

A new bi-fidelity model reduction method for Bayesian inverse problems

Na Ou*

Lijian Jiang[†]

Guang Lin[‡]

ABSTRACT

This work presents a new bi-fidelity model reduction approach to the inverse problem under the framework of Bayesian inference. A low-rank approximation is introduced to the solution of the corresponding forward problem and admits a variable separation form in terms of stochastic basis functions and physical basis functions. The calculation of stochastic basis functions is computationally predominant for the low-rank expression. To significantly improve the efficiency of constructing the low-rank approximation, we propose a bi-fidelity model reduction based on a novel variable separation method, where a low-fidelity model is used to compute the stochastic basis functions and a high-fidelity model is used to compute the physical basis functions. The low-fidelity model has lower accuracy but efficient to evaluate compared with the high-fidelity model, it accelerates the derivative of recursive formulation for the stochastic basis functions. The high-fidelity model is computed in parallel for a few samples scattered in the stochastic space when we construct the high-fidelity physical basis functions. The required number of forward model simulations in constructing the basis functions is very limited. The bi-fidelity model can be constructed efficiently while retaining good accuracy simultaneously. In the proposed approach, both the stochastic basis functions and physical basis functions are calculated using the model information. This implies that a few basis functions may accurately represent the model solution in high-dimensional stochastic spaces. The bi-fidelity model reduction is applied to Bayesian inverse problems to accelerate posterior exploration. A few numerical examples in time-fractional derivative diffusion models are carried out to identify the smooth field and channel structured field in porous media in the framework of Bayesian inverse problems.

Keywords: Bayesian inversion, model reduction, multi-fidelity model, fractional equation

1 Introduction

Uncertainties exist inherently in mathematical and physical models, which comes from the models inputs and parameters. The uncertainties can propagate through the model and greatly

*School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha, 410114, China. Email:oyoungla@163.com

[†]School of Mathematical Sciences, Tongji University, Shanghai, 200092, China. Email: ljjiang@tongji.edu.cn

[‡]Department of Mathematics, School of Mechanical Engineering, Department of Statistics (by Courtesy), Purdue University, West Lafayette, IN 47907-2067, USA. Email:guanglin@purdue.edu, Corresponding author.

affect the prediction of the model. To better predict the models outputs, we need to estimate the model inputs and parameters based on some limited measurements [1, 2]. This comes down to solve the inverse problems, which arise frequently in assorted scientific and engineering applications.

The Bayesian approach [3, 4, 5] is one of the methods in solving inverse problems, it incorporates uncertainties in observations and prior information by Bayesian rule and gives the posterior density of the unknowns, which enables us to characterize the parameters and quantify the uncertainties of the system. The appropriate prior density imposed often works as the penalty functional regularization in addressing the ill-posed problems, which may be resulted by the sparse and indirect observation. Although the posterior density can be expressed explicitly by the likelihood function and prior, the nonlinearity of the parameter-to-observation map and lack of analytical form of the forward model make it difficult to utilize the expression straightforwardly. Many numerical approaches such as Markov chain Monte Carlo (MCMC) methods [6, 7] are used to sample the posterior distribution, some efficient MCMC techniques [8, 9, 10, 11, 12] have been developed to deal with the degeneration of the standard MCMC algorithms. The major challenge of applying this method is that hundreds of thousands or even millions of large-scale PDE-based simulations may be required to characterize the posterior distribution, which leads the Bayesian approach to be computationally prohibitive.

One attempt to accelerate Bayesian inference in computationally intensive inverse problems is to construct surrogates [13] of the stochastic forward model. The surrogate may be obtained by the generalized polynomial chaos (gPC)-based stochastic method [14, 15, 16, 17, 18, 19], Gaussian process [20, 21, 22] or projection-type reduced order models [23, 24], etc. We note that the number of polynomial basis functions grows with an exponential rate as the dimension of the unknowns increases, though the number of forward model simulation required is less than the number of polynomial basis functions for sparsity constricted stochastic collocation methods, it increases as increasing the dimension of the unknowns, to pursue the accuracy. The novel variable-separation (NVS) method proposed in [25] provide a variable-separated form approximation of the functions dependent with the stochastic inputs, where the stochastic basis functions can be expressed explicitly, escaping from the limitation of polynomial basis functions. Though recursive formulations can be derived for stochastic functions, it would be extremely time-consuming when the degree of freedom for the model or the number of the separated term is large. The multi-fidelity model [26, 27, 28] technique aims for reducing the computational time. It is widely used in uncertainty propagation, inference, and optimization [29] to improve efficiency. The work proposed in [30, 8] proposes a two-stage delayed acceptance MCMC sampling in solving Bayesian inverse problems, where the candidate sample has to be accepted with the low-fidelity model approximated posterior first before it passes on to be accepted or not with the posterior approximated by the high-fidelity model. The term fidelity can refer to the accuracy of the computer code or simulator compared to the exact function it approximates [31]. Generally, large numbers of the forward model runs are simulated by the low-fidelity models for the sake of speedup, the high-fidelity models are used to guarantee accuracy and/or convergence.

We provide a new bi-fidelity model reduction method in this paper. We run a low-fidelity model to obtain the expression of stochastic basis functions, which is the predominant computational capacity in obtaining the variable separation form of the solution. The iterative method we use requires a limited number of forward model simulations. The low-fidelity model is computationally cheaper to evaluate but not necessarily as accurate as the high-fidelity one. It reduces the

calculation complexity and storage of the residual and matrices involved in obtaining the stochastic recursive formulations. The high-fidelity model can then run in parallel since the corresponding samples have been selected out, the bases in physical space are updated basing on stochastic basis functions. The proposed method can deal with high stochastic dimensional problems efficiently due to the speedup of the low-fidelity model. It provides a global approximation of the states, the observation can be extracted out easily from the form of the bi-fidelity model, i.e. the surrogate model can be invoked conveniently and efficiently. Besides, the credible/prediction intervals of states at time or location that are not observed can also be obtained once posterior samplers provided.

To demonstrate the proposed model reduction method, we apply it to time-dependent stochastic systems, e.g., fractional-differential equation, and study the involved inverse problems. The fractional-differential equation is an effective mathematical model in describing many physical phenomenons, such as fluid mechanics, earthquakes and contaminant transport in aquifers [32], the parameter identification inverse problems have also been intensively studied [33, 34]. We successively obtain the separated representations of the solution when inputs come from the fractional order, source location and the permeability field. The surrogate models are used to approximate the forward model in nonlinear inverse problems, numerical examples show that the performance of the surrogate strongly depends on the number of the separated terms and the low-fidelity model.

In particular, the bi-fidelity model is constructed with a channel structured permeability field, we use the discrete cosine transformation (DCT) [35] to parameterize the field, which can abstract the essential part of data and therefore reduce the dimension of parameters. TV-Gaussian (TG) prior proposed in [36] is imposed as the penalty term here, where TV (total variation) term detects the edges and Gaussian distributions guarantee the well-posedness of the posterior. We take advantage of the efficiency surrogate model construction and use the proposed approach flexibly for the 400-dimensional inverse problem. A computational cheaper low-fidelity model based bi-fidelity model is constructed first, which is used for Bayesian inference, the global structure property of permeability field can be captured roughly with high dimensional unknowns. Informed by the observed data, we sort the posterior variance of parameters and reduce its dimension according to this order. Then we do inference to the new parameter with reduced dimension, the corresponding surrogate bi-fidelity model is built based on a more accurate low-fidelity than the one used before. We find the local structure, e.g., jumps and edges can be recovered better with the new dimension-reduced parameters.

The outline of this paper is as follows. Section 2 introduces the Bayesian inference and the pCN-MCMC algorithm. Section 3 describes the proposed model reduction method, the application of NVS method to time-dependent equation and the new bi-fidelity model reduction method. Some numerical examples are presented in Section 4, the proposed approach is applied to 40 and 400-dimensional inverse problems. Some conclusions and comments are made finally.

2 Bayesian inference for inverse problems

Let (Ω, F, P) be a complete probability space, where Ω is the event space, F the σ -algebra, and P the probability measure. Consider the following stochastic differential equation:

$$\mathcal{D}_t^{\alpha_\xi} u(x, t; \xi(\omega)) - \text{div} \left(\kappa(x; \xi(\omega)) \nabla u(x, t; \xi(\omega)) \right) = f(x, t; \xi(\omega)), \quad x \in \mathcal{O}, \quad t \in [0, T], \quad (2.1)$$

where $f(x, t; \xi(\omega))$ is the source term. $\xi(\omega) : \Omega \rightarrow \mathbb{R}^{n_\xi}$ is the vector of uncertainties, it may be used to describe the coefficient properties, initial/boundary conditions, or source locations. The fractional derivative α_ξ with the subscript ξ represents that it is also part of the uncertainties. $\mathcal{D}_t^{\alpha_\xi}$ may be a *Riemann-Liouville type* or *Caputo type* fractional derivative operator with respect to time, where the *Caputo* time-fractional derivative of the given function $u(x, t)$ with order α_ξ is defined as [37]

$${}^c\mathcal{D}_t^{\alpha_\xi} u = \frac{1}{\Gamma(m - \alpha_\xi)} \int_0^t \frac{\partial^m u(x, s)}{\partial s^m} \frac{ds}{(t - s)^{\alpha_\xi + 1 - m}},$$

and *Riemann-Liouville* time-fractional derivative of $u(x, t)$ is given by

$${}_{RL}\mathcal{D}_t^{\alpha_\xi} u = \frac{1}{\Gamma(m - \alpha_\xi)} \frac{d^m}{dt^m} \int_0^t \frac{ds}{(t - s)^{\alpha_\xi + 1 - m}}.$$

where $\Gamma(\cdot)$ is the Γ function and m is a positive integer satisfying $m - 1 \leq \alpha_\xi < m$. For practical models, we need to quantify the uncertainties of model (2.1) by some observations or measurements.

In the paper, we use Bayesian inference to estimate the unknown inputs, e.g., α_ξ and ξ . We consider the case of additive noise ζ with probability density function $\pi(\zeta)$, the measurement data can then be expressed by

$$d = \mathbf{G}(\xi) + \zeta,$$

where ξ is a vector of the model parameters or inputs and $\mathbf{G}(\xi) = g \circ u(x, t; \xi) \in \mathbb{R}^{n_d}$ is the model response measured, g is the observation operator and n_d is the dimension of the observations. We assume that ζ is independent of ξ , then the conditional probability density for the measurement data d given the unknown ξ , i.e., the likelihood function is given by

$$\pi(d|\xi) = \pi(d - \mathbf{G}(\xi)). \quad (2.2)$$

We use Bayesian inference to solve the inverse problem. This approach gives not only a point estimation but also a probability distribution. This is an advantage of the Bayesian methods over the standard regularization methods. In the framework of Bayesian inference, both ξ and d are random variables. Then the posterior probability density of ξ can be derived by the Bayesian rule,

$$\pi(\xi|d) \propto \pi(d|\xi)\pi(\xi), \quad (2.3)$$

where $\pi(\xi)$ is the prior distribution with available prior information before the data is observed, it can be hybrid, e.g., $\pi(\xi)$ can be the hybrid of Gaussian density and total variation (TV) penalty, which has been proved to be well-posed in [36]. The data is embodied by the likelihood function $\pi(d|\xi)$ in the Bayesian formulation. For the convenience of notation, we will use $\pi^d(\xi)$ to denote the posterior density $\pi(\xi|d)$ and $L(\xi)$ to denote the likelihood function $\pi(d|\xi)$. Then (2.3) can be written as

$$\pi^d(\xi) \propto L(\xi)\pi(\xi). \quad (2.4)$$

The vector ζ is assumed to be independent and identically distributed (i.i.d.) Gaussian random vector with mean zero and standard deviation σ ,

$$\zeta \sim \mathcal{N}(0, \sigma^2 \mathbb{I}),$$

where $\mathbb{I}$ is the identity matrix of size $n_d \times n_d$. Then the likelihood $L(\xi)$ defined in (2.2) is given by

$$L(\xi) = (2\pi\sigma^2)^{-\frac{n_d}{2}} \exp\left(-\frac{\|d - \mathbf{G}(\xi)\|_2^2}{2\sigma^2}\right), \quad (2.5)$$

where $\|\cdot\|_2$ refers to the Euclidean norm. We note that it is not necessary to compute the normalized term in (2.4) under regular circumstance.

