Computers & Geosciences 108 (2017) 21–32

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

# Implicit sampling combined with reduced order modeling for the inversion of vadose zone hydrological data

Yaning Liu*, George Shu Heng Pau, Stefan Finsterle

*Earth & Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

ARTICLE INFO

*Keywords:*
Inverse modeling
Implicit sampling
Vadose zone
TOUGH2
Polynomial chaos expansion
Sparse Bayesian learning

ABSTRACT

Bayesian inverse modeling techniques are computationally expensive because many forward simulations are needed when sampling the posterior distribution of the parameters. In this paper, we combine the implicit sampling method and generalized polynomial chaos expansion (gPCE) to significantly reduce the computational cost of performing Bayesian inverse modeling. There are three steps in this approach: (1) find the maximizer of the likelihood function using deterministic approaches; (2) construct a gPCE-based surrogate model using the results from a limited number of forward simulations; and (3) efficiently sample the posterior distribution of the parameters using implicit sampling method. The cost of constructing the gPCE-based surrogate model is further decreased by using sparse Bayesian learning to reduce the number of gPCE coefficients that have to be determined. We demonstrate the approach for a synthetic ponded infiltration experiment simulated with TOUGH2. The surrogate model is highly accurate with mean relative error that is <0.035% in predicting saturation and <0.25% in predicting the likelihood function. The posterior distribution of the parameters obtained using our proposed technique is nearly indistinguishable from the results obtained from either an implicit sampling method or a Markov chain Monte Carlo method utilizing the full model.

## 1. Introduction

Hydrological models are crucial to the understanding and description of water cycles. Hydrological model parameters, such as site-specific material properties and process-related parameters, as well as boundary conditions and site geometry play a major role in the model's ability to predict the hydrological states. These parameters can be large scale, highly uncertain and difficult to measure (Abubakar et al., 2009; Liu and Gupta, 2007). Thus, inverse modeling is typically performed to infer the model parameter values based on the sparse observations of some observables, by matching the numerical model (the forward model) to measured data at discrete spatial and temporal points.

Inverse modeling techniques, in general, fall into two categories: deterministic and probabilistic inversions. The deterministic approach aims to find a single set of parameter values that represent the "best fit" given the observations and a criterion that measures the closeness between the model response and the observations. Therefore, the essence of a deterministic inversion is the minimization of the objective function, which measures the difference between the model and the observed data. Commonly used objective functions are derived based on the assumptions that the model is correct and the errors in the measured data are normally or exponentially distributed, resulting in the maximum likelihood and $L_1$ estimators, respectively (Tarantola, 2004). A broader selection of objective functions can be found in Finsterle and Najita (1998) and Schoups and Vrugt (2010). For detailed theory and computational methods for deterministic inversions, the readers are referred to Vogel (2002), Neto and da Silva Neto (2012), and Ramm (2005). Deterministic inversion usually requires some form of regularization since inverse problems are often ill-posed (Kabanikhin, 2008). The well-known Tikhonov regularization, for instance, is typically used in least-square problems. In addition, estimation of the inversion uncertainty within a deterministic framework requires strong assumptions about the error structure of the observation and parameters, and about the linearity of the forward problem (Carrera and Neuman, 1986).

Accurate characterization of the inversion uncertainty is desirable in many applications as it provides an informed representation of the distribution of parametric uncertainty that can be propagated through the forward model, as is often done in uncertainty quantification (UQ) (Mondal et al., 2010; Liu et al., 2015a). With probabilistic inversion methods, the inversion result is presented in the form of joint probability distributions instead of a single set of parameter values. A common way to achieve this is through the Bayesian probability theory, which relates the parameter posterior distribution conditioned on the observations to the product of the prior distribution

* Corresponding author.
*E-mail addresses:* yaningliu@lbl.gov (Y. Liu), gpau@lbl.gov (G.S.H. Pau), safinsterle@lbl.gov (S. Finsterle).

http://dx.doi.org/10.1016/j.cageo.2017.04.001
Received 28 May 2016; Received in revised form 18 February 2017; Accepted 3 April 2017
Available online 06 April 2017
0098-3004/ 

and likelihood function. Since the resulting posterior densities can be complex, they are typically discretely represented by a set of samples obtained through sampling methods. With the Markov chain Monte Carlo (MCMC) method (e.g., Andrieu et al., 2003), these samples are generated by an acceptance-rejection approach. However, evaluating the acceptance-rejection criteria in MCMC, which involves a forward simulation for each proposed sample, can be extremely inefficient if the fraction of the rejected samples is large. The "burn-in" period for a MCMC sampling can also be long, resulting in a large number of samples being wasted. In addition, the significant (high-probability) region of the posterior distribution may be small, which further decreases the efficiency of MCMC sampling. Recent advances in MCMC methods address some of these difficulties. For example, Vrugt et al. (2009) proposed a differential evolution adaptive Metropolis scheme (DREAM) that explores the optimized proposal distribution in parallel and extends its applicability to estimating multi-modal posterior densities; Goodman and Weare (2010) suggest a family of multi-particle MCMC samplers with an affine invariance property that can offer significantly improved performance over standard single-particle methods; and Martin et al. (2012) present a stochastic Newton MCMC method by constructing a proposal density based on local Gaussian approximation that is especially efficient for large-scale inversions.

As a powerful alternative, particle filters (PF) (Liu and Chen, 1998; Arulampalam et al., 2002) are sequential Monte Carlo methods used in data assimilation to update the discrete representation of posteriors of the state variables in the form of particles (samples) with associated weights as new observable data become available. PF are "embarrassingly parallel" and the computational complexity is independent of the dimensionality of the system. However, PF can suffer from "particle collapse" (sample impoverishment), where only a small fraction of the particles have non-negligible weights. A large number of particles are thus needed for a meaningful approximation of the posterior.

Recently, a variant of PF, named implicit particle filter (IPF) was developed by Chorin and Tu (2009) and Chorin et al. (2010, 2013) as a remedy for sample impoverishment. IPF searches samples in the high probability region of the posterior by connecting the target particles with a reference distribution through a mapping of one's choice. The quality of the particles can therefore be significantly enhanced, as the proportion of the particles with non-negligible weights increases, and the overall number of particles needed is reduced. A number of improvements and applications of IPF have been made since its inception. Morzfeld et al. (2012) proposed a random map procedure and applied it to assimilating data for a stochastic Lorenz attractor; Morzfeld and Chorin (2012) applied IPF to geomagnetic data assimilation with partial noise; Atkins et al. (2013) established the connection of IPF with variational data assimilation. More recently, Morzfeld et al. (2015) implemented IPF for estimating parameters for subsurface flow modeled by Darcy's law; they observed a faster convergence compared to Metropolis MCMC. We henceforth use the term "implicit sampling" adopted therein in the case of inverse modeling independent of time, while "implicit particle filter" is used for the data assimilation processes where the systems are dynamic and the parameters are time-dependent.

Implicit sampling (IS) involves the mapping from samples drawn from a different distribution, called the reference distribution, to the target particles, which may require solving a nonlinear equation for each given sample, depending on the type of the mapping chosen, and thus requires a large number of forward simulations. In view of this, we propose using a reduced order model (ROM) that serves as a surrogate for the forward model (see Razavi et al. (2012), Pau et al. (2014), Liu et al. (2016a, 2016b)), among others). The ROM is constructed with an initial set of forward simulations (training set), and subsequently substitutes the forward model. We develop our ROM based on the widely used generalized polynomial chaos expansion (gPCE) (Xiu and Karniadakis, 2002). The construction of gPCE requires determining the coefficients of the expansion terms once the type of the polynomial basis and expansion order are selected. However, the optimal polynomial expansion order is not a priori known. A low-order expansion may not accurately represent the response surface. On the other hand, a high-order expansion leads to exponentially large number of expansion terms. As a consequence, the number of forward simulations needed to estimate the gPCE coefficients also increases exponentially. The error associated with the overall gPCE can also increase since the estimation errors associated with the estimated gPCE coefficients may increase substantially as the expansion order increases.

However, the gPCE coefficients can be sparse, i.e., only a small number of the coefficients are non-zero. The sparsity is due to the following reasons: higher-order parameter interactions may not exist (Rabitz et al., 1999); the model response is smooth and so are the higher-order derivatives, leading to a fast decrease in the magnitudes of the coefficients as the polynomial orders increase; and the model response is, by nature, the superposition of only a sparse subset of all the polynomial bases up to a given order. Since the set of non-negligible gPCE coefficients is not a priori known, we can pose the problem as a sparse Bayesian learning (SBL) problem (Sargsyan et al., 2014) (also known as relevance vector machine and Bayesian compressive sensing (Tipping, 2001; Tipping and Faul, 2003)), where the model outputs are characterized by a hierarchical form of Gaussian likelihood and prior. Babacan et al. (2010) further demonstrated that using Laplace prior to model the sparsity improved the performance. With SBL, the sparsity is obtained by updating one gPCE coefficient at a time using a greedy algorithm that iteratively selects the most contributing coefficients until a prescribed stopping criterion is reached. The initial set of coefficients is set to be empty, and an efficient algorithm to update the coefficient set, either by including a new one, revising the value of an existing coefficient, or deleting an existing one, is described in Tipping and Faul (2003). Sargsyan et al. (2014) empirically demonstrated that good estimates of the gPCE coefficients can be obtained if the number of training samples is about five times that of non-zero coefficients. Depending on the ratio of the number of non-zero coefficients to the number of coefficients required by a particular expansion order, the number of forward model simulations to construct the ROM can be greatly reduced.

