2005 Using Bayesian statistics in the estimation of heat source in radiation. *International Journal of Heat and Mass Transfer* **48**, 15–29.

Wiener N 1938 The homogeneous chaos. *Am. J. Math.* **60**, 897–936.

Willcox K and Peraire J 2002 Balanced model reduction via the proper orthogonal decomposition. *AIAA Journal* **40**(11), 2223–2230.

Xiu D 2009 Fast numerical methods for stochastic computations: a review. *Comm. Comput. Phys.* **5**, 242–272.

Xiu D and Hesthaven JS 2005 High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comp.* **27**(3), 1118–1139.

Xiu D and Karniadakis G 2002 The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comp.* **24**(2), 619–644.

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