We use pCN-MCMC algorithm to explore the posterior distribution, which is listed in Algorithm 1. The credible interval and the prediction interval for the quantities of interests can be calculated based on the posterior samples. The building of the Markov chain requires repeated evaluations of PDE (2.1), which leads to expensive computational cost for large-scale systems. In order to accelerate the MCMC sampling process, we use a new bi-fidelity model reduction method to solve the forward problem. Denote the corresponding reduced forward observation operator as $\mathbf{G}_{\mathbf{N}}$, where $\mathbf{N}$ represents factors that control the accuracy of the surrogate model, e.g. it can be the low-fidelity model and number of separated terms in our paper. We then have the approximated posterior

$$\pi_{\mathbf{N}}^d(\xi) \propto \exp\left(-\frac{\|d - \mathbf{G}_{\mathbf{N}}(\xi)\|_2^2}{2\sigma^2}\right) \pi(\xi).$$

The Kullback-Leibler (KL) divergence [38] is used to measure the difference between the approximated posterior and the reference one. For probability density functions $\pi_{\mathbf{N}}^d(\xi)$ and $\pi^d(\xi)$, KL divergence is defined by

$$D_{KL}(\pi_{\mathbf{N}}^d \parallel \pi^d) = \int \pi_{\mathbf{N}}^d(\xi) \log \frac{\pi_{\mathbf{N}}^d(\xi)}{\pi^d(\xi)} d\xi.$$

D_{KL} measures the difference between the two probability distributions and is non-negative. It vanishes if and only if $\pi_{\mathbf{N}}^d = \pi^d$. Theorem 2.1 gives an estimate of the KL divergence, we will confirm the performance of our proposed method by some numerical examples, and study the effects of the low-fidelity model and the number of separated terms on the convergence of the approximated posterior.

Theorem 2.1. (Theorem 3.1 of [39].) Suppose the functions $\mathbf{G}$ and $\mathbf{G}_{\mathbf{N}}$ are under some assumption, and the observational error has an i.i.d. Gaussian distribution. Then the approximation posterior $\pi_{\mathbf{N}}^d$ and the true posterior density π^d are close with respect to the Kullback-Leibler distance, there is a constant C independent of $\mathbf{N}$, such that

$$D_{KL}(\pi_{\mathbf{N}}^d \parallel \pi^d) \leq \frac{C}{\sigma^4} \|\mathbf{G}(\xi) - \mathbf{G}_{\mathbf{N}}(\xi)\|_{L_{\pi}^2}^2.$$

Remark 2.1. The assumptions functions $\mathbf{G}$ and $\mathbf{G}_{\mathbf{N}}$ satisfied in Theorem 2.1 are: the forward operator $\mathbf{G}$ satisfies $\sup_{\xi} \|\mathbf{G}(\xi)\|_2 < \infty$, which is true in many applications. The surrogate $\mathbf{G}_{\mathbf{N}}$ satisfies $\sup_{\xi} \|\mathbf{G}(\xi) - \mathbf{G}_{\mathbf{N}}(\xi)\|_2 \rightarrow 0$ as $\mathbf{N} \rightarrow \infty$. Since $\mathbf{N}$ represents the low-fidelity model we used in constructing surrogates and the number of separated terms in the variable separation form, $\mathbf{N} \rightarrow \infty$ means the increase of separated terms and the improvement of the accuracy of the low-fidelity model in this paper.

Algorithm 1 pCN-MCMC method

Initial: Generate $\xi^{(0)}$ from $\mathcal{N}(\bar{\xi}, \Sigma)$, set $\beta \in (0, 1)$, chain length N_c .

$qdata = cell(1, N_c)$;

Run the forward model for input $\xi^{(0)}$ and set $\mathcal{Y}^{(0)} = \mathbf{G}(\xi^{(0)})$;

for $i=1:N_c$

pCN proposal, generate ξ from

$$\xi = \sqrt{1 - \beta^2} \xi^{(0)} + (1 - \sqrt{1 - \beta^2}) \bar{\xi} + \beta \varepsilon, \quad \text{with } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

Run the forward model and set $\mathcal{Y} = \mathbf{G}(\xi)$;

Calculate the acceptance probability

$$\alpha(\xi, \xi^{(0)}) = \min \left\{ 1, \frac{L(\xi)}{L(\xi^{(0)})} \right\}.$$

if $\text{rand}(1) < \alpha$

$\xi^{(0)} = \xi$;

$\mathcal{Y}^{(0)} = \mathcal{Y}$;

end if

$qdata(i) = \xi^{(0)}$;

end for

3 Proposed approach

We use the new bi-fidelity model reduction method to construct surrogates for Eq. (2.1), and the solution is approximated as the form

$$u(x, t; \xi) \approx \tilde{u}_N(x, t; \xi) = \sum_{i=1}^N U_i^H(x, t) \eta_i^L(\xi), \quad (3.6)$$

where stochastic basis functions $\{\eta_i^L(\xi)\}_{i=1}^N$ with respect to the stochastic inputs are obtained by low-fidelity models, the physical basis $\{U_i^H(x, t)\}_{i=1}^N$ are calculated by high-fidelity model, the superscript 'L' and 'H' represent the low and high fidelity models, respectively. The application of low-fidelity model makes the relevant matrix during the NVS iterations calculated faster. The high-fidelity basis is updated based on the selected samples, which can be carried on in parallel. The new bi-fidelity model reduction method demands significantly less computational cost and is accuracy preserving simultaneously compared with the original forward model. The surrogate model constructed efficiently is applied to the nonlinear inverse problems involved with Eq. (2.1) in the framework of the Bayesian inference.

3.1 NVS method for time-dependent equation

We focus on the fractional derivative operator $\mathcal{D}_t^{\alpha_\xi}$ with $\alpha_\xi \in (0, 1) \cup (1, 2)$ since the fractional order is separable when we treat the fractional order as part of our stochastic inputs. The

separated form can be derived approximately based on their definitions, i.e., the discretization of the fractional derivative can be written as

$$\mathbf{D}_{t_n}^{\alpha_\xi} u(x, t; \xi) = \sum_{j=1}^n c_j(\xi) u(x, t_j), \quad (3.7)$$

where $0 = t_1 < t_2 \cdots < t_{n_T} = T$, $t_{n_T} = (n_T - 1)\Delta t$, and $n = 1, \dots, n_T$. We denote the dependence of c_j on the fractional order α_ξ as $c_j(\xi)$ for clarity, e.g., when $\mathcal{D}_t^{\alpha_\xi}$ is the *Caputo type* fractional derivative operator, and $\alpha_\xi \in (0, 1)$, c_j 's have the following expression

$$c_j(\xi) := \begin{cases} -\frac{\Delta t^{-\alpha_\xi} [(n-1)^{1-\alpha_\xi} - (n-2)^{1-\alpha_\xi}]}{\Gamma(2-\alpha_\xi)}, & j = 1, \\ -\frac{\Delta t^{-\alpha_\xi} [2(n-j)^{1-\alpha_\xi} - (n-1-j)^{1-\alpha_\xi} - (n+1-j)^{1-\alpha_\xi}]}{\Gamma(2-\alpha_\xi)}, & j = 2, \dots, n-1, \\ \frac{\Delta t^{-\alpha_\xi}}{\Gamma(2-\alpha_\xi)}, & j = n, \end{cases}$$

more details about the expression of $c_j(\xi)$ can be referred in [40] (chapter 2.2).

We will show the application of the NVS method to the time-dependent equation. A weak formulation of Eq. (2.1) can be written as: find u such that

$$(\mathcal{D}_t^{\alpha_\xi} u(\xi), v; \xi) + a(u(\xi), v; \xi) = b(v; \xi), \text{ for } \forall v \in \mathcal{V},$$

where the Hilbert space $\mathcal{V} := H^1(\mathcal{O})$, $a(\cdot, \cdot; \xi)$ and $b(\cdot; \xi)$ are a bilinear form and a linear form on $\mathcal{V}$, respectively. The inner product $(\cdot, \cdot; \xi)$ is defined as

$$(w, v; \xi) = (w(\xi), v(\xi))_{L^2(\mathcal{O})}.$$

We assume the coefficient $\kappa(x; \xi)$ and the source term $f(x, t; \xi)$ have the variable separated form, then the bilinear form $a(\cdot, \cdot; \xi)$ and the associated linear form $b(\cdot; t, \xi)$ are affine with respect to ξ i.e.,

$$\begin{cases} a(w, v; \xi) = \sum_{p=1}^{m_a} \kappa_p(\xi) a_p(w, v), \quad \forall w, v \in \mathcal{V}, \quad \forall \xi \in \Omega, \\ b(v; t, \xi) = \sum_{q=1}^{m_b} f_q(\xi) b_q(v; t), \quad \forall v \in \mathcal{V}, \quad \forall \xi \in \Omega, \end{cases} \quad (3.8)$$

where $\kappa_p(\xi)$ and $f_q(\xi)$ are stochastic functions with respect to ξ , $a_p : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ is a symmetric bilinear form, and $b_q : \mathcal{V} \rightarrow \mathbb{R}$ is continuous functional, they are independent of ξ . If $\kappa(x; \xi)$ and $f(x, t; \xi)$ are not affine with respect to ξ , the Empirical Interpolation Method (EIM)[41, 42] can be used, the NVS method [25] can also be used to obtain their affine forms.

Let $\mathcal{V}_h \subset \mathcal{V}$ be given finite dimensional approximation space. Then the fully discretized approximation for Eq. (2.1) is given by

$$\sum_{j=1}^n c_j(\xi) (u^j(\xi), v; \xi) + a(u^n(\xi), v; \xi) = b^n(v; \xi), \quad \forall v \in \mathcal{V}_h, \text{ for } n = 1, \dots, n_T.$$

where $u^j(\xi) = u(x, t_j; \xi)$, $j = 1, \dots, n$, and $b^n(v; \xi) = b(v; t_n, \xi)$. We seek for the numerical solution of Eq. (2.1) under the form (3.6) where $U_i(x, t) \in \mathcal{V}_h$ are deterministic basis functions and $\eta_i(\xi)$ are stochastic functions, for $i = 1, \dots, N$. Let the residual for the NVS procedure be

$$e^j(\xi) := u^j(\xi) - u_{k-1}^j(\xi), \text{ for } j = 1, \dots, n, \text{ and } n = 1, \dots, n_T.$$

we then have for $\forall v \in \mathcal{V}_h$,

$$\sum_{j=1}^n c_j(\xi)(e^j(\xi), v; \xi) + a(e^n(\xi), v; \xi) = b^n(v; \xi) - \sum_{j=1}^n c_j(\xi)(u_{k-1}^j(\xi), v; \xi) - a(u_{k-1}^n(\xi), v; \xi).$$

Let $r^n(v; \xi) \in \mathcal{V}_h^*$ (the dual space of $\mathcal{V}_h$) be the residual

$$r^n(v; \xi) = \begin{cases} b^n(v; \xi), & k = 1, \\ b^n(v; \xi) - \sum_{j=1}^n c_j(\xi)(u_{k-1}^j(\xi), v; \xi) - a(u_{k-1}^n(\xi), v; \xi), & k \geq 2, \end{cases} \quad (3.9)$$

we then have

$$\sum_{j=1}^n c_j(\xi)(e^j(\xi), v; \xi) + a(e^n(\xi), v; \xi) = r^n(v; \xi), \quad \forall v \in \mathcal{V}_h. \quad (3.10)$$

By Riesz representation theory, there exists $\hat{e}^n(\xi) \in \mathcal{V}_h$ such that

$$(\hat{e}^n(\xi), v)_{\mathcal{V}} = r^n(v; \xi), \quad \forall v \in \mathcal{V}_h. \quad (3.11)$$

Then we can rewrite the error residual Eq. (3.10) as

$$\sum_{j=1}^n c_j(\xi)(e^j(\xi), v; \xi) + a(e^n(\xi), v; \xi) = (\hat{e}^n(\xi), v)_{\mathcal{V}}, \quad \forall v \in \mathcal{V}_h.$$

Consequently, the dual norm of the residual $r^n(v; \xi)$ can be evaluated through the Riesz representation

$$\|r^n(v; \xi)\|_{\mathcal{V}_h^*} := \sup_{v \in \mathcal{V}_h} \frac{r^n(v; \xi)}{\|v\|_{\mathcal{V}}} = \|\hat{e}^n(\xi)\|_{\mathcal{V}}, \quad (3.12)$$

the computation of the residual is crucial to the novel NVS procedure.

We define the error estimator for the solution by

$$\Delta_k(\xi) := \sum_{n=1}^{n_T} \|\hat{e}^n(\xi)\|_{\mathcal{V}} \Delta t, \quad (3.13)$$

we note that $\hat{e}^n(\xi) \in \mathcal{V}$ is related to $r^n(v; \xi)$ through Eq. (3.12), for $n = 1, \dots, n_T$.