In the present work, we implement implicit sampling in inverse modeling for a synthetic ponded infiltration experiment. The goal of the inversion is to determine the permeability distribution of the vadose zone based on saturation measurements. We also demonstrate that the performance of the inversion can be further enhanced by constructing and utilizing a reduced order model based on generalized polynomial chaos expansion and sparse Bayesian learning. The proposed inversion method is compared to a state-of-the-art Markov chain Monte Carlo simulator described in Goodman and Weare (2010).

The rest of this paper is structured as follows. In the next section, we introduce the vadose zone hydrological forward model, followed by a mathematical description of the implicit sampling method and details about the construction of the reduced order model using generalized polynomial chaos expansion and sparse Bayesian learning. In Section 4, the main results and discussion of the inversion of the hydrological model are presented. In the end, we conclude the paper with possible improvements for future work.

## 2. A hydrological inverse modeling problem

The hydrological problem used to demonstrate our approach is a synthetic field experiment shown in Fig. 1. An infiltration pond releases water into a heterogeneous but structured vadose zone whose water table is at the depth of 3 m. The saturation distribution is initially in gravity-capillary equilibrium, and the infiltration rate is controlled in order for the water level in the pond to stay at 2 cm for 1 day. After that, the experiment proceeds without infiltration for another day. Water saturation is measured at 36 monitoring points (circles in Fig. 1) initially and every 2 h from 34th hour till the end of day 2; the amount of water flowing out of the pond is also measured at the same times. The subsurface flow is modeled by Richards' equation (Richards, 1931) as implemented in the integral finite difference simulator TOUGH2

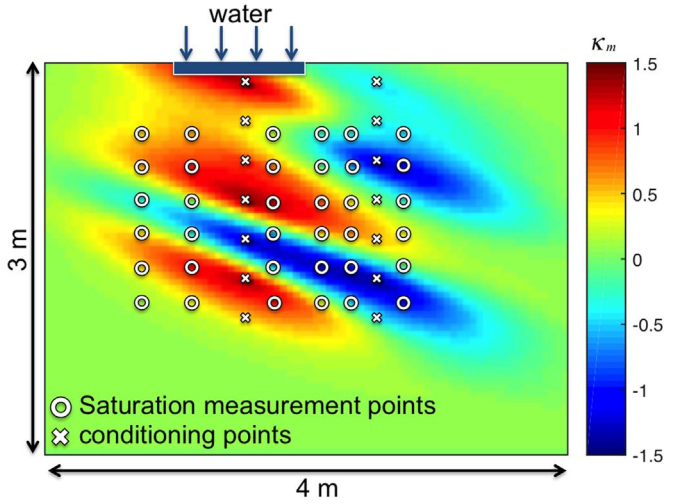


**Fig. 1.** Setup of the synthetic field experiment. Circles are the locations of the observation points and crosses are those of the pilot points. The background is the reference log-permeability modifier field.

(Pruess et al., 2012) with prescribed van Genuchten relative permeability and capillary pressure functions (van Genuchten, 1980). The computational domain is of size 4 m×3 m and is discretized with a constant grid spacing of 0.05 m.

The objective of the field experiment is to determine the permeability distribution of the vadose zone based on the measurements described above. We parameterize the distribution of the permeability using geostatistics and the pilot-point method (RamaRao et al., 1995; Gómez-Hernánez et al., 1997). Details of the pilot-point inverse modeling approach are described elsewhere, for example in Finsterle and Kowalsky (2008). Fourteen pilot points, for which the locations are shown in Fig. 1 (crosses), serve as conditioning points for Kriging (Deutsch and Journel, 1997). The permeability at a particular grid cell is described as

$$\kappa = \kappa_{\text{ref}} \times 10^{\theta} \tag{1}$$

where $\kappa_{\text{ref}} = 10^{-13}$ is a reference permeability value and $\theta$ is the exponent of a permeability modifier. Kriging is performed on $\theta$ based on a spherical semivariogram model with fixed parameters. We denote the 14 pilot point log-permeability modifiers as $\boldsymbol{\theta} = \{\theta_i, 1 \le i \le 14\}$, and the hydrological inverse problem consists of determining the values of $\boldsymbol{\theta}$. To generate the measurement data used in this experiment, we first prescribe the reference values ("true values") of $\boldsymbol{\theta}$, which are shown in Table 1. The resulting log-permeability modifier field after Kriging is shown in Fig. 1. We then run a forward simulation with the prescribed reference values and the corresponding saturation field at the end of the experiment is shown in Fig. 2. In addition, we added a zero mean Gaussian noise with a standard deviation of 0.01 for the saturation measurements and 0.05 for the volume of water flowing out of the pond. The number of observations used for calibration is 333 (37 locations and 9 times). We prescribe a Gaussian prior with alternating 0.1 and −0.1 as the means and a standard deviation of 1 for all parameters (also shown in Table 1), and the corresponding prior log-permeability modifier field is shown in Fig. 3. Compared to Fig. 1, where the reference log-permeability modifier field is shown, the prior permeability field shows a very different pattern.

**Table 1**
Reference log-permeability modifiers and the prior distributions at the 14 pilot points.

| pilot point # | true value | prior mean | prior standard deviation |
|---|---|---|---|
| 1 | 1.5 | 0.1 | 1.0 |
| 2 | 0.5 | −0.1 | 1.0 |
| 3 | 1.0 | 0.1 | 1.0 |
| 4 | 1.5 | −0.1 | 1.0 |
| 5 | −1.5 | 0.1 | 1.0 |
| 6 | 1.5 | −0.1 | 1.0 |
| 7 | 0.5 | 0.1 | 1.0 |
| 8 | −0.5 | −0.1 | 1.0 |
| 9 | −0.5 | 0.1 | 1.0 |
| 10 | −1.5 | −0.1 | 1.0 |
| 11 | −0.001 | 0.1 | 1.0 |
| 12 | 0.5 | −0.1 | 1.0 |
| 13 | −1.5 | 0.1 | 1.0 |
| 14 | −0.5 | −0.1 | 1.0 |

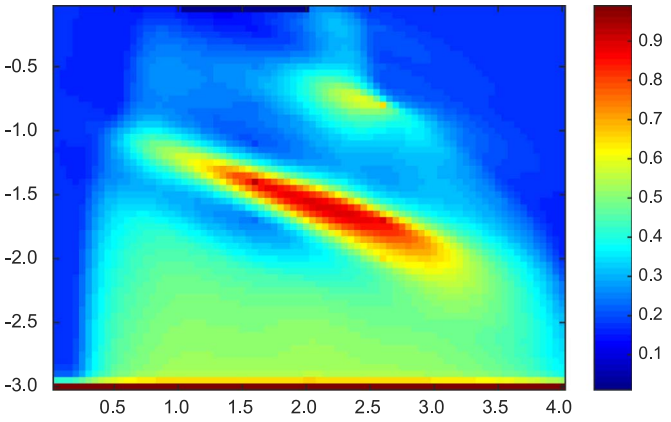


**Fig. 2.** Reference saturation at the end of the experiment.

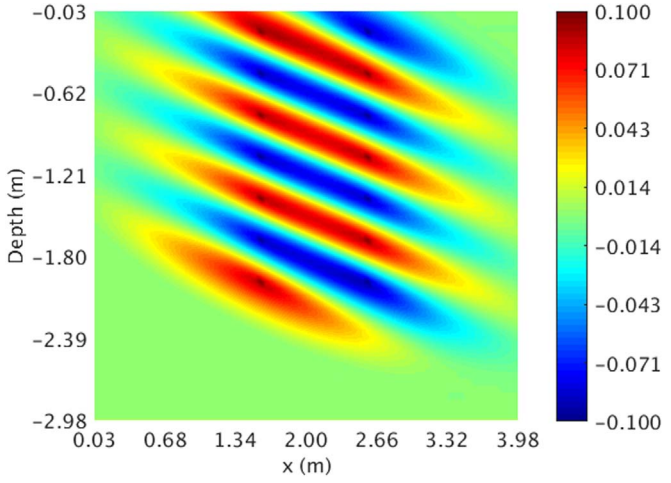


**Fig. 3.** Prior log-permeability modifier field.

## 3. Methodology

### 3.1. Implicit sampling

Assume the $k$-dimensional observations $\boldsymbol{D}$, forward model $\boldsymbol{f}$ and $d$-dimensional unknown input parameters $\boldsymbol{\theta}$ are related by

$$\boldsymbol{D} = \boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{\epsilon} \tag{2}$$

where $\boldsymbol{\epsilon}$ is the random noise. The central task of Bayesian parameter inversion is to obtain the conditional probability distribution of the parameters given the observations, $p(\boldsymbol{\theta}|\boldsymbol{D})$, known as the posterior density (or simply posterior). The well-known Bayes' theorem infers the posterior by combining the knowledge of the parameters prior to data being available, $p(\boldsymbol{\theta})$, the prior distribution (or just prior), and the measure of how likely a set of parameter values are given the observations $p(\boldsymbol{D}|\boldsymbol{\theta})$, the likelihood function (likelihood):

$$p(\boldsymbol{\theta}|\boldsymbol{D}) \propto p(\boldsymbol{\theta})p(\boldsymbol{D}|\boldsymbol{\theta}). \tag{3}$$