By (3.8) and (3.9), the residual can be expressed by

$$r^n(v; \xi) = \sum_{q=1}^{m_b} f_q(\xi) b_q(v; t_n) - \sum_{j=1}^n c_j(\xi) \sum_{i=1}^{k-1} \eta_i(\xi)(U_i^j, v) - \sum_{i=1}^{k-1} \eta_i(\xi) \sum_{p=1}^{m_a} \kappa_p(\xi) a_p(U_i^n, v). \quad (3.14)$$

where $U_i^j = U_i(x, t_j)$, $j = 1, \dots, n_T$. By (3.14) and (3.11), we have

$$(\hat{e}^n(\xi), v)_{\mathcal{V}} = \sum_{q=1}^{m_b} f_q(\xi) b_q(v; t_n) - \sum_{j=1}^n c_j(\xi) \sum_{i=1}^{k-1} \eta_i(\xi)(U_i^j, v) - \sum_{i=1}^{k-1} \eta_i(\xi) \sum_{p=1}^{m_a} \kappa_p(\xi) a_p(U_i^n, v).$$

This implies that

$$\hat{e}^n(\xi) = \sum_{q=1}^{m_b} f_q(\xi) \mathcal{C}_q^n + \sum_{j=1}^n c_j(\xi) \sum_{i=1}^{k-1} \eta_i(\xi) \mathcal{J}_i^j + \sum_{i=1}^{k-1} \eta_i(\xi) \sum_{p=1}^{m_a} \kappa_p(\xi) \mathcal{L}_i^{p,n}, \quad (3.15)$$

where $\mathcal{C}_q^n$ is the Riesz representation of $b_q(v, t_n)$, i.e., $(\mathcal{C}_q^n, v)_{\mathcal{V}} = b_q(v, t_n)$ for any $v \in \mathcal{V}$, $1 \leq q \leq m_b$, $1 \leq n \leq n_T$. Similarly, $\mathcal{J}_i^j$ is the Riesz representation of (U_i^j, v) , i.e., $(\mathcal{J}_i^j, v)_{\mathcal{V}} = -(U_i^j, v)$ for any $v \in \mathcal{V}$, where $1 \leq i \leq k-1$, $1 \leq j \leq n$, $1 \leq n \leq n_T$. $\mathcal{L}_i^{p,n}$ is the Riesz representation of $a_p(U_i^n, v)$, i.e., $(\mathcal{L}_i^{p,n}, v)_{\mathcal{V}} = -a_p(U_i^n, v)$ for any $v \in \mathcal{V}$ and $1 \leq p \leq m_a$, $1 \leq i \leq k-1$, $1 \leq n \leq n_T$. The Eq. (3.15) gives rise to

$$\begin{aligned} \|\hat{e}^n(\xi)\|_{\mathcal{V}}^2 &= \sum_{q=1}^{m_b} \sum_{q'=1}^{m_b} f_q(\xi) f_{q'}(\xi) (\mathcal{C}_q^n, \mathcal{C}_{q'}^n)_{\mathcal{V}} + \sum_{j=1}^n c_j(\xi) \sum_{i=1}^{k-1} \eta_i(\xi) \\ &\quad \times \left\{ 2 \sum_{q=1}^{m_b} f_q(\xi) (\mathcal{C}_q^n, \mathcal{J}_i^j)_{\mathcal{V}} + \sum_{j'=1}^n c_{j'}(\xi) \sum_{i'=1}^{k-1} \eta_{i'}(\xi) (\mathcal{J}_i^j, \mathcal{J}_{i'}^{j'})_{\mathcal{V}} \right\} \\ &\quad + \sum_{i=1}^{k-1} \eta_i(\xi) \sum_{p=1}^{m_a} \kappa_p(\xi) \times \left\{ 2 \sum_{q=1}^{m_b} f_q(\xi) (\mathcal{C}_q^n, \mathcal{L}_i^{p,n})_{\mathcal{V}} \right. \\ &\quad \left. + 2 \sum_{j'=1}^n c_{j'}(\xi) \sum_{i'=1}^{k-1} \eta_{i'}(\xi) (\mathcal{J}_{i'}^{j'}, \mathcal{L}_i^{p,n})_{\mathcal{V}} + \sum_{i'=1}^{k-1} \eta_{i'}(\xi) \sum_{p'=1}^{m_a} \kappa_{p'}(\xi) (\mathcal{L}_i^{p,n}, \mathcal{L}_{i'}^{p',n})_{\mathcal{V}} \right\}. \end{aligned} \quad (3.16)$$

$\|\hat{e}^n(\xi)\|_{\mathcal{V}}$ and $\Delta_k(\xi)$ can be efficiently computed when offline-online procedure is applied. In offline stage, we compute and storage the quantities independent with stochastic inputs. We store $(\mathcal{C}_q^n, \mathcal{C}_{q'}^n)_{\mathcal{V}}$, $(\mathcal{C}_q^n, \mathcal{J}_i^n)_{\mathcal{V}}$, $(\mathcal{J}_i^n, \mathcal{J}_{i'}^n)_{\mathcal{V}}$, $(\mathcal{C}_q^n, \mathcal{L}_i^{p,n})_{\mathcal{V}}$, $(\mathcal{J}_i^n, \mathcal{L}_i^{p,n})_{\mathcal{V}}$ and $(\mathcal{L}_i^{p,n}, \mathcal{L}_{i'}^{p',n})_{\mathcal{V}}$ for online stage, where $1 \leq i, i' \leq k-1$, $1 \leq q, q' \leq m_b$, $1 \leq p, p' \leq m_a$, $1 \leq n \leq n_T$. In the online stage, we evaluate $\|\hat{e}^n(\xi)\|_{\mathcal{V}}$ and hence $\Delta_k(\xi)$ for any ξ using Eqs. (3.16) and (3.13).

Let Ξ be a small sample set taken from $\xi(\Omega)$. At step k , we choose

$$\xi^{(k)} := \begin{cases} \text{randomly chosen in } \Xi, & k = 1, \\ \arg \max_{\xi \in \Xi} \Delta_k(\xi), & k \geq 2. \end{cases}$$

Let $e(\xi) = e_h(x, t) e_{\xi}(\xi)$, and $e_h(x, t)$ be the solution of Eq. (3.10) with $\xi = \xi^{(k)}$, we take $U_k(x, t) = e_h(x, t)$ in (3.6), then for $n = 1, \dots, n_T$,

$$e_{\xi}(\xi) \left[\sum_{j=1}^n c_j(\xi) (e_h^j, v) + \sum_{p=1}^{m_a} \kappa_p(\xi) a_p(e_h^n, v) \right] = r^n(v; \xi), \quad (3.17)$$

where $e_h^j = e_h(x, t_j)$, $j = 1, \dots, n$. We take $v = e_h^J$ in Eq. (3.17), where

$$J = \arg \max_{n=1, \dots, n_T} r^n(e_h^n; \xi^{(k)}),$$

then it follows that

$$e_{\xi}(\xi) = \frac{b^J(e_h^J; \xi) - \sum_{i=1}^{k-1} \eta_i(\xi) R_i(\xi)}{R_k(\xi)}, \quad (3.18)$$

where

$$b^J(e_h^J; \xi) = \sum_{q=1}^{m_b} f_q(\xi) b_q(e_h^J, t_J),$$

$$R_i(\xi) = \sum_{j=1}^J c_j(\xi)(U_i^j, e_h^J) + \sum_{p=1}^{m_a} \kappa_p(\xi) a_p(U_i^J, e_h^J), \quad i = 1, \dots, k,$$

we take $\eta_k(\xi) = e_\xi(\xi)$ in Eq. (3.6).

Proposition 3.1. *Let the selected samples be $\{\xi^{(k)}\}_{k=1}^N$ during the NVS procedure, then for $k = 1, \dots, N$.*

$$\eta_k(\xi^{(j)}) := \begin{cases} 0, & j < k, \\ 1, & j = k. \end{cases}$$

This proposition can be proved according to expressions of $\{\eta_i(\xi)\}_{i=1}^N$ and $\{U_i(x, t)\}_{i=1}^N$.

As can be seen from Eq. (3.18), $\eta_k(\xi)$ depends on $\{\eta_i(\xi)\}_{i=1}^{k-1}$ computed previously while they are stochastic functions with respect to ξ . Let $\mathbf{b}^k := [b_1(e_h^J, t_J), \dots, b_{m_b}(e_h^J, t_J)]$, $\mathbf{f}(\xi) := [f_1(\xi), \dots, f_{m_b}(\xi)]^T$, $\mathbf{c}^k(\xi) := [c_1(\xi), \dots, c_J(\xi)]^T$, $\kappa_\xi(\xi) := [\kappa_1(\xi), \dots, \kappa_{m_a}(\xi)]^T$, and $\boldsymbol{\eta}^k(\xi) := [\eta_1(\xi), \dots, \eta_{k-1}(\xi)]$,

$$\mathbf{Z}^k := \begin{bmatrix} (U_1^1, U_k^J) & \dots & (U_1^J, U_k^J) \\ \vdots & \ddots & \vdots \\ (U_k^1, U_k^J) & \dots & (U_k^J, U_k^J) \end{bmatrix}. \quad (3.19)$$

$$\mathbf{A}^k := \begin{bmatrix} a_1(U_1^J, U_k^J) & \dots & a_{m_a}(U_1^J, U_k^J) \\ \vdots & \ddots & \vdots \\ a_1(U_k^J, U_k^J) & \dots & a_{m_a}(U_k^J, U_k^J) \end{bmatrix}. \quad (3.20)$$

Then the matrix form of $\eta_k(\xi)$ can be written as

$$\eta_k(\xi) = \frac{\mathbf{b}^k \mathbf{f}(\xi) - \boldsymbol{\eta}^k(\xi) \left[\mathbf{Z}_1^k \mathbf{c}^k(\xi) + \mathbf{A}_1^k \kappa_\xi(\xi) \right]}{\mathbf{z}^k \mathbf{c}^k(\xi) + \mathbf{a}^k \kappa_\xi(\xi)}.$$

where $\mathbf{Z}_1^k, \mathbf{A}_1^k$ are the first $k-1$ rows of matrix $\mathbf{Z}^k, \mathbf{A}^k$, respectively. $\mathbf{z}^k$ and $\mathbf{a}^k$ are the last row of matrix $\mathbf{Z}^k$ and $\mathbf{A}^k$. All the inner product in physical space can be computed and saved at each step, the simulation for ξ is efficient due to the affine form. Algorithm 2 presents the pseudo-code for the NVS procedure to approximate the solution to Eq. (2.1).

Algorithm 2 The NVS method for time-dependent equation

Input: The fractional system, training sample set Ξ taken from the prior, and the error tolerance Δ_ε

Output: The separated representation $u_N(x, t; \xi) = \sum_{i=1}^N U_i(x, t) \eta_i(\xi)$

$k = 1;$

$\xi = \xi^{(k)};$

$r^n(v; \xi) = b^n(v; \xi), \text{ for } n = 1, \dots, n_T;$

while $k < \infty$

Calculate $U_k(x, t) = e_h(x, t)$ by solving (3.10) with ξ ;

Take $v = e_h^J$, where $J = \arg \max_{n=1, \dots, n_T} r^n(e_h^n, \xi^{(k)});$

Calculate and storage inner product in physical space in matrix (3.19) and (3.20);

if $\Delta_k(\xi) < \Delta_\varepsilon$

$N = k;$

break;

end if

Update Ξ with $\Xi \setminus \xi^{(k)};$

Take $\xi^{(k)} = \arg \max_{\xi \in \Xi} \Delta_k(\xi);$

Set $\xi = \xi^{(k)};$

Update the residual $r^n(v; \xi) = b^n(v; \xi) - \sum_{i=1}^k U_i(x, t_n) \eta_i(\xi), \text{ for } n = 1, \dots, n_T;$

$k = k + 1;$

end while

3.2 High-fidelity approximation via basis update

Though the offline-online procedure can improve the efficiency of calculating Eq.(3.16), inner products in physical space should be calculated and stored at each iteration. For example, when we finish the $k - 1$ th iteration and obtain the first $k - 1$ physical and stochastic basis functions, we need to calculate the involved products between $U_{k-1}(x, t)$ and $\{U_i(x, t)\}_{i=1}^{k-1}$, of which the computational cost depends on the degree of freedom when solving the PDEs. For large-scale systems, the computational time for the offline stage may be significant. For cases that the number of the separated terms N required are large for model accuracy, iterations of the NVS method would be time-consuming. i.e., these lead much burden to the calculation of the residual r , the computation of matrix $\mathbf{Z}^k$ and $\mathbf{A}^k$, for $k = 1, \dots, N$. Following [28, 43], we construct a bi-fidelity model, of which stochastic basis functions are built from low-fidelity model, but physical basis is calculated by the high-fidelity model.

We rearrange the NVS method for the time-dependent equation (2.1) before presenting our proposed method. Let $\mathcal{A}_k := (\mathcal{D}_t^{\alpha_{\xi^{(k)}}} \cdot, v; \xi^{(k)}) - a(\cdot, v; \xi^{(k)})$, where v is from the space $\mathcal{V}$. $\{\xi^{(k)}\}_{k=1}^N$

are samples selected during the iteration of NVS, then we have

$$\begin{aligned}\mathcal{A}_1 U_1 &= (f, v; \xi^{(1)}) \\ \mathcal{A}_2(U_2 + U_1 \eta_1(\xi^{(2)})) &= (f, v; \xi^{(2)}) \\ &\vdots \\ \mathcal{A}_N(U_N + \sum_{i=1}^{N-1} U_i \eta_i(\xi^{(N)})) &= (f, v; \xi^{(N)}).\end{aligned}$$

The equations above can be derived from the procedure of the NVS method, due to the property concluded in Proposition 3.1, they can be written uniformly as

$$\mathcal{A}_k(\sum_{i=1}^k U_i \eta_i(\xi^{(k)})) = (f, v; \xi^{(k)}),$$

for $k = 1, 2, \dots, N$, if we denote $\mathring{U}_k(x, t; \xi^{(k)})$ as the solution of the full model

$$\mathcal{A}_k \mathring{U}_k = (f, v; \xi^{(k)}), \quad (3.21)$$

then the relationship between $\{U_i(x, t)\}_{i=1}^N$ and $\{\mathring{U}_i(x, t; \xi^{(i)})\}_{i=1}^N$ is

$$\sum_{i=1}^k U_i \eta_i(\xi^{(k)}) = \mathring{U}_k(x, t; \xi^{(k)}). \quad (3.22)$$

Denote the approximate solution obtained from low-fidelity model as

$$u_N^L(x, t; \xi) = \sum_{i=1}^N U_i^L(x, t) \eta_i^L(\xi),$$

where $\{\eta_i^L(\xi)\}_{i=1}^N$ is stochastic basis functions learned from the low-fidelity model, the relevant matrix $\mathbf{Z}^k$ and $\mathbf{A}^k$ are computed by vectors with relative small size, which is corresponding to the low-fidelity model. We then replace the basis vectors $\{U_i^L(x, t)\}_{i=1}^N$ with $\{U_i^H(x, t)\}_{i=1}^N$, which are obtained from the high-fidelity model corresponding to the same selected samples $\{\xi_L^{(k)}\}_{k=1}^N$, i.e., we derive the high-fidelity basis vectors by equations

$$\sum_{i=1}^k U_i^H \eta_i^L(\xi_L^{(k)}) = \mathring{U}_k^H(x, t; \xi_L^{(k)}). \quad \text{for } k = 1, \dots, N, \quad (3.23)$$

where $\{\mathring{U}_i^H(x, t; \xi^{(i)})\}_{i=1}^N$ are full solutions of equations (3.21) with the corresponding high-fidelity model, they can be solved in parallel since samples $\{\xi_L^{(k)}\}_{k=1}^N$ has been selected out by the low-fidelity model. As $\{\mathring{U}_i^H\}_{i=1}^N$ and stochastic basis functions $\{\eta_i^L(\xi)\}_{i=1}^N$ have been obtained before, $\{U_i^H(x, t)\}_{i=1}^N$ can be easily calculated by the recursive formulation (3.23). The bi-fidelity model can then be established as (3.6).

Hence, compared with the full high-fidelity model, the bi-fidelity model uses the low-fidelity model to construct stochastic basis functions during the NVS iteration procedure, which can improve the efficiency of calculating and storage of residual r . Besides, the offline stage in computing

$\|\hat{e}^n(\xi)\|_{\mathcal{V}}$ and $\Delta_k(\xi)$ can also be accelerated due to the small size of the low-fidelity model, i.e. the efficiency of selecting samples has been improved. The implementation of $\{\eta_i^L(\xi)\}_{i=1}^N$ and linearity of the time-dependent equation makes the update of the high-fidelity basis faster since parallelization can be carried on. Algorithm 3 describes the procedure for construction of the bi-fidelity surrogate model.

Algorithm 3 A new bi-fidelity model reduction method.

Input: The time-dependent system

Output: The bi-fidelity model $\tilde{u}_N(x, t; \xi) = \sum_{i=1}^N U_i^H(x, t) \eta_i^L(\xi)$

Run Algorithm 2 with low-fidelity time-dependent model, obtain stochastic basis functions $\{\eta_i^L(\xi)\}_{i=1}^N$ and the selected samples $\{\xi_L^{(k)}\}_{k=1}^N$;

Run high-fidelity time-dependent model with $\{\xi_L^{(k)}\}_{k=1}^N$ in parallel and obtain $\{\mathring{U}_i^H\}_{i=1}^N$ according to Equation (3.21);

Calculate the high-fidelity basis $\{U_i^H(x, t)\}_{i=1}^N$ according to Equation (3.23);

Update the low-fidelity model with the high-fidelity basis and obtain the new bi-fidelity model.

For the choice of the low-fidelity model used in constructing the bi-fidelity model, we summarize some helpful conclusions from paper [43]. Let $\{\xi^i\}_{i=1}^n$ be samples i.i.d taken from the prior, we denote the corresponding realizations of the full high-fidelity and bi-fidelity models at the same input samples as $\mathring{\mathbf{U}}^H, \tilde{\mathbf{U}}_N \in \mathbb{R}^{N_H \times n}$, where N_H is the degree of freedom of the high-fidelity model. The corresponding realizations of the full low-fidelity and NVS approximation for low-fidelity models at the same input samples are denoted as $\mathring{\mathbf{U}}^L, \mathbf{U}_N^L \in \mathbb{R}^{N_L \times n}$, respectively. N_L is the degree of freedom of the low-fidelity model. $\mathbf{C}_L \in \mathbb{R}^{N \times n}$ is the matrix of which the columns are realizations of η^L , the stochastic functions learned from the low-fidelity model. $\mathbf{C}_L^S \in \mathbb{R}^{N \times N}$ represents the realizations of η^L at the selected points $\{\xi_L^{(k)}\}_{k=1}^N$, with the corresponding order. According to Proposition 3.1, we have that $\mathbf{C}_L^S$ is an upper triangular matrix whose diagonal elements are 1. We note $\{\xi_L^{(k)}\}_{k=1}^N$ is part of $\{\xi^i\}_{i=1}^n$, i.e., columns of $\mathbf{C}_L^S$ are part of $\mathbf{C}_L$. Then as sought by the author in [43], the condition $\mathring{\mathbf{U}}^L$ (in relation to $\mathring{\mathbf{U}}^H$) should satisfy is:

Theorem 3.1. For a finite $\tau \geq 0$, define

$$\varepsilon(\tau) := \lambda_{\max} \left(\mathbb{T}(\mathring{\mathbf{U}}^H) \mathring{\mathbf{U}}^H - \tau \mathbb{T}(\mathring{\mathbf{U}}^L) \mathring{\mathbf{U}}^L \right),$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix, $\mathbb{T}(\cdot)$ is the transformation of matrix. If $\varepsilon(\tau)$ is small enough then there exists a matrix $\mathbf{T}$ with bounded $\|\mathbf{T}\|$ such that for a matrix $\mathbf{E}$ with small $\|\mathbf{E}\|$, the error in the approximation can be represented as

$$\mathring{\mathbf{U}}^H = \mathbf{T} \mathring{\mathbf{U}}^L + \mathbf{E}, \quad (3.24)$$

and

$$\|\mathring{\mathbf{U}}^H - \tilde{\mathbf{U}}_N\| \leq (1 + \|(\mathbf{C}_L^S)^{-1}\| \|\mathbf{C}_L\|) \|\mathbf{E}\| + \|\mathbf{T}\| \|\mathring{\mathbf{U}}^L - \mathbf{U}_N^L\|. \quad (3.25)$$

where $\|\cdot\|$ denotes the matrix induced l_2 norm.

The proof of the inequation is provided in the Appendix. $\|\mathring{\mathbf{U}}^L - \mathbf{U}_N^L\|$ is small given that $\mathring{\mathbf{U}}^L$ is the realization of the full low-fidelity models, when $\|(\mathbf{C}_L^S)^{-1}\|\|\mathbf{C}_L\|$ is bounded, (3.25) suggests an accurate bi-fidelity error estimate. We will demonstrate the boundedness of $\|(\mathbf{C}_L^S)^{-1}\|\|\mathbf{C}_L\|$ by numerical experiments in our paper. As the approximation error can be expressed by Eq.(3.25), the accuracy of the proposed method is effected by the accuracy of the variable separation approximation of the low-fidelity model, i.e. the term $\|\mathring{\mathbf{U}}^L - \mathbf{U}_N^L\|$. One can still use the error tolerance Δ_ε to determine the number of separated terms N when constructing stochastic functions $\{\eta_i^L(\xi)\}_{i=1}^N$, hence the bi-fidelity model.

Remark 3.1. *We note that the bi-fidelity technology can also be used in separating variables for random fields with explicit expression, e.g., when the coefficient or the source term is not affine, using NVS method to get the affine form, the bi-fidelity technology can greatly improve the efficiency of NVS procedure.*

4 Numerical examples

In this section, we consider the *Caputo* time-fractional partial differential equations with fractional order $\alpha_\xi \in (0, 1)$

$${}^c D_t^{\alpha_\xi} u - \text{div}(\kappa(x)\nabla u) = f(x, t), \quad x \in \Omega, t \in (0, T], \quad (4.26)$$

to present the performance of the proposed method. The equation is subjected to an appropriate boundary condition and initial condition. Here $\kappa(x)$ is a spatial varied diffusion coefficient, $f(x, t)$ is a source (or sink) term. For practical models, the model inputs such as $\kappa(x)$ and α_ξ may be unknown, in order to make better predictions, we need to estimate them by some measurements. We focus on the inverse problems for the fractional diffusion model (4.26), the unknown order, coefficient and the source location are estimated under the framework of Bayesian inference. The space domain considered here is $\mathcal{O} := [0, 1] \times [0, 1]$, and the initial condition is set as $u(x, 0) = 0$. Measurement data are generated by the using finite element method in a fine grid with time step $\Delta t = 0.001$, and the measurement noise is set to be $\sigma = 0.01$.

4.1 Identify the source location and fractional derivative α_ξ

Consider the case when there is no flow on the boundaries, the coefficient is constant and $\kappa(x) = 1$. For the first example, the fractional derivative is given by $\alpha_\xi = 0.5$, and the end time is set as $T = 0.15$. Assume the source term has the form

$$f(x, t; \xi) = 10 \exp \left(- \frac{(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2}{2 \times 0.1^2} \right) [1 - h(t - T_1)],$$

where (ξ_1, ξ_2) denotes the location of the one point source, h is the heaviside function, here T_1 is prescribed as 0.05. We need to identify the source location (ξ_1, ξ_2) , and we assume the uniform distribution as its prior density, i.e., $\xi_i \sim \mathcal{U}(0, 1), i = 1, 2$.

Firstly, we run a low-fidelity model with grid size 20×20 to obtain stochastic basis functions $\{\eta_i^L(\xi)\}_{i=1}^N$ and the corresponding samples $\{\xi_L^{(k)}\}_{k=1}^N$. Since the degree of freedom of the low-fidelity model is small, we don't apply the NVS method to approximate the stochastic source term

Strategies	N	Construction		Online per sample	Re
		low time	high time		
Bi-model	20	$1.0406 \times 10^1 s$	$3.8163 \times 10^1 s$	$1.7196 \times 10^{-4} s$	2.3173×10^{-2}
	30	$1.1397 \times 10^1 s$	$5.5039 \times 10^1 s$	$1.9342 \times 10^{-4} s$	9.4496×10^{-3}
	40	$1.2849 \times 10^1 s$	$7.2316 \times 10^1 s$	$1.9442 \times 10^{-4} s$	5.2028×10^{-3}
High-vs-model	20	—	$1.8213 \times 10^2 s$	$4.2435 \times 10^{-3} s$	2.2508×10^{-2}
	30	—	$2.8694 \times 10^2 s$	$4.4362 \times 10^{-3} s$	9.1729×10^{-3}
	40	—	$4.5455 \times 10^2 s$	$4.7051 \times 10^{-3} s$	4.3759×10^{-3}

Table 1: Comparison of the construction and online calling CPU time, relative mean error Re for different model reduced methods, with different numbers of separated terms.

in separated form, but use the original expression and calculate the approximated solution for any random realization by some low-dimensional products. Here the time step is set as $\Delta t = 0.002$, and there are 300 samples randomly draw from the prior in the training sample set Ξ . For each selected samples, we then solve the forward models with grid size 100×100 and finer time step $\Delta t = 0.001$, which can be proceeded in parallel. The high-fidelity model bases are calculated according to Eq. (3.23). By updating the low-fidelity surrogate model with the high-fidelity bases, The bi-fidelity model can be established.

In Table 1, we compare the construction and online calling CPU time, relative mean error Re for different model reduced methods, with different numbers of separated terms. The Bi-model refers to the bi-fidelity model and the High-vs-model refers to the variable-separation approximation obtained by the high-fidelity model (with grid size 100×100 and time step 0.001) only. The low-time represents the time cost of the low-fidelity model, the high time is the corresponding time cost of the high-fidelity model, high-fidelity vectors are calculated in parallel with 2 cores for the Bi-model. The relative mean error is defined as

$$Re = \sqrt{\frac{\int_{\Omega} \int_{[0,T]} \int_{\mathcal{O}} (\hat{u}^H(x,t;\xi) - u_r(x,t;\xi))^2 dx dt d\xi}{\int_{\Omega} \int_{[0,T]} \int_{\mathcal{O}} \hat{u}^H(x,t;\xi)^2 dx dt d\xi}},$$

where $\hat{u}^H(x,t;\xi)$ represents the full high-fidelity model solution and $u_r(x,t;\xi)$ is obtained by the reduced models, i.e. the Bi-model or High-vs-model. For the same number of separated terms, the Bi-model can achieve almost the same relative mean error level as the High-vs-model, while it costs less time in the reduced model construction stage than the High-vs-model as expected. Since the force term is not affine, the Bi-model is more efficient in the online stage than the High-vs-model.