In general, $p(\boldsymbol{\theta}|\boldsymbol{D})$ cannot be sampled directly. One widely used method that facilitates such sampling is importance sampling. In the importance sampling framework, we instead sample from a new, easy-to-sample probability distribution $\pi(\boldsymbol{\theta})$, called importance distribution. The resulting $N$ posterior samples $\boldsymbol{\theta}_i$, $i = 1,\ldots,N$, are then weighted by

$$w_i = \frac{p(\boldsymbol{\theta}_i)p(\boldsymbol{D}|\boldsymbol{\theta}_i))}{\pi(\boldsymbol{\theta}_i)}, \quad i = 1, 2,\ldots,N \tag{4}$$

to correct the bias from taking samples from the importance distribution rather than the original distribution (Lu and Zhang, 2003). A resampling procedure (Arulampalam et al., 2002) can be used to eliminate samples with small weights and obtain a set of samples that discretely represents the posterior distributions of $\boldsymbol{\theta}$. Unlike MCMC, the samples in importance sampling are independent and thus can be embarrassingly parallelized. Nonetheless, the importance function must be chosen carefully, or else the sampling can be inefficient.

In implicit sampling (Chorin and Tu, 2009; Morzfeld et al., 2012, 2015; Chorin et al., 2013), an importance distribution is constructed so that it has large values where the posterior is large. This is realized by computing the maximizer of $p(\boldsymbol{\theta}|\boldsymbol{D})$, i.e., the maximum a posteriori (MAP) of $\boldsymbol{\theta}$. If the prior and likelihood are exponential functions (as they often are in applications), the MAP can be found by minimizing the objective function

$$F(\boldsymbol{\theta}) = -\log(p(\boldsymbol{\theta})p(\boldsymbol{D}|\boldsymbol{\theta})). \tag{5}$$

Once the minimization problem is solved, one generates samples in the neighborhood of the minimizer $\boldsymbol{\mu} = \text{argmin}F$ as follows. A sequence of reference variables $\boldsymbol{\xi}_i$, $i = 1,\ldots,N$ are first sampled from a reference probability distribution $g(\boldsymbol{\xi})$, and subsequently each target posterior sample $\boldsymbol{\theta}_i$ is obtained by solving the equation

$$F(\boldsymbol{\theta}_i) - \phi = G(\boldsymbol{\xi}_i) - \gamma, \tag{6}$$

where $\phi = \min_{\boldsymbol{\theta}} F$, $G(\boldsymbol{\xi}) = -\log(g(\boldsymbol{\xi}))$ and $\gamma = \min_{\boldsymbol{\xi}} G$. The sample weights are

$$w_i = J(\boldsymbol{\theta}_i) \tag{7}$$

where $J$ is the Jacobian of the bijective map $\boldsymbol{\xi} \to \boldsymbol{\theta}$ (Morzfeld et al., 2015). Note that the sequence of samples $\boldsymbol{\theta}_i$ obtained by solving Eq. (6) are in the neighborhood of the MAP $\boldsymbol{\mu}$, since the right-hand side is small if $\boldsymbol{\xi}_i$'s are sampled close to the minimizer of $G$. Thus, Eq. (6) maps a likely $\boldsymbol{\xi}$ to a likely $\boldsymbol{\theta}$.

However, solving the mapping Eq. (6) is nontrivial, noting that it is in general nonlinear. One strategy, named "linear maps", is inspired by the approximation of $F$ by its second-order Taylor expansion around the MAP $\boldsymbol{\mu}$

$$F_0(\boldsymbol{\theta}) = \phi + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})^T H(\boldsymbol{\theta} - \boldsymbol{\mu}) \tag{8}$$

where $H$ is the Hessian matrix at $\boldsymbol{\mu}$. For an uncorrelated standard Gaussian reference variable, Eq. (6) is transformed to

$$F(\boldsymbol{\theta}_i) - \phi = \frac{1}{2}\boldsymbol{\xi}_i^T\boldsymbol{\xi}_i. \tag{9}$$

Substituting $F(\boldsymbol{\theta}_i)$ in Eq. (9) with $F_0(\boldsymbol{\theta}_i)$, the mapping now takes the simple form

$$\boldsymbol{\theta}_i = \boldsymbol{\mu} + L^{-T}\boldsymbol{\xi}_i \tag{10}$$

where $L$ is a lower triangular matrix obtained from the Cholesky decomposition of $H$. Accounting for the error of approximating the objective function $F$ by $F_0$, the weights of the samples are computed as

$$w_i \propto \exp(F_0(\boldsymbol{\theta}_i) - F(\boldsymbol{\theta}_i)). \tag{11}$$

The computational cost of implementing implicit sampling with linear maps includes minimizing the objective function Eq. (5) as well as evaluating the weights in Eq. (11), where one sample needs one forward model simulation. Morzfeld et al. (2012) described a "random maps" procedure in which an additional variable representing the search direction was introduced in Eq. (10). Despite its potential to further improve sample quality, a nonlinear equation needs to be solved for each sample, which incurs higher computational cost. In this work, we did not observe a clear improvement with random maps, and thus we only demonstrate the results for linear maps.

### 3.2. Sparse Bayesian learning based generalized polynomial chaos expansion

Polynomial chaos (PC) was originally developed to expand functions in terms of Hermite polynomial bases. Xiu and Karniadakis (2002) generalized the bases to the Askey family of orthogonal polynomials and named it generalized polynomial chaos expansion (gPCE). The gPCE for a second-order random process $f(\boldsymbol{\theta})$ takes the form

$$f(\boldsymbol{\theta}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d} f_{\boldsymbol{\alpha}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) \tag{12}$$

where $\boldsymbol{\theta}$ holds $d$-dimensional random parameters, $\boldsymbol{\alpha}$ is a $d$-dimensional nonnegative integer index tuple $\boldsymbol{\alpha} = (\alpha_1, \alpha_2,\ldots,\alpha_d)^T$, $\Phi_{\boldsymbol{\alpha}}$ are the multi-dimensional orthogonal polynomials whose individual one-dimensional components are of orders $\boldsymbol{\alpha}$, i.e., $\Phi_{\boldsymbol{\alpha}} = \phi_{\alpha_1}\phi_{\alpha_2}\ldots\phi_{\alpha_d}$, assuming the type of the polynomial $\phi$ for each dimension is identical, and $f_{\boldsymbol{\alpha}}$ are the PC coefficients corresponding to index $\boldsymbol{\alpha}$.

In the Askey family, for example, Legendre polynomials are used as the bases for uniform distributions, Hermite polynomials for Gaussian distributions, and Laguerre polynomials for Gamma distributions, to achieve the optimal convergence. In practice, the gPCE (Eq. (12)) is truncated to a finite number of terms, e.g., by specifying the total degree of the expansion, $P_{\boldsymbol{pc}}$, such that the dimensionality of $\boldsymbol{\alpha}$, $|\boldsymbol{\alpha}| = \alpha_1 + \alpha_2 + \cdots + \alpha_d$, is no larger than $P_{\boldsymbol{pc}}$. As a result, the total number of expansion terms $N_{\boldsymbol{pc}}$ is

$$N_{pc} = \frac{(d + P_{pc})!}{d!\,P_{pc}!}. \tag{13}$$

A number of techniques to evaluate the PC coefficients have been developed in the literature. For a review of some of the techniques, we refer the readers to Xiu (2010). However, these techniques are subject to the "curse of dimensionality", i.e., the number of gPCE terms $N_{\boldsymbol{pc}}$ grows rapidly as the number of parameters $d$ and the expansion order $P_{\boldsymbol{pc}}$ increase, resulting in the need for a much larger number of model simulations to construct the gPCE.