We study the effect of the low-fidelity model and the number of the separated terms on the approximation of the bi-fidelity model. The errors achieved by the low-fidelity model alone $\|\hat{\mathbf{U}}^L - \mathbf{U}_N^L\|$ and the one achieved by the bi-fidelity model $\|\hat{\mathbf{U}}^H - \tilde{\mathbf{U}}_N\|$ against the number of the separated terms are plotted in Figure 1, where columns of the involved matrices are calculated with respect to the training sample set Ξ , i.e., $n = 300$. Stochastic basis functions are constructed by 20×20 and 40×40 low-fidelity models, respectively. All errors decrease as we increase the number of separated terms. Though for the errors $\|\hat{\mathbf{U}}^L - \mathbf{U}_N^L\|$ low-fidelity models achieved alone, the 20×20

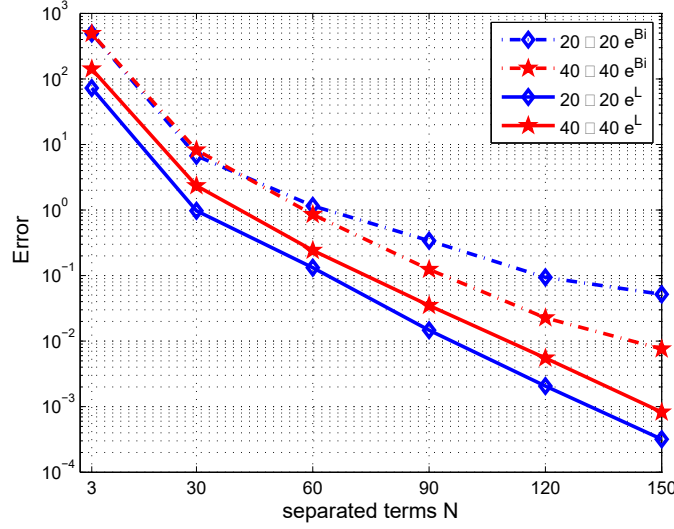


Figure 1: Approximation of the model with respect to the number of the separated terms by different strategies: bi-fidelity model and low-fidelity model. Dashed line represents the bi-fidelity model approximation error $\|\hat{\mathbf{U}}^H - \tilde{\mathbf{U}}_N\|$ (denoted by e^{Bi}); solid line shows the error $\|\hat{\mathbf{U}}^L - \mathbf{U}_N^L\|$ (denoted by e^L).

model leads to error level lower than the 40×40 one at the same number of the separated terms, top curves show that the bi-fidelity model constructed with the 40×40 low-fidelity model is more accurate than the one constructed by the 20×20 low-fidelity model in general.

As can be seen from Figure 2, the values of $\|(\mathbf{C}_L^S)^{-1}\| \|\mathbf{C}_L\|$ are bounded for the two different low-fidelity models. The values of $\|\mathbf{C}_L\|$ are stable. The trend of $\|(\mathbf{C}_L^S)^{-1}\| \|\mathbf{C}_L\|$ are almost the same as the trend of $\|(\mathbf{C}_L^S)^{-1}\|$. For the 40×40 low-fidelity model, the terms $\|(\mathbf{C}_L^S)^{-1}\| \|\mathbf{C}_L\|$ and $\|\hat{\mathbf{U}}^L - \mathbf{U}_N^L\|$ are larger than the ones estimated by the 20×20 low-fidelity model mostly, while it reaches better bi-fidelity model approximation. This is resulted by the smaller $\|\mathbf{E}\|$ and $\|\mathbf{T}\|$, which are determined by the relationship between the low-fidelity model and the high-fidelity one. It verifies numerically that the bi-fidelity model approximation is effected by the low-fidelity variable-separation approximation and the low-fidelity model used in constructing the bi-fidelity model, which coincides with the result of Theorem 3.1.

The ground true parameter values is set as $\xi = (0.25, 0.75)$ in the example. Observation data are generated by adding independent random noise $\mathcal{N}(0, \sigma^2)$ to the solution at a uniform 6×6 sensor network. At each sensor location, measurements are taken at time $t = 0.05$ and 0.1 , which correspond to a total of 72 measurements. Figure 3 shows the difference between the approximate and the reference posterior distribution with different separated terms $N = 5, 20, 50$ and 75 , where the low-fidelity model used is 20×20 , the solid and dashed lines denote the reference and the surrogate posterior, respectively. The better agreement between $\pi_N^d(\xi)$ and $\pi^d(\xi)$ is observed by increasing the separated terms, the approximate posterior is almost the same as the reference on when the surrogate bi-fidelity model is constructed with 75 separated terms.

The bi-fidelity model is used repeatedly to obtain samplers via pCN-MCMC algorithm, in which parameter β is set as 0.01 and the length of the Markov chain is 3×10^4 , only the last

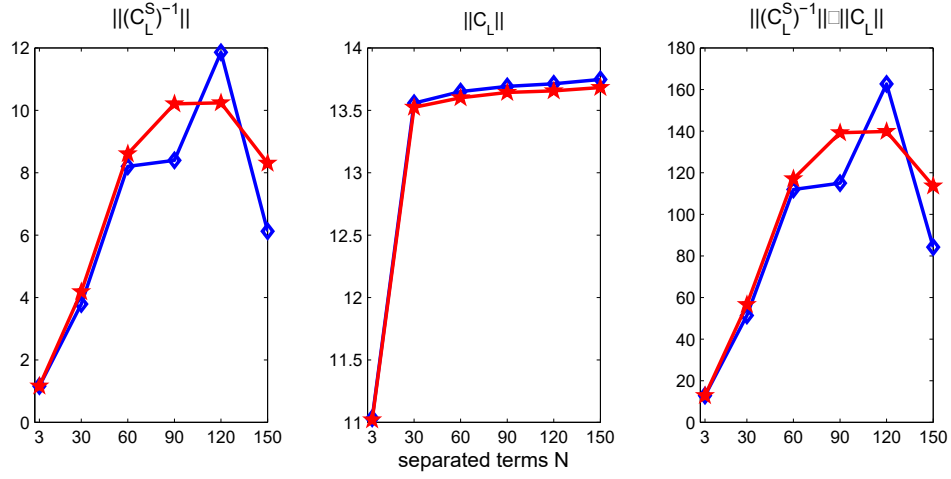


Figure 2: Values of $\|(C_L^S)^{-1}\|$, $\|C_L\|$ and $\|(C_L^S)^{-1}\| \|C_L\|$ with respect to the number of the separated terms: line with marker 'diamond' represents the values obtained by the 20×20 low-fidelity model; line with marker 'pentagram' represents the values obtained by the 40×40 low-fidelity model.

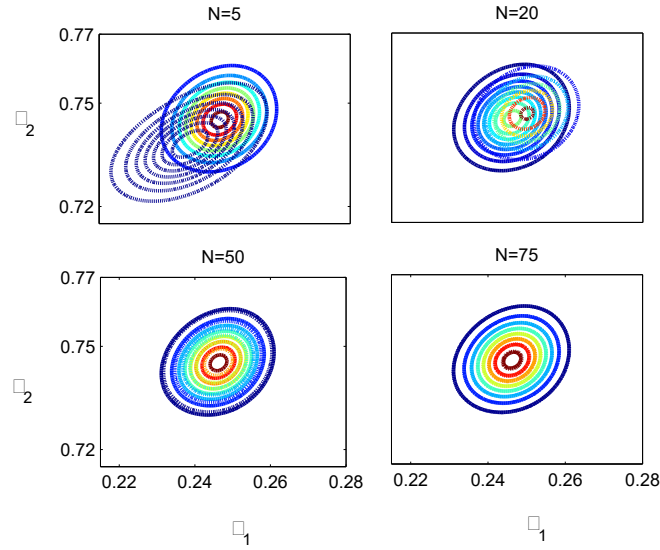


Figure 3: Contours of the posterior with the separated terms $N = 5, 20, 50$ and 75 . The solid lines represent the reference posterior and the dashed lines show the approximated one.

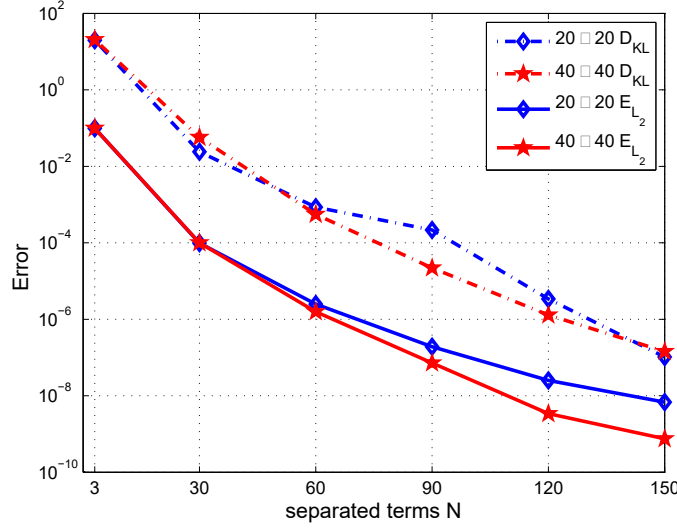


Figure 4: Approximation of the forward model and the posterior density with respect to the number of the separated terms. Dashed line represents the Kullback-Leibler divergence $D_{KL}(\pi_{\mathbf{N}}^d || \pi^d)$ (denoted by D_{KL}); solid line shows the L^2_{π} error $\|\mathbf{G} - \mathbf{G}_{\mathbf{N}}\|_{L^2_{\pi}}^2$ (denoted by E_{L_2}).

2×10^4 realizations are used to compute the relevant statistical quantities. The KL divergence D_{KL} between the approximated posterior and the reference posterior are discussed here to measure the difference.

First, we consider the effect of the separated terms N on the performance. In Figure 4, the approximation of the bi-fidelity surrogate model is plotted against the increasing numbers of separated terms when stochastic basis functions are constructed by 20×20 and 40×40 low-fidelity model, respectively. From this figure, the model approximation error $\|\mathbf{G} - \mathbf{G}_{\mathbf{N}}\|_{L^2_{\pi}}^2$ decreases as the rank of approximation increases for both low-fidelity models. The top curve of the figure shows that the posterior density of KL divergence D_{KL} decreases as we increase the rank of approximation, this is consistent with the result shown in Theorem 2.1. Figure 4 also shows when separated terms fixed, a surrogate model of which stochastic basis functions constructed by the 40×40 low-fidelity model approximates the forward model better. The KL divergence presents the same results. The more accurate the low-fidelity model is, the less the difference between the approximate posterior and the reference one will be. The figures confirm that the accuracy of the surrogate posterior depends both on the separated terms N and the low-fidelity model.

It would be interesting to recover the fractional order α_{ξ} in the model (4.26) simultaneously with the point location. To this end, we keep the same conditions. But the unknown inputs are $\xi = [\alpha_{\xi}; \xi_1, \xi_2]$, and $T = 0.2$. We also set $T_1 = T$, which implies that the source is always active during the considered time interval $[0, T]$. The ground true value of α_{ξ} is 0.3, measurements are taken at the same locations and at time $t = 0.05, 0.1, 0.15$. We find that the NVS method is sensitive to the training sample set when α_{ξ} is also unknown.

We use the 30×30 low-fidelity model to obtain the corresponding interpolation rule in constructing the bi-fidelity model. When the surrogate model with rank $N = 120$ is implemented to approximate the likelihood, the approximate posterior is explored with $\beta = 0.0033$, and the ac-

ceptance rate is 31.55%. Figure 5 shows the one- and two- dimensional posterior marginals of ξ . The posterior of α_ξ centers around 0.3. The posterior support of location is relatively wider than α_ξ 's. The modes appear uncorrelated and mutually independent based on the shape of their two-dimensional marginals.

The forward model approximation error and KL divergence are plotted against the number of separated terms in Figure 6. For the first 160 separated terms, the difference in forward model drops about 4 order of magnitude, with the corresponding 6 orders of magnitude decreasing in D_{KL} for both surrogate models. The 40×40 low-fidelity model based surrogate performs little better than 30×30 based one as the separated terms increases to 200. Generally, the posterior density of Kullback-Leibler divergence D_{KL} decreases twice as fast as $\|\mathbf{G} - \mathbf{G}_N\|_{L^2_\pi}^2$.

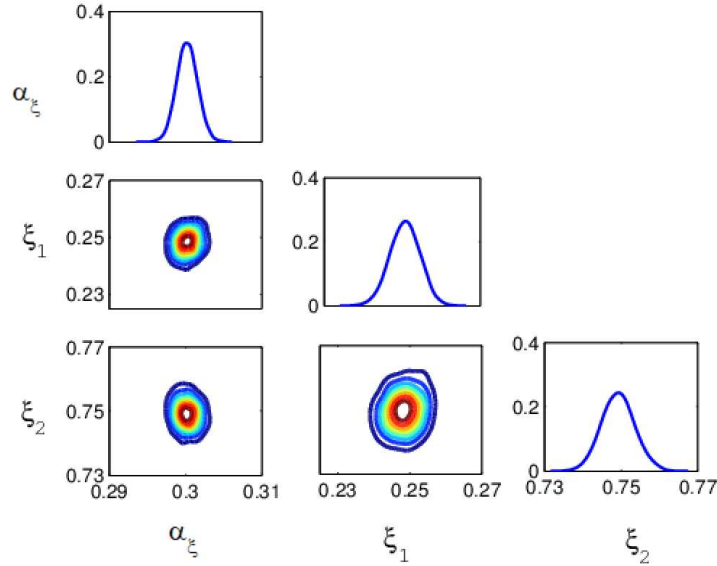


Figure 5: one-dimensional and 2-dimensional posterior marginals of $[\alpha_\xi; \xi_1; \xi_2]$.

4.2 Identify the smooth permeability field

Consider the problem with the mixed boundary condition, where the homogeneous Dirichlet boundary condition is imposed on $x = 0$ and $x = 1$, and there is no flow on the other lateral boundaries. The end time is $T = 0.25$, the fractional derivative is set as $\alpha_\xi = 0.5$, and the source term is set as $f = 30$. The random field is given by the truncated Karhunen-Loève expansion with the first n_ξ terms,

$$\kappa(x; \xi) = \bar{\kappa}_c + \sum_{i=1}^{n_\xi} \sqrt{\zeta_i} \varphi_i(x) \xi_i, ,$$

where $(\zeta_i, \varphi_i(x))$ are eigenpairs of the Gaussian covariance kernel

$$\text{Cov}(x_1, y_1; x_2, y_2) = \sigma_\xi^2 \exp \left(-\frac{|x_1 - x_2|^2}{2l_x^2} - \frac{|y_1 - y_2|^2}{2l_y^2} \right), \quad (4.27)$$

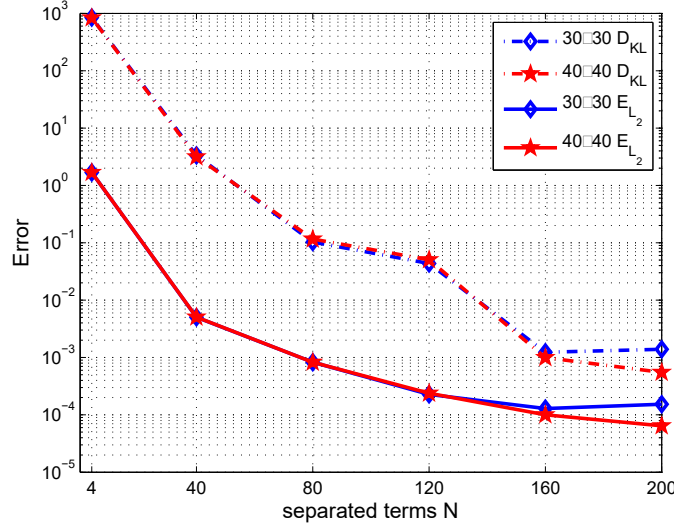


Figure 6: Approximation of the forward model and the posterior density with respect to the number of separated terms. Dashed line represents the Kullback-Leibler divergence $D_{KL}(\pi_N^d || \pi^d)$ (denoted by D_{KL}); solid line shows the L_π^2 error $\|\mathbf{G} - \mathbf{G}_N\|_{L_\pi^2}^2$ (denoted by E_{L_2}).

here $l_x = 0.2$, $l_y = 0.3$, $\sigma_\xi^2 = 1$, $\bar{\kappa}_c = 10$ is the mean of the random field. And the number of truncated terms is $n_\xi = 40$.

The ground true parameter ξ is randomly drawn from the standard multivariate normal distribution. Measurements are taken from 0.02 to 0.21, with the time interval 0.01, and the locations are distributed on a uniform 7×6 grid of the domain $[0.1, 0.85] \times [0, 0.9]$. We use Gaussian distribution $\mathcal{N}(0, \lambda \mathbb{I})$ as the prior. The parameter λ from the prior is treated as a hyperparameter here. We choose inverse Gamma distribution as its hyperprior density, i.e., we have the joint posterior density

$$\pi(\xi, \lambda | d) \propto \exp\left(-\frac{\|d - \mathbf{G}(\xi)\|_2^2}{2\sigma^2}\right) \exp\left(-\frac{\|\xi\|_2^2}{2\lambda}\right) \lambda^{-\frac{n_\xi}{2} - a - 1} \exp\left(-\frac{b}{\lambda}\right),$$

which yields an inverse Gamma full conditional distribution

$$\pi(\lambda | \xi, d) = \mathcal{IG}\left(a + \frac{n_\xi}{2}, b + \frac{\|\xi\|_2^2}{2}\right).$$

pCN-MCMC method is used to obtain the samplers of ξ and Gibbs method is employed to update the hyperparameter λ during the iteration. Here $a = 0.01$ and $b = 0.01$ for this example.

The bi-fidelity surrogate model is constructed based on a 20×20 low-fidelity and a 100×100 high-fidelity model. We note the upscaling technique [44] is used here when low-fidelity model is constructed, since the Karhunen-Loève expansion is generated to match with the finer grid. For each coarse domain K as shown in Figure 7, we solve the local problem

$$\text{div}(\kappa(x) \nabla \phi_j) = 0, \quad (4.28)$$

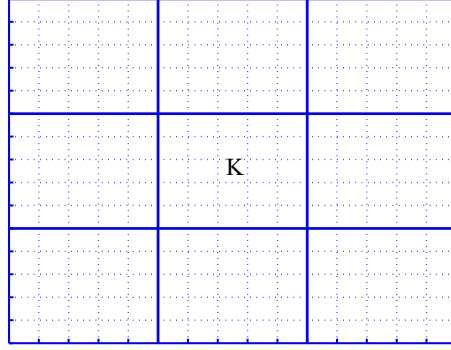


Figure 7: Schematic description of fine and coarse grids. Solid lines illustrate a coarse-scale partitioning, while dashed lines show a fine-scale partitioning within coarse-grid cells.

with some coarse-scale boundary conditions. Here $\kappa(x)$ denotes the fine-scale permeability field. The boundary conditions are given by $\phi_j = 1$ and $\phi_j = 0$ on the opposite sides along the direction e_j and no-flow boundary conditions on the other lateral sides. The coarse-scale permeability tensor is then given by

$$(\kappa^*(x)e_j, e_i) = \frac{1}{|K|} \int_K (\kappa(x) \nabla \phi_j, e_i) dx,$$

where ϕ_j is the solution of (4.28) with prescribed boundary conditions. Once the upscaled absolute permeability $\kappa^*(x)$ is obtained, the original equation can be solved on the coarse grid without changing the form of relative permeability curves.

When the approximation rank is set as $N = 100$, the corresponding surrogate model is implemented to explore the approximate posterior. Parameter β is set as 0.075 and the length of the Markov chain is 5×10^4 . The last 2×10^4 realizations are retained. Figure 8 shows the reference permeability field and some posterior realizations of it. The general shapes of the realizations match the reference field, especially for spatial locations where the values are high. Though they are not almost the same, each realization capture part characteristic of the target profile.

The 95% credible intervals for the model response at $u(x = 0.55, y, t = 0.05)$, $u(x, y = 0.45, t = 0.1)$ and $u(x = 0.25, y = 0.45, t)$ are constructed. This uncertainty is added to the estimated error variance to construct prediction intervals. As illustrated in Figure 9, the reference values are located in the credible intervals in general, measurements are almost contained in the predictive intervals. For $u(x, y = 0.45, t = 0.1)$, both the credible interval and prediction interval become tight as x gets closer to the end points, which results from the deterministic Dirichlet boundary condition at $x = 0$ and $x = 1$. For $u(x = 0.25, y = 0.45, t)$, the studied intervals get loose as time moves on, which means that the uncertainty associated with the model fit and predictions grows, as the uncertainty from the input ξ propagates.

The accuracy of the surrogate model is discussed here. As shown in Figure 10, the forward model error drops dramatically during the first hundred separated terms for both low-fidelity based surrogate model, the difference in forward model drops more than two orders of magnitude from $N = 10$ to $N = 50$ and continues to fall towards $N = 150$. The 40×40 low-fidelity based surrogate model decreases faster than the 20×20 low-fidelity based one after 150 separated terms, the KL

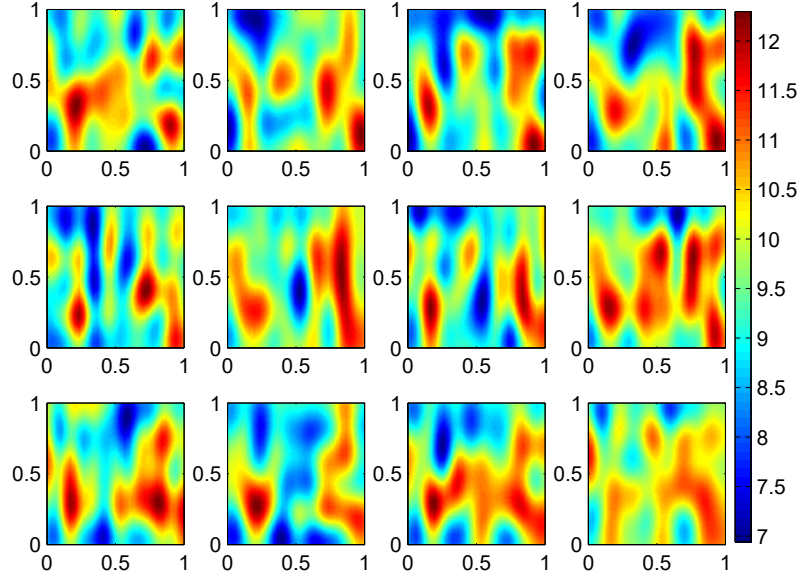


Figure 8: Reference and posterior realization of the permeability field, the left upper one is the reference one and the others are posterior realizations.

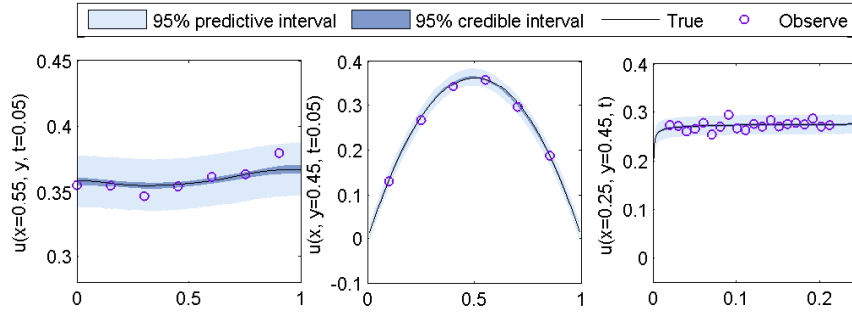


Figure 9: Data, point estimates, and 95% credible and prediction intervals produced by the Bayesian analysis for $u(x=0.55, y, t=0.05)$, $u(x, y=0.45, t=0.1)$ and $u(x=0.25, y=0.45, t)$, respectively.

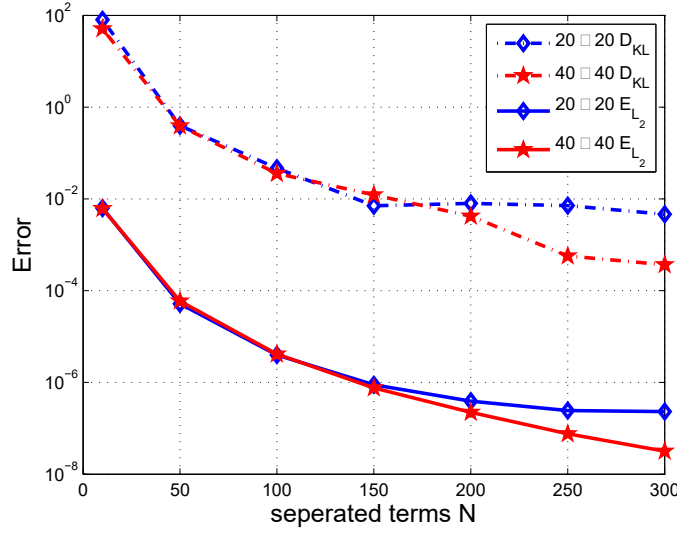


Figure 10: Approximation of the forward model and the posterior density with respect to the number of separated terms. Dashed line shows the Kullback-Leibler divergence $D_{KL}(\pi_{\mathbf{N}}^d || \pi^d)$ (denoted by D_{KL}); solid line presents the L_{π}^2 error $\|\mathbf{G} - \mathbf{G}_{\mathbf{N}}\|_{L_{\pi}^2}^2$ (denoted by E_{L_2}).

divergence curves present the same trend with the forward model error in general. For the 100×100 forward model with 40-dimensional unknown parameters, the D_{KL} can achieve magnitude 10^{-3} accuracy with the bi-fidelity surrogate model, where the 40×40 low-fidelity model requires to solve only for 300 times.

4.3 Identify the channel structured permeability field

Consider the problem with the same conditions as example 4.2, but the permeability field is channel-structured as a priori and the source term is set as

$$f(x) = 15 \exp \left(- \frac{(x_1 - 0.4)^2 + (x_2)^2}{2 \times 0.3^2} \right).$$

DCT is used to parameterize the unknown permeability field, which uses real cosine functions and transforms information into coefficients of cosine functions, more details on DCT and Inverse DCT can be found on [35].