Recently Sargsyan et al. (2014) applied the sparse Bayesian learning (SBL) technique, developed by Tipping (2001) and Tipping and Faul (2003) and enhanced by Babacan et al. (2010), to the construction of gPCE. With SBL, the determination of the gPCE coefficients is treated as a regression problem, as the sparsity of the coefficients is taken into consideration. Therefore, the total number of coefficients to solve for is far fewer than the total number of gPCE coefficients, which alleviates the "curse of dimensionality" to a great extent if the model in question is inherently sparse.

To cast gPCE as an SBL problem, we solve the regression problem for the model output

$$\boldsymbol{f} = \Phi\boldsymbol{f}_{\boldsymbol{\alpha}} + \boldsymbol{\epsilon} \tag{14}$$

where $\boldsymbol{f} = (f(\boldsymbol{\theta}_1), f(\boldsymbol{\theta}_2),\ldots,f(\boldsymbol{\theta}_{N_o}))^T$ are the model outputs evaluated at $N_o$ sets of parameters, $\Phi$ is a matrix of size $N_o$ by $N_{\boldsymbol{pc}}$, with elements $\Phi_{i,\boldsymbol{\alpha}} = \Phi_{\boldsymbol{\alpha}}(\boldsymbol{\theta}_i)$, $\boldsymbol{f}_{\boldsymbol{\alpha}}$ is the vector of all the $N_{\boldsymbol{pc}}$ gPCE coefficients $f_{\boldsymbol{\alpha}}$, and $\boldsymbol{\epsilon}$ is the error vector. Therefore, the gPCE problem Eq. (12) is converted to the regression problem given by Eq. (14).

In SBL, the errors $\boldsymbol{\epsilon}$ are conventionally modeled as independent Gaussian variables with zero mean and variance $\sigma^2$, which is estimated iteratively. Hence, the Gaussian likelihood for the target $\boldsymbol{f}$ is

$$p(\boldsymbol{f}|\boldsymbol{f}_{\boldsymbol{\alpha}}, \sigma^2) = (2\pi)^{-N_o/2}\sigma^{-N_o}\exp\left(-\frac{\|\boldsymbol{f} - \Phi\boldsymbol{f}_{\boldsymbol{\alpha}}\|^2}{2\sigma^2}\right). \tag{15}$$

The coefficients $f_{\alpha}$ are hierarchically modeled as a Gaussian prior with zero means and variances $\gamma = \{\gamma_{\alpha}\}_{\alpha} = (\gamma_1, \gamma_2, \ldots, \gamma_{N_{pc}})$:

$$p(f_{\alpha}|\gamma) = (2\pi)^{-N_{pc}/2} \prod_{\alpha} \gamma_{\alpha}^{-1/2} \exp\left(-\frac{f_{\alpha}^2}{2\gamma_{\alpha}}\right). \tag{16}$$

Combining the likelihood and prior by Bayes' rule, the posterior distribution of $f_{\alpha}$ given $\gamma$

$$p(f_{\alpha}|f, \gamma, \sigma^2) = \frac{p(f|f_{\alpha}, \sigma^2)p(f_{\alpha}|\gamma)}{p(f|\gamma, \sigma^2)} \tag{17}$$

can be shown to be Gaussian, i.e., $f_{\alpha} \sim \mathcal{N}(\mu, \Sigma)$, where

$$\mu = \{\mu_{\alpha}\}_{\alpha} = \sigma^{-2}\Sigma\Phi^T f \tag{18}$$

and

$$\Sigma = (\Sigma_{\alpha,\alpha'}) = \left(\mathrm{diag}\left(\gamma_1^{-1}, \gamma_2^{-1}, \ldots, \gamma_{N_{pc}}^{-1}\right) + \sigma^{-2}\Phi^T\Phi\right)^{-1}. \tag{19}$$

The parameters $\gamma_1, \gamma_2, \ldots, \gamma_{N_{pc}}$ are found by maximizing the marginal likelihood

$$\mathcal{L}(\gamma) = \log p(f|\gamma, \sigma^2) = -\frac{1}{2}(N_o \log 2\pi + \log|C| + f^T C^{-1} f) \tag{20}$$

where $C = \sigma^2 I + \Phi\mathrm{diag}\left(\gamma_1, \gamma_2, \ldots, \gamma_{N_{pc}}\right)\Phi^T$.

The vector $\gamma$ controls the sparsity of $f_{\alpha}$. If $\gamma_{\alpha}$ is estimated to be zero, then the corresponding gPCE coefficient $f_{\alpha} = 0$. Tipping and Faul (2003) provided an efficient algorithm to locate a single $f_{\alpha}$ that causes the largest increase in $\mathcal{L}(\gamma)$ in each iteration, take the corresponding action (add, modify or delete), and derive a subset of $f_{\alpha}$ containing only the nonzero elements, until a prescribed stopping criterion is satisfied.

For any given set of parameters $\theta^*$, we can therefore predict the corresponding output with the ROM, which is a Gaussian process with mean

$$m(\theta^*) = \sum_{|\alpha| \le P_{pc}} \mu_{\alpha}\Phi_{\alpha}(\theta^*) \tag{21}$$

and covariance

$$K(\theta^*, \theta') = \sum_{|\alpha| \le P_{pc}, |\alpha'| \le P_{pc}} \Phi_{\alpha}(\theta^*)\Sigma_{\alpha,\alpha'}\Phi_{\alpha'}(\theta'), \tag{22}$$

where $\mu_{\alpha}$ and $\Sigma_{\alpha,\alpha'}$ are obtained in Eqs. (18) and (19). We hereafter denote the ROM based on SBL and gPCE, Eqs. (21) and (22), by SBLgPCE. The mean of SBLgPCE is used to predict the model output, and the variance $K(\theta^*, \theta^*)$ can provide an error estimate of the ROM.

### 3.3. General frameworks of coupling SBLgPCE with IS

In this section, we introduce three different approaches to use SBLgPCE within the implicit sampling algorithm. We give their detailed workflows and their pros and cons. Implicit sampling involves two main steps, namely computing the MAP and mapping samples from reference distribution to posterior distribution. To reduce the computational cost of both steps, one can construct the SBLgPCE for the forward model $f$ at the very beginning, and use SBLgPCE subsequently in both the optimization and mapping steps, which leads to the first approach:

*Approach 1.*

- Generate $N_t$ training input parameters $\theta_i$, $i = 1, \ldots, N_t$ from the prior parameter space and evaluate the forward model $f$ at the parameters to obtain $N_t$ sets of output values, $f_{obs}^i$, $i = 1, \ldots, N_t$, each $f_{obs}^i$ being the outputs at all the observation points.
- Construct gPCE for each observation point with a selected orthogonal polynomial base, e.g., Legendre polynomials, in the form of Eq. (12) and evaluate the bases $\Phi_{\alpha}(\theta)$. Use SBL to evaluate the coefficients $f_{\alpha}$ and the SBLgPCE ROM is obtained.
- Use the ROM to approximate the forward model in the optimization of Eq. (5) to obtain $\mu$, $\phi$ and $H$.
- For each reference sample, map the samples by Eq. (10) and use the ROM to predict the model outputs and eventually the objective function value $F$ to compute the sample weight with Eq. (11).

Approach 1 has the advantage that once the ROM is constructed, it can be used to enhance the efficiency of both steps in IS. However, the training input parameters for the ROM are generated from the prior parameter space, which is usually larger than the posterior space. As a result, a large number of training samples for the ROM may be needed to achieve an acceptable accuracy. In addition, optimizing the objective function in IS with a ROM may result in a different minimizer from that with the original model.

To address these issues, one can perform the optimization step in IS with the original model. The SBLgPCE ROM is then constructed in the vicinity of the MAP by approximating the posterior space using a Gaussian distribution centered around the MAP. This leads to the next two approaches, which differ in the output the ROM is constructed for. In Approach 2, the ROM represents the relationship between the input parameters and the objective function, whereas in Approach 3, it represents the relationship between the input parameters and the model outputs.

*Approach 2.*

- Perform the optimization step of IS with the original model to obtain $\mu$, $\phi$ and $H$.
- Generate $N_t$ training input parameters $\theta_i$, $i = 1, \ldots, N_t$ from the vicinity of the MAP $\mu$, and evaluate the forward model $f$ at the parameters to obtain the model outputs and eventually the $N_t$ objective function values $F_i$, $i = 1, \ldots, N_t$ from Eq. (5).
- Construct one single SBLgPCE for the input-output relationship between $\theta$ and $F$.
- For each reference sample, map the samples by Eq. (10) and use the ROM to directly predict objective function value $F$ to compute the sample weight with Eq. (11).

*Approach 3.*

- Perform the optimization step of IS with the original model to obtain $\mu$, $\phi$ and $H$.
- Generate $N_t$ training input parameters $\theta_i$, $i = 1, \ldots, N_t$ from the vicinity of the MAP $\mu$, and evaluate the forward model $f$ at the parameters to obtain $N_t$ sets of output values, $f_{obs}^i$, $i = 1, \ldots, N_t$, each $f_{obs}^i$ being the outputs at all the observation points.
- Construct SBLgPCE for each observation point to approximate the input-output relationship between $\theta$ and $f$.
- For each reference sample, map the samples by Eq. (10) and use the ROM to predict the model outputs and eventually the objective function value $F$ to compute the sample weight with Eq. (11).