The channeled coefficient is parameterised as

$$\kappa(x; \xi) = \bar{\kappa}_c + \tilde{\kappa}(x; \xi),$$

where $\bar{\kappa}_c$ is a constant expected field, and $\kappa(x; \xi)$ is discretized as

$$\tilde{\kappa}(n_x, n_y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \alpha(i) \alpha(j) \xi(i, j) \cos \left[\frac{\pi(2n_x - 1)(i - 1)}{2N_x} \right] \cos \left[\frac{\pi(2n_y - 1)(j - 1)}{2N_y} \right], \quad (4.29)$$

$\alpha(i) = \begin{cases} \frac{1}{\sqrt{N_x}}, & i = 1, \\ \sqrt{\frac{2}{N_x}}, & \text{otherwise,} \end{cases} \quad \text{and} \quad \alpha(j) = \begin{cases} \frac{1}{\sqrt{N_y}}, & j = 1, \\ \sqrt{\frac{2}{N_y}}, & \text{otherwise,} \end{cases}$
 where $n_x = 1, \dots, N_x$, $n_y = 1, \dots, N_y$, and $N_p = N_x N_y$. Denote the 1-dimension DCT basis matrix as $I_x \in \mathbb{R}^{N_x \times N_x}$ with elements

$$[I_x]_{i,n_x} = \alpha(i) \cos \left[\frac{\pi(2n_x - 1)(i - 1)}{2N_x} \right],$$

and $I_y \in \mathbb{R}^{N_y \times N_y}$ with elements

$$[I_y]_{j,n_y} = \alpha(j) \cos \left[\frac{\pi(2n_y - 1)(j - 1)}{2N_y} \right],$$

then 2-dimension DCT basis can be represented as

$$\Phi = I_x \otimes I_y,$$

where $\otimes$ means the Kronecker product. Due to the arrangement of the DCT's cosine functions, the lower frequency focus on the upper-left part, i.e., a set of $n_\xi < N_p$ DCT-weighting coefficients can reproduce the original image, we use the truncated DCT basis functions to approximate $\tilde{\kappa}$, the equation (4.29) can be rewritten in vector form

$$\tilde{\kappa} = \tilde{\Phi} \xi,$$

where $\tilde{\Phi} \in \mathbb{R}^{N_p \times n_\xi}$ is the first n_ξ columns of matrix Φ , $\xi \in \mathbb{R}^{n_\xi}$ is the corresponding coefficient.

Before conducting the inference, we need to quantify the prior information of ξ . Firstly, since $\tilde{\kappa}$ is spatial-dependent, we naturally choose Gaussian prior for it. Given that the permeability field is channel structured, to describe the jump better, we impose TV norm penalty for this inverse problem, i.e. the prior density for $\pi(\tilde{\kappa})$ is given by

$$\pi(\tilde{\kappa}) \propto \exp \left(-\frac{\tilde{\kappa}^T \Sigma^{-1} \tilde{\kappa}}{2} \right) \exp(-\gamma \|\tilde{\kappa}\|_{TV}),$$

where Σ is the covariance matrix. The TV norm of a 2-D image $\tilde{\kappa} \in \mathbb{R}^{N_p}$ is approximated by

$$\|\tilde{\kappa}\|_{TV} = \sum_{i=1}^2 \|\mathbf{R}_i \tilde{\kappa}\|_1,$$

$\|\cdot\|_1$ is l_1 norm and $\mathbf{R}_i$, $i = 1, 2$ is a finite difference matrix in direction i . The second-order finite-difference strategy is used here, e.g., in 1-D problem, $\mathbf{R}_i$ is defined as

$$\mathbf{R}_i := \begin{bmatrix} -1 & 1 & & & \\ -\frac{1}{2} & 0 & \frac{1}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{2} & 0 & \frac{1}{2} \\ & & & -1 & 1 \end{bmatrix}.$$

Substitute equation (4.29) in, we have the hybrid prior for ξ

$$\pi(\xi) \propto \exp\left(-\frac{\xi^T \tilde{\Phi}^T \Sigma^{-1} \tilde{\Phi} \xi}{2}\right) \exp(-\gamma \|\tilde{\Phi} \xi\|_{TV}),$$

The covariance matrix Σ is generated by Eq. (4.27) with $\sigma_\xi = 0.1$, $l_x = l_y = 0.05$, $\bar{\kappa}_c$ is set as 1.5 in this example. The unknown permeability field is uniformly discretised into 40×40 pixels, i.e., the dimension of the original coefficient is $N_p = 1600$, when DCT applied, we set $n_\xi = 400$. Measurements are taken the same as in example 4.2.

We construct the bi-fidelity model based on the 40×40 low-fidelity model, the bases are calculated by 120×120 high-fidelity model. The surrogate model built with $N = 100$ separated terms is used to approximate the likelihood function. The regularization parameter γ from the TV prior is set as $\gamma = 100$, and the pCN-MCMC parameter $\beta = 0.035$ in this example. The length of the Markov chain is 8×10^4 , the last 3×10^4 realizations are retained. We compare MCMC mean estimate with the reference one and present the posterior standard deviation in the Figure 11 (a)(c), the region of high values has been captured, the curves of the channel have also been described generally.

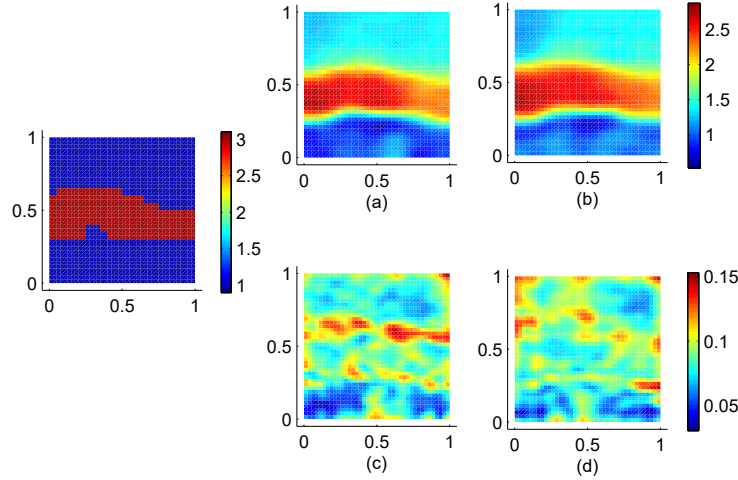


Figure 11: The reference channel structured permeability field (the left), posterior mean (the first row) and posterior standard deviation (the second row) of $\kappa(x)$, the estimates (a) and (c) are based on the posterior samplers of ξ , and (b) and (d) are based on the posterior samplers of ζ .

We sort the parameters in descending order according to $\text{var}(\xi_i|d)$, the resorted parameter is denoted as $\tilde{\xi}$. As can be seen in Figure 12, posterior samplers of the parameters $\tilde{\xi}_i$ and its variance are plotted with the new order, the red pentagram is $\text{var}(\tilde{\xi}_i|d) \times 10$, we multiply the variance with 10 to make samplers $\tilde{\xi}_i$ and $\text{var}(\tilde{\xi}_i|d)$ under the same scale to be presented in the same figure. we then do inference to the first n_ζ terms of $\tilde{\xi}$, where

$$\frac{\sum_{i=1}^{n_\zeta} \text{var}(\tilde{\xi}_i|d)}{\sum_{i=1}^{n_\xi} \text{var}(\xi_i|d)} = 0.95,$$

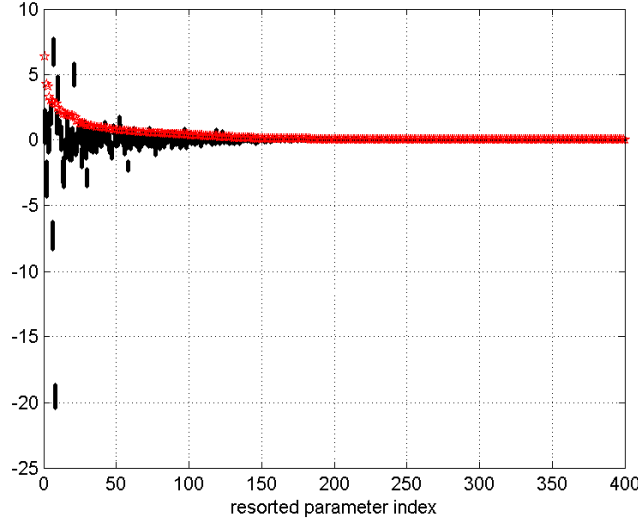


Figure 12: Resorted posterior samplers of parameters $\tilde{\xi}_i$ and its variance, the pentagram is $\text{var}(\tilde{\xi}_i|d) \times 10$, black dot is the posterior samplers.

and $n_\zeta = 152$ is estimated from the posterior samples in this example. We denote the dimension reduction parameter as ζ , and construct bi-fidelity surrogate model with respect to it.

The low-fidelity model is then built with a 80×80 grid, we build the surrogate model with $N = 100$ separated terms, the regularization parameter and pCN-MCMC parameter (the β) are set to be the same. The MCMC mean estimate and posterior standard deviation are presented in Figure 11 (b)(d), ζ based mean estimates also capture the channel structure in general. Compared with ξ based estimates, the deviation around the edges has been reduced.

Figure 13 shows the histogram of the reference, 0.25 quantile, mean and 0.95 quantile estimate of the permeability. The reference has a bimodal distribution in the histogram, subfigures (a)(b)(c) are estimated from the posterior samplers of ξ and (d)(e)(f) are from the posterior samplers of ζ . The high frequency value concentrates around 1 and 3 for the estimates, this distribution property is more obvious for ζ based estimates, which is lead by relative sufficient data information as dimension of ζ is smaller than ξ .

We also discuss the influence of parameter reduction on interval estimates of responses. We construct the 95% credible intervals for model response at $u(x = 0.4, y, t = 0.05)$, $u(x, y = 0.6, t = 0.1)$ and $u(x = 0.25, y = 0.3, t)$. As shown in Figure 14, the true response are not totally contained in credible intervals, this may be result from the DCT truncation. However, the measurement data are almost contained in predictive intervals for both cases, they can both explain the data.

5 Conclusion

This paper presented a new bi-fidelity model reduction method and its application in Bayesian inverse problems for the time-fractional equation. We obtain stochastic basis functions by the low-

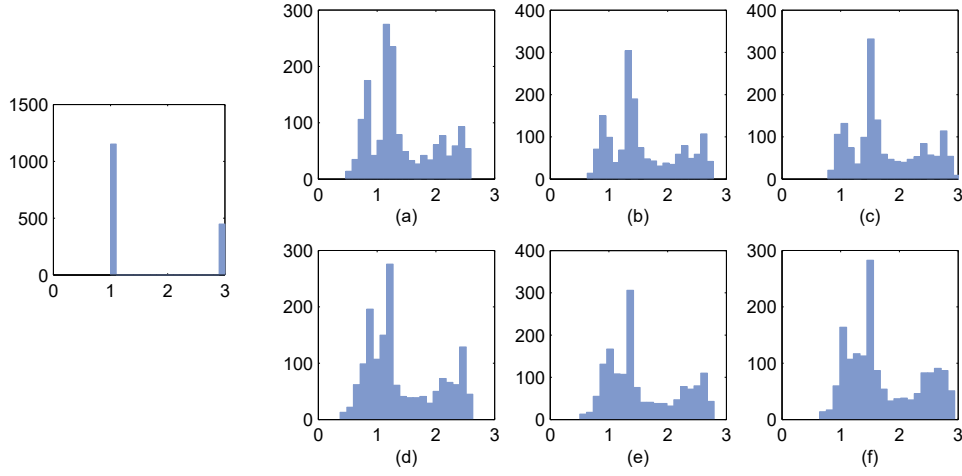


Figure 13: Permeability histogram: reference (left) and estimates based on the posterior samplers of ξ (the first row) and ζ (the second row), (a) and (d) are 0.25 quantile, (b) and (e) are mean, (c) and (f) are 0.95 quantile.