Approach 2 has the advantage that only one single SBLgPCE needs to be constructed, since the output for the ROM is the objective function, which is a scalar, while in Approach 3 the total number of ROMs that need to be constructed is the number of observation points. The workflows of IS, SBLgPCE and the three approaches to coupling both are summarized in Fig. 4.

Note that embarrassing parallelism, i.e., little or no communication between parallel tasks, can be exploited in the frameworks discussed above. In all the approaches, the derivative information needed for the optimization process in IS is often estimated by finite difference approximations, which can be distributed as parallel tasks. In the cases of high-dimensional parameters, the parallelization brings even more computational gains. The mapping procedure in IS for each reference sample is also independent and can be distributed without any

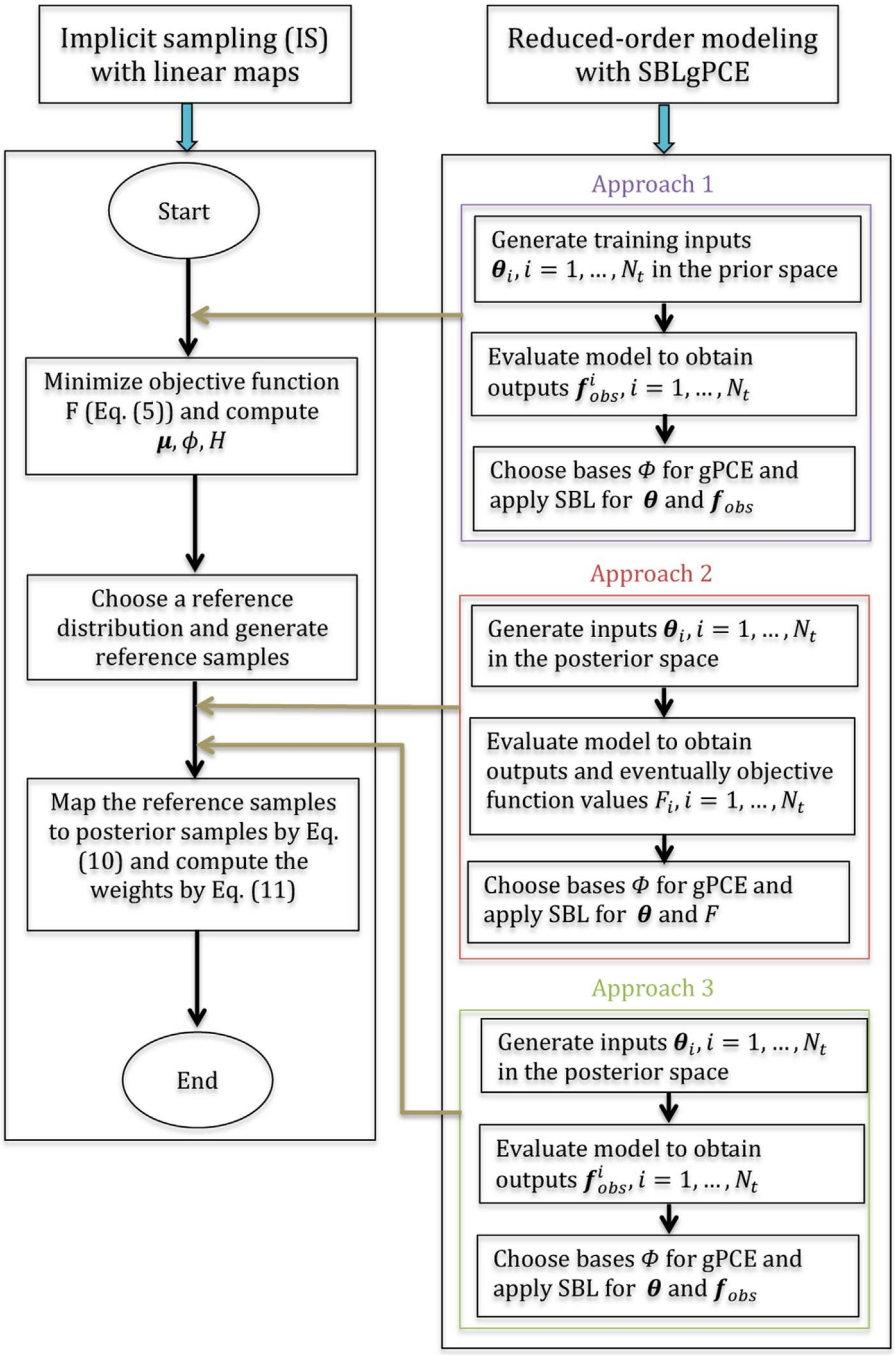


**Fig. 4.** The flowchart of IS, SBLgPCE and their coupling.

communications. In Approach 1 and 3, the SBLgPCE ROMs are constructed for the outputs at all the observation points. The computational cost for the ROM construction is considerable if the number of observation points is large. However, each ROM is independent and hence can be constructed in parallel as well.

## 4. Results and discussion

In this section, we first look at the results obtained from IS. We compare the results to those obtained using MCMC and then describe three different ways to build the ROM for implicit sampling. The ROM that gives the highest accuracy is identified. Finally, we show that the use of ROM with IS leads to an efficient estimate of the posterior distribution.

### 4.1. Implicit sampling

To minimize the objective function $F$ and obtain the MAP, the Levenberg-Marquardt algorithm is used since Eq. (5) is a least-squares minimization problem. The Jacobian matrix is numerically estimated by finite differences, and the Hessian matrix is then approximated from the Jacobian matrix. The optimization process costs 330 TOUGH2 simulations.

Once the MAP is obtained, posterior samples and weights are generated via Eqs. (10) and (11) in parallel from independent standard Gaussian reference variables. To determine the number of samples that can accurately represent the posterior, we examine the convergence behavior of the posterior densities for $N = 200$, 500, 1000, 2000, 5000 and 10000, where $N$ is the number of samples. Fig. 5 shows that the marginal posterior densities for 5000 and 10000 are indistinguishable;

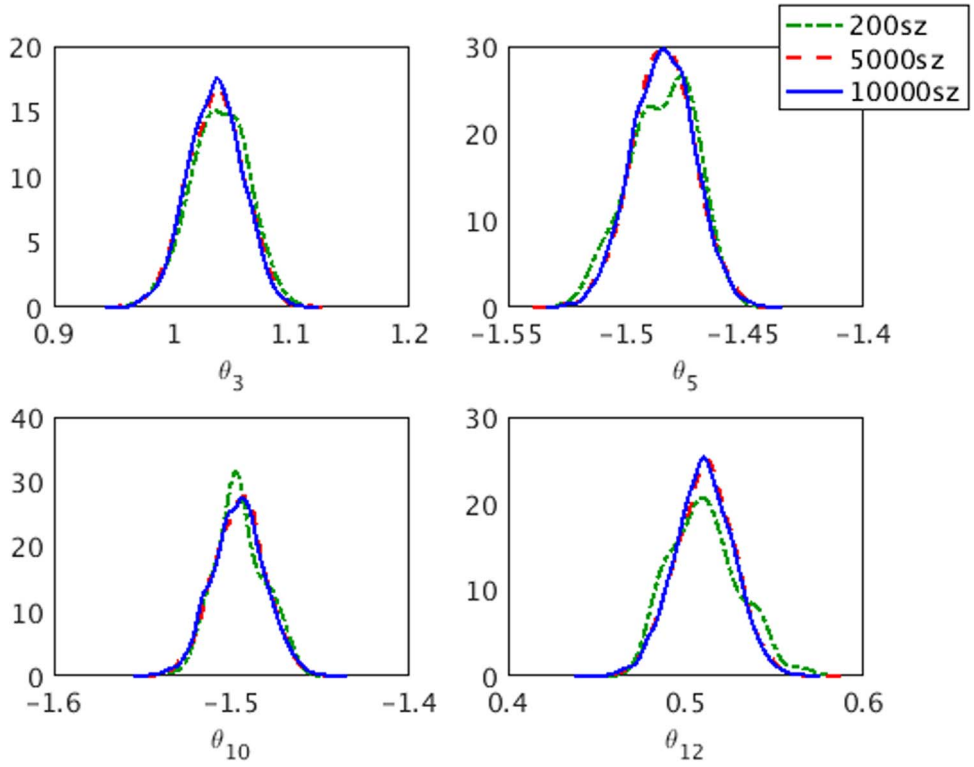


**Fig. 5.** Convergence behavior of marginal posteriors estimated from three sets of implicit samples of sizes $N$=200 (green dash-dot line), 5000 (red dashed line) and 10,000 (blue solid line) for four parameters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for clarity, we show results for four arbitrarily selected $\theta_i$ although all $\theta_i \in \boldsymbol{\theta}$ showed similar convergence behavior. A sample size of 200 is clearly insufficient since the marginal posterior densities for $N$=200 differ significantly from $N$=10000. We note that the marginal posterior densities from $N$=500, 1000, and 2000 have not converged as well but are left out from Fig. 5 for clarity. For the rest of the analyses, we will use $N$=5000.

Fig. 6 shows the marginal posteriors of all 14 parameters. The reference parameters lie within the high-probability regions of the marginal posterior densities. The ranges of the marginal posteriors are also smaller compared to the corresponding marginal prior distributions given in Table 1. Fig. 7 (left) plots the weighted mean of the log-permeability modifier fields corresponding to all the samples $\boldsymbol{\theta}$. In comparison to Fig. 1, it can be concluded that the log-permeability modifier field is accurately recovered. Fig. 7 (right) shows the standard deviations for the posterior log-permeability modifier fields. It is clear that $\theta_8$ (the top right log-permeability modifier in Fig. 1) has the highest uncertainty resulting from poor sensitivity to the observations. This is because the saturation plume does not pass through the vicinity of $\theta_8$'s location.

To measure the quality of the samples, we estimate the effective sample size by

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} (w_i)^2}. \tag{23}$$

The larger the estimated effective sample size is, the less severely the sample quality degenerates. In the ideal case where the weights are uniformly $1/N$, $N_{\text{eff}} = N$, meaning all the samples are effective, while in the worst scenario where only one sample is weighted nonzero, $N_{\text{eff}} = 1$. In our example, the ratios of the effective sample size (to the total number of samples) are all around 30% for all the samples sizes studied, which can be considered sufficient.

To further investigate the accuracy and efficiency of implicit sampling, we also implement MCMC sampling for the synthetic experiment as a comparison. The recently developed MCMC ensemble sampler (Goodman and Weare, 2010), is implemented in parallel on the Cori supercomputer system of National Energy Research Scientific Computing Center (NERSC) with the Python package emcee (Foreman-Mackey et al., 2013). The MCMC ensemble sampler consists of a number of "walkers" (similar to the Metropolis-Hastings chains but with dependence on each other) with affine invariance property. The method is inspired by and analogous to the Nelder-Mead simplex algorithm, which solves ill-conditioned optimization problems by evolving many copies of the system toward a local minimum. Due to the affine invariance, the ensemble sampler can effectively sample from highly anisotropic target distributions with negligible computational overhead.

We choose 126 walkers, and advance the sampler for 2000 steps, with a total number of $2.52 \times 10^5$ TOUGH2 forward model simulations. The mean acceptance rate of the MCMC sampling is 33% as provided by the emcee package, comparable to the ratio of effective sample size in implicit sampling. The ensemble sampler does not converge for all of the parameters since some still have high autocorrelations. Due to the computational cost, we only show the MCMC results for some of the converged parameters.

Fig. 8 compares the marginal posterior densities obtained from implicit sampling with 5000 sample size and the MCMC ensemble sampler for four of the parameters. The first 500 steps of the MCMC sampler are removed to eliminate the burn-in period, and 189,000 samples are obtained by combining the remaining 1500 steps and all of the walkers to plot the densities. The posteriors from implicit sampling (green solid line) show good agreement with those from MCMC (red dash-dot line). In this case, however, with 5000 samples, implicit sampling has converged for all parameters, while the MCMC ensemble sampler needs more steps to converge for all the parameters even with $O(10^5)$ simulations. Thus, implicit sampling requires <2% of the number of samples required by MCMC, even after considering the initial optimization step. In addition, all 5000 samples used in IS can be run in parallel, and thus the reduction in the actual wallclock time can be even more significant. For this example, it is clear that IS is more efficient than a MCMC ensemble sampler.

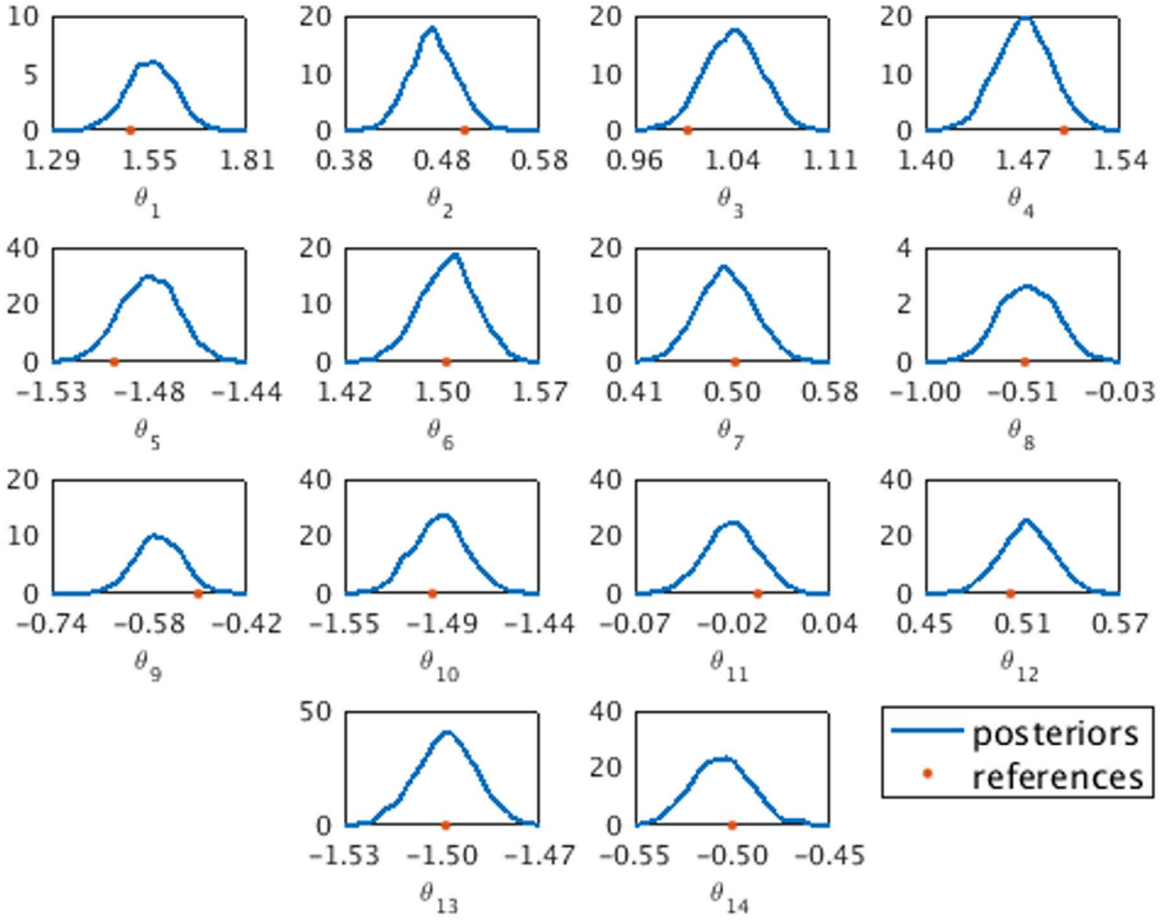


**Fig. 6.** Marginal posterior distributions (posteriors) obtained from implicit sampling with sample size 5000, and the reference values (references) for all 14 parameters.

### 4.2. *ROM for* f *trained in the prior parameter space*

First, we consider constructing an SBLgPCE ROM for the TOUGH2 model for each of the 333 observation points. The training data are sampled in the uniform parameter space $[-3, 3]^{14}$, which is two to three standard deviations away from the prior means and covers most of the prior parameter space. The goal is to use the resulting ROM in the initial optimization step to determine MAP and the subsequent sampling in the neighborhood of the MAP. We denote the resulting ROM by ROM1. We train ROM1 using 1200 samples and validate it for a different set of 400 samples; however, they fail to accurately approximate the TOUGH2 model, as reported in Liu et al. (2015b), where the relative root mean square errors are shown to range from 1.8% to 33.8%. The reason is that the input parameter space is so big that more training samples are needed to construct the ROM. The large errors in the prediction of the model output at the observation point are further amplified by the exponential function in the estimation of objective function $F$ in Eq. (5) and eventually lead to inaccurate weights through Eq. (11).