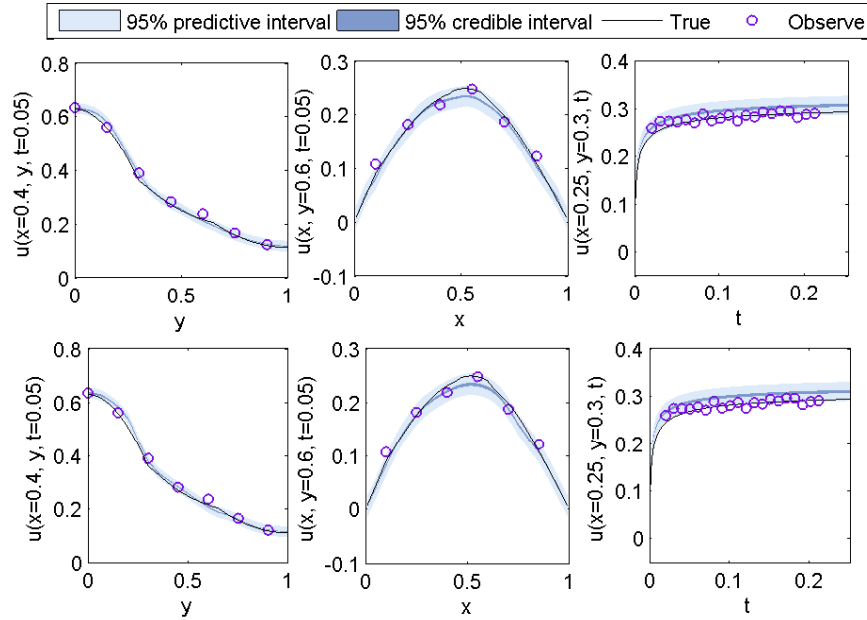


Figure 14: Data, point estimates, and 95% credible and prediction intervals produced by the Bayesian analysis for $u(x=0.4, y, t=0.05)$, $u(x, y=0.6, t=0.1)$ and $u(x=0.25, y=0.3, t)$. The first row is estimated based on the posterior samplers of ξ and the second row is obtained from the posterior samplers of ζ .

fidelity model, and the NVS method is used. The efficiency of the low-fidelity model leads to the reduction of time in calculating residual and matrices involved in the recursive formulation for stochastic basis functions. The bases in the physical space are obtained by running the high-fidelity model in parallel, which is due to the mechanism of the NVS method and the linearity of the fractional equation. The presented strategy leads to an accurate approximation of the full-order forward model and gives an approximated posterior density, which is easier to be evaluated than the original posterior. The numerical examples confirm that the approximated posterior matches the reference one very well by using the proposed method, the accuracy of the surrogate posterior depends both on the separated terms N and the low-fidelity model.

We demonstrate our approach in the time-fractional equation inversion problems: estimating the source location, the time-fractional order and inferring the permeability field with a different structure. The new bi-fidelity model reduction method constructs the random space basis according to the model itself, inheriting the advantage of the NVS method, i.e., it escapes from the limitation of polynomial basis and can be used to high-dimensional stochastic problems. The implementation of multi-fidelity models accelerates the construction of stochastic basis functions while the accuracy can still be preserved. When constructing the low-fidelity model for the unknown permeability field, the upscaling technique is used, for the target profile with high oscillation, GMSFEM [24] can be used to approximate the solutions better. We also plan to do some mathematical analysis on the convergence of the proposed method in our future work.

Appendix

Proof: Denote $\mathring{\mathbf{U}}_{1:N}^L = [\mathring{U}_1^L, \mathring{U}_2^L, \dots, \mathring{U}_N^L] \in \mathbb{R}^{N_L \times N}$, $\mathring{\mathbf{U}}_{1:N}^H = [\mathring{U}_1^H, \mathring{U}_2^H, \dots, \mathring{U}_N^H] \in \mathbb{R}^{N_H \times N}$ the set of full low- and high- fidelity model solution with respect to samples $\{\xi_L^{(k)}\}_{k=1}^N$, respectively. According to Eq.(3.24), we have

$$\mathring{\mathbf{U}}_{1:N}^H = \mathbf{T}\mathring{\mathbf{U}}_{1:N}^L + \mathbf{E}^S, \quad (5.30)$$

where $\mathbf{E}^S$ is matrix of errors corresponding to samples $\{\xi_L^{(k)}\}_{k=1}^N$, of which columns are part of $\mathbf{E}$. Let $\mathbf{U}_{1:N}^L = [U_1^L, U_2^L, \dots, U_N^L] \in \mathbb{R}^{N_L \times N}$, $\mathbf{U}_{1:N}^H = [U_1^H, U_2^H, \dots, U_N^H] \in \mathbb{R}^{N_H \times N}$ be the set of low- and high- fidelity basis vectors, respectively. Due to Eqs. (3.22) and (3.23), we have

$$\mathring{\mathbf{U}}_{1:N}^L = \mathbf{U}_{1:N}^L \mathbf{C}_L^S,$$

$$\mathring{\mathbf{U}}_{1:N}^H = \mathbf{U}_{1:N}^H \mathbf{C}_L^S.$$

Substitute them into Eq. (5.30), we obtain

$$\mathbf{U}_{1:N}^H - \mathbf{T}\mathbf{U}_{1:N}^L = (\mathbf{C}_L^S)^{-1} \mathbf{E}^S,$$

finally, we have

$$\begin{aligned} \|\mathring{\mathbf{U}}^H - \tilde{\mathbf{U}}_N\| &\leq \|\mathring{\mathbf{U}}^H - \mathbf{T}\mathring{\mathbf{U}}^L\| + \|\mathbf{T}\mathring{\mathbf{U}}^L - \mathbf{T}\mathbf{U}_N^L\| + \|\mathbf{T}\mathbf{U}_N^L - \tilde{\mathbf{U}}_N\| \\ &\leq \|\mathring{\mathbf{U}}^H - \mathbf{T}\mathring{\mathbf{U}}^L\| + \|\mathbf{T}\| \|\mathring{\mathbf{U}}^L - \mathbf{U}_N^L\| + \|\mathbf{T}\mathbf{U}_{1:N}^L \mathbf{C}_L - \mathbf{U}_{1:N}^H \mathbf{C}_L\| \\ &\leq \|\mathbf{E}\| + \|\mathbf{T}\| \|\mathring{\mathbf{U}}^L - \mathbf{U}_N^L\| + \|(\mathbf{C}_L^S)^{-1}\| \|\mathbf{E}^S\| \|\mathbf{C}_L\| \\ &\leq (1 + \|(\mathbf{C}_L^S)^{-1}\| \|\mathbf{C}_L\|) \|\mathbf{E}\| + \|\mathbf{T}\| \|\mathring{\mathbf{U}}^L - \mathbf{U}_N^L\|. \end{aligned}$$

Acknowledgments

We acknowledge the support of Chinese NSF 11471107. We also thank the discussion with Qiuqi Li. Lin would like to acknowledge the support from National Science Foundation (DMS-1555072, DMS-1736364, and DMS-1821233).

References

- [1] Evans SN, Stark PB. Inverse problems as statistics. *Inverse problems* 2002; 18(4): R55.
- [2] Wang YP, Tao SL, Chen Q. Retrieving the variable coefficient for a nonlinear convection-diffusion problem with spectral conjugate gradient method. *Inverse Problems in Science & Engineering* 2015; 23(8): 1342-1365.
- [3] Congdon P. *Bayesian statistical modelling*. 704. John Wiley & Sons . 2007.
- [4] Stuart AM. Inverse problems: a Bayesian perspective. *Acta Numerica* 2010; 19: 451–559.
- [5] Kaipio J, Somersalo E. *Statistical and computational inverse problems*. 160. Springer Science & Business Media . 2006.
- [6] Liu JS. *Monte Carlo strategies in scientific computing*. Springer Science & Business Media . 2008.
- [7] Brooks S, Gelman A, Jones G, Meng XL. *Handbook of markov chain monte carlo*. CRC press . 2011.
- [8] Efendiev Y, Datta-Gupta A, Ginting V, Ma X, Mallick B. An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. *Water Resources Research* 2005; 41(12): 12423.
- [9] Efendiev Y, Jin B, Michael P, Tan X. Multilevel Markov chain Monte Carlo method for high-contrast single-phase flow problems. *Communications in Computational Physics* 2015; 17(1): 259–286.
- [10] Leimkuhler B, Matthews C, Weare J. Ensemble preconditioning for Markov chain Monte Carlo simulation. *Statistics and Computing* 2017: 1–14.
- [11] Emerick AA, Reynolds AC. Combining the Ensemble Kalman Filter With Markov-Chain Monte Carlo for Improved History Matching and Uncertainty Characterization. *Spe Journal* 2012; 17(2): 418-440.
- [12] Martin J, Wilcox LC, Burstedde C, Ghattas O. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing* 2012; 34(3): A1460–A1487.
- [13] Biegler L, Biros G, Ghattas O, et al. 7. *Surrogate and Reduced-Order Modeling: A Comparison of Approaches for Large-Scale Statistical Inverse Problems*. John Wiley & Sons, Ltd . 2010.

- [14] Xiu D, Karniadakis GE. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing* 2002; 24(2): 619–644.
- [15] Marzouk Y, Xiu D. A stochastic collocation approach to Bayesian inference in inverse problems. 2009.
- [16] Yan L, Guo L. Stochastic Collocation Algorithms Using L₁-Minimization for Bayesian Solution of Inverse Problems. *SIAM Journal on Scientific Computing* 2015; 37(3): A1410–A1435.
- [17] Jakeman JD, Eldred MS, Sargsyan K. Enhancing l₁-minimization estimates of polynomial chaos expansions using basis selection. *Journal of Computational Physics* 2015; 289: 18–34.
- [18] Yang X, Lei H, Baker NA, Lin G. Enhancing sparsity of Hermite polynomial expansions by iterative rotations. *Journal of Computational Physics* 2016; 307: 94–109.
- [19] Jiang L, Ou N. Multiscale model reduction method for Bayesian inverse problems of subsurface flow. *Journal of Computational and Applied Mathematics* 2017; 319: 188–209.
- [20] Rasmussen CE, Williams CK. *Gaussian process for machine learning*. MIT press . 2006.
- [21] Bilonis I, Zabaras N. Multidimensional adaptive relevance vector machines for uncertainty quantification. *SIAM Journal on Scientific Computing* 2012; 34(6): B881–B908.
- [22] Bilonis I, Zabaras N. Solution of inverse problems with limited forward solver evaluations: a Bayesian perspective. *Inverse Problems* 2013; 30(1): 015004.
- [23] Hinze M, Volkwein S. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. In: Springer. 2005 (pp. 261–306).
- [24] Efendiev Y, Galvis J, Hou TY. Generalized multiscale finite element methods (GMsFEM). *Journal of Computational Physics* 2013; 251: 116–135.
- [25] Li Q, Jiang L. A Novel Variable-Separation Method Based on Sparse and Low Rank Representation for Stochastic Partial Differential Equations. *SIAM Journal on Scientific Computing* 2017; 39(6): A2879–A2910.
- [26] Robinson TD, Eldred MS, Willcox KE, Haines R. Surrogate-Based Optimization Using Multifidelity Models with Variable Parameterization and Corrected Space Mapping. *Aiaa Journal* 2008; 46(11): 2814–2822.
- [27] Perdikaris P, Karniadakis GE. Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond. *Journal of The Royal Society Interface* 2016; 13(118): 20151107.
- [28] Zhu X, Narayan A, Xiu D. Computational Aspects of Stochastic Collocation with Multifidelity Models. *Society for Industrial & Applied Mathematics* 2014; 2(1): 444463.
- [29] Peherstorfer B, Willcox K, Gunzburger M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review* 2018; 60(3): 550–591.

- [30] Christen JA, Fox C. Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical statistics* 2005; 14(4): 795–810.
- [31] Picheny V, Ginsbourger D, Richet Y, Caplin G. Quantile-based optimization of noisy computer experiments with tunable precision. *Technometrics* 2013; 55(1): 2–13.
- [32] Benson DA, Wheatcraft SW, Meerschaert MM. Application of a fractional advection-dispersion equation. *Water Resources Research* 2000; 36(6): 1403–1412.
- [33] Bondarenko AN, Ivaschenko DS. Numerical methods for solving inverse problems for time fractional diffusion equation with variable coefficient. *Journal of Inverse and Ill-Posed Problems* 2009; 17(5): 419–440.
- [34] Miller L, Yamamoto M. Coefficient inverse problem for a fractional diffusion equation. *Inverse Problems* 2013; 29(7): 075013.
- [35] Rao KR, Yip P. *Discrete cosine transform: algorithms, advantages, applications*. Academic press . 2014.
- [36] Yao Z, Hu Z, Li J. A TV-Gaussian prior for infinite-dimensional Bayesian inverse problems and its numerical implementations. *Inverse Problems* 2016; 32(7): 075006.
- [37] Vogel CR. *Computational methods for inverse problems*. SIAM . 2002.
- [38] Marzouk YM, Najm HN. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics* 2009; 228(6): 1862–1902.
- [39] Yan L, Zhang YX. Convergence analysis of surrogate-based methods for Bayesian inverse problems. *Inverse Problems* 2017; 33(12): 125001.
- [40] Li C, Zeng F. *Numerical methods for fractional calculus*. Chapman and Hall/CRC . 2015.
- [41] Barrault M, Maday Y, Nguyen NC, Patera AT. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathématique* 2004; 339(9): 667–672.
- [42] Grepl MA, Maday Y, Nguyen NC, Patera AT. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis* 2007; 41(3): 575–605.
- [43] Hampton J, Fairbanks H, Narayan A, Doostan A. Parametric/ Stochastic Model Reduction: Low-Rank Representation, Non-Intrusive Bi-Fidelity Approximation, and Convergence Analysis. *arXiv preprint arXiv:1709.03661* 2017.
- [44] Durlofsky LJ, Jones RC, Milliken WJ. A nonuniform coarsening approach for the scale-up of displacement processes in heterogeneous porous media. *Advances in Water Resources* 1997; 20(5): 335–347.