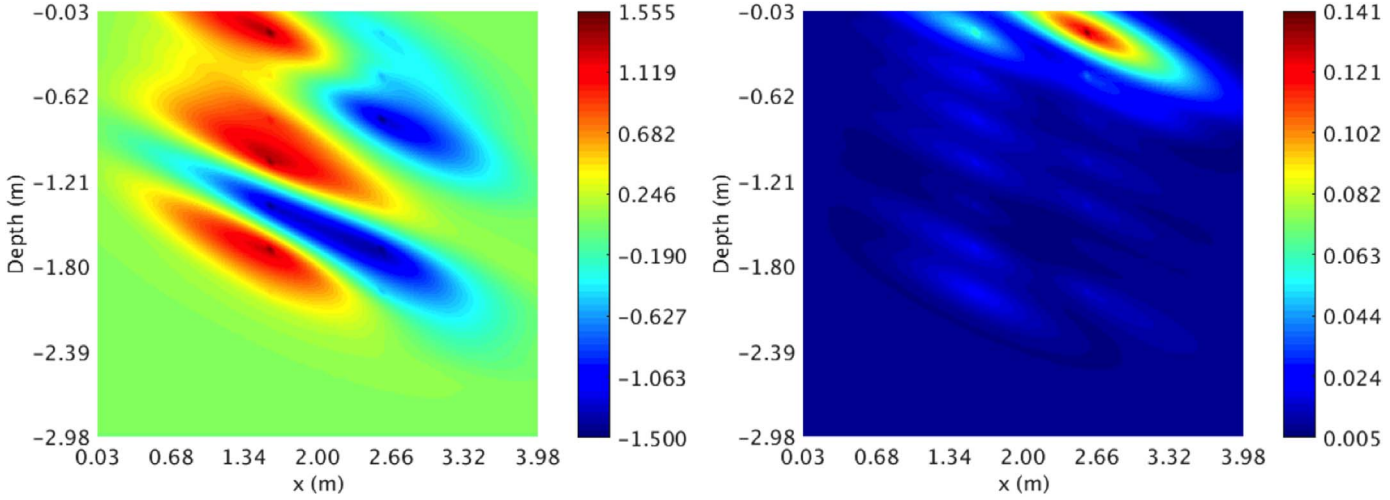


**Fig. 7.** The mean (left) and standard deviation (right) of the log-permeability modifier fields corresponding to the posterior distribution obtained from implicit sampling with sample size 5000.

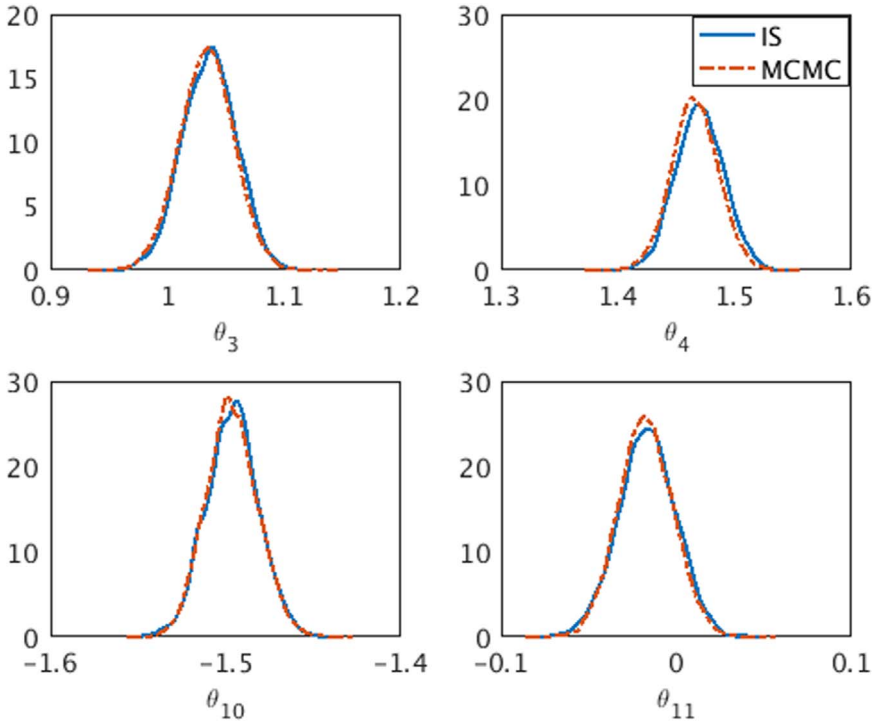


**Fig. 8.** Comparison of the marginal posterior densities obtained from 5000 implicit samples (solid line, IS) and those from the MCMC ensemble sampler with 126 walkers each having 1500 samples (dash-dot line, MCMC). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

### 4.3. ROM for F, with samples drawn from the posterior parameter space

Next, we consider performing reduced-order modeling with SBLgPCE directly for the objective function $F$, after MAP is obtained. The training samples are generated within three standard deviations of MAP, estimated from the diagonal of the Hessian matrix, to cover the majority of the posterior density. We call this model ROM2. Since the training samples used to construct ROM2 are generated in the neighborhood of MAP, it can achieve higher accuracy than ROM1 for the purpose of sampling near MAP. In addition, only a single ROM needs to be constructed as opposed to 333 independent ROMs for ROM1.

We train the ROM with a set of 1000 training data that are obtained from running the full model (TOUGH2 simulations), and validate the ROM for the same 5000 samples used in Section 4.1. The total degree of the polynomial expansion is set to be $P_{pc} = 5$, resulting in $N_{pc} = 11628$ polynomial terms. The stopping criteria for the SBL is that the increase of the marginal likelihood Eq. (20) is smaller than $10^{-5}$. Fig. 9 shows the histogram, as well as the box-and-whisker plot of the relative errors between the true objective function values $F$ (evaluated by TOUGH2) and the ROM2-predicted values $F_{\mathrm{ROM2}}$ for the 5000 validation cases. The errors range from 3.83e − 7 to 0.13 with a mean and median of 4.6e − 3 and 2.9e − 3, respectively, showing a significant improvement in the accuracy compared to ROM1. We will examine the performance of using ROM2 in the characterization of posterior densities with implicit sampling in Section 4.5.

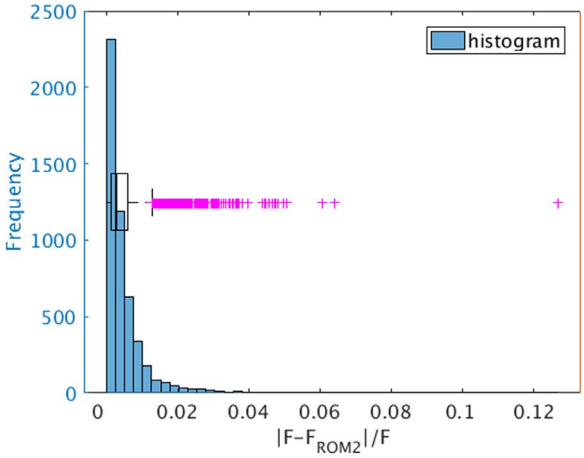


**Fig. 9.** The histogram and horizontal box-and-whisker plot of the relative errors between the objective function values computed by TOUGH2 (F) and those computed by ROM2 ($F_{\mathrm{ROM2}}$). The "+" markers are outliers.

### 4.4. ROM for f, with samples drawn from the posterior parameter space

Finally, we construct a ROM for the TOUGH2 model at each of the observation points (similar to ROM1), but in the posterior parameter space (similar to ROM2); we name it ROM3. The training and validation samples of ROM3 are the same as those used in Section 4.3. The total degree of the polynomial expansion is again chosen as $P_{pc} = 5$, and thus $N_{pc} = 11628$. For traditional gPCE methods, a data set containing at least $N_{pc}$ training samples, which is far greater than 1000 used in this example, is needed for an accurate estimation of the gPCE coefficients. Thus, the capability of SBL to detect the sparse structure of the gPCE coefficients allows fewer training samples to be used.

Using the same stopping criteria for SBL as in ROM2, we explore in Fig. 10 the sparsity of the gPCE's coefficients for ROM3 for the saturation value at each of the observation point. The sparsity is defined as

$$\text{sparsity} = \frac{\sum_{\alpha} \mathbf{1}_{\{\gamma_\alpha = 0\}}}{N_{pc}} = 1 - \frac{\sum_{\alpha} \mathbf{1}_{\{\gamma_\alpha \neq 0\}}}{N_{pc}} \tag{24}$$

where the notation $\mathbf{1}_{\{\cdot\}}$ represents the indicator function. In other words, sparsity is the ratio of the number of zero gPCE coefficients to the total number of coefficients. The left y-axis in Fig. 10 displays the sparsity, sorted in ascending order, for all the individual SBLgPCE of

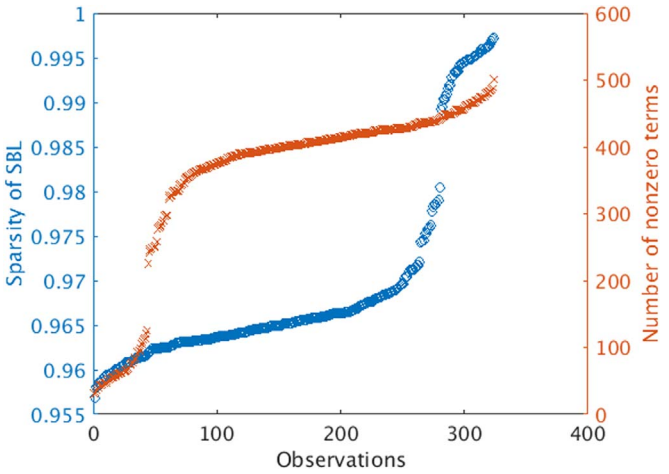


**Fig. 10.** Sparsity (left y-axis) and the number of nonzero gPCE coefficients (right y-axis) of ROM3, sorted in ascending order.

ROM3. The sparsity varies from 0.957 to 0.997, indicating at least 95% of the gPCE coefficients are zeros for all cases. The right y-axis plots the corresponding numbers of nonzero gPCE terms, which range from only 32–502.

Fig. 11 (left) shows the mean relative errors between TOUGH2- and ROM3-simulated saturation values for the 5000 validation cases at all observation points and times. The mean errors range from $O(10^{-4})$ to $O(10^{-3})$, with the average value over all the observations being 0.034%. The right figure shows the maximum errors, which are one order of magnitude larger than the mean errors. The accurate prediction of saturation at the observation points provides a solid basis for the accurate approximation of the objective function values, and by extension, the sample weights.

Based on the predictions of saturation by ROM3, we compute, according to Eq. (5), the objective function values, which subsequently determine the weights of the implicit samples. Fig. 12 displays the histogram of the relative errors between the objective function values computed by TOUGH2 (F) and ROM3 ($F_{\mathrm{ROM3}}$). The relative errors range from 1.90e − 7 to 4.50e − 2, with a mean of 2.2e − 3 and 1.2e − 3 as the median. In comparison to Fig. 9, the maximum relative error is significantly reduced, and the number of cases where the relative errors are >1% is reduced from 1041 to 289. For ROM3, although 333 SBLgPCEs are needed instead of just a single one, the ROMs are independent and are constructed in an embarrassingly parallel way. The enhanced accuracy of ROM3 results in a more accurate characterization of the posterior as shown next.

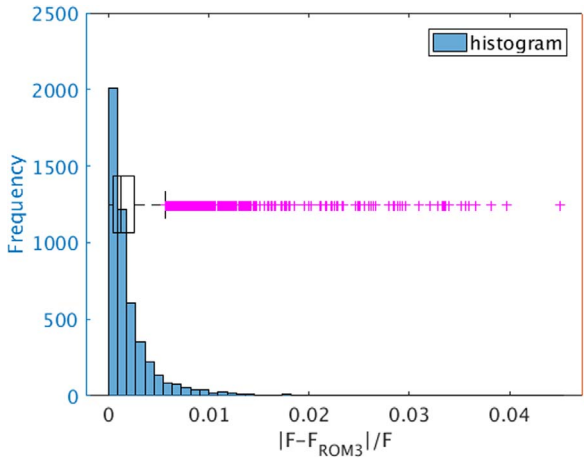


**Fig. 12.** Histogram and horizontal box-and-whisker plot of the relative errors between the objective function values computed by TOUGH2 (F) and those by ROM3 ($F_{\mathrm{ROM3}}$). The "+" markers are the outliers.

### 4.5. Incorporating the ROM in IS

With the predicted objective function values from ROM2 and ROM3, we are able to compute the sample weights according to Eq. (11) and perform the resampling procedure to obtain the posterior distributions. Fig. 13 demonstrates the marginal posteriors of the same four parameters as shown in Fig. 8, obtained from the full TOUGH2 model, ROM2 and ROM3. The marginal posteriors from ROM2 are in good agreement with those from the full model and they almost overlap each other, indicating that ROM3 is capable of accurately characterizing the posterior density in implicit sampling. On the other hand, ROM2 fails to match the inversion results obtained from the full model; the distinction between the marginal posteriors are clear, especially for $\theta_3$ and $\theta_4$.

The reason for the failure of ROM2 is that the few predicted $F$ with relatively large errors (e.g., those greater than 5%) can cause the normalized sample weights to be completely different from the true weights. The possible reason for the insufficient accuracy of ROM2 is that the nonlinearity of $F$ is added to an already nonlinear forward model, making the ROM more difficult to learn the input-output relationship. This behavior was also observed in Zeng et al. (2016).

We measure the computational efficiency of the proposed procedure by the number of forward simulations. The computational cost of forward simulations is typically significantly higher than the computational cost of determining the gPCE coefficients. In addition, ROM predictions based on gPCE have negligible computational cost. For $N$=5000, the use of ROM3 has reduced the number of forward

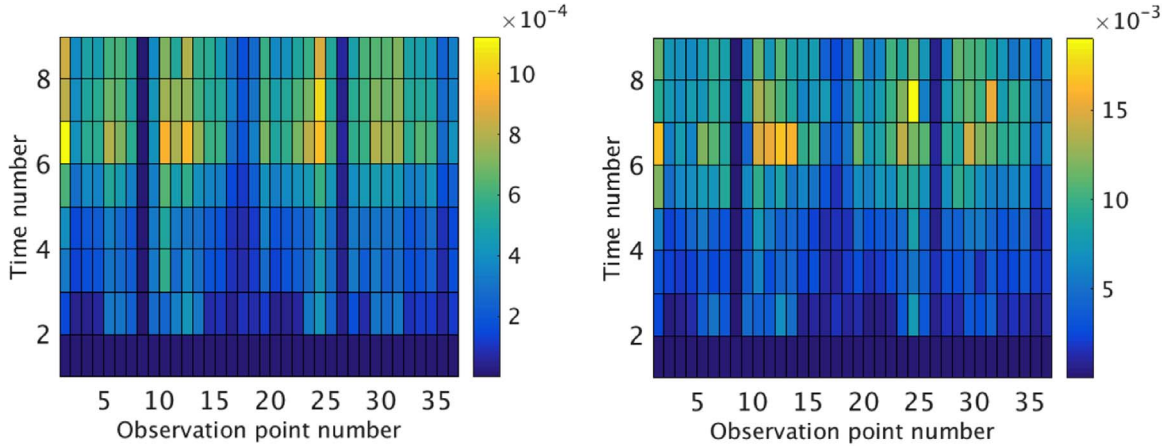


**Fig. 11.** The mean (left) and maximum (right) relative saturation errors between TOUGH2 and ROM3 for the 5000 validation cases at all observation points ($x$-axis) and times ($y$-axis).

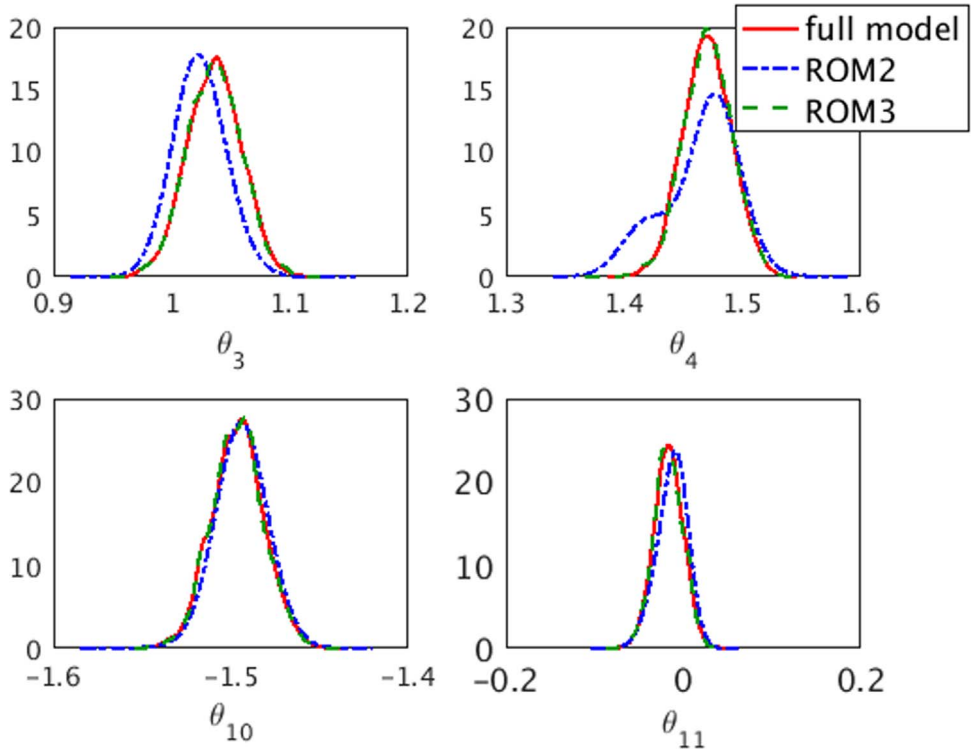


**Fig. 13.** Comparison of the marginal posteriors of four parameters obtained with implicit sampling using the TOUGH2 full model, ROM2 and ROM3.

simulations required by about 80%. Clearly, the reduction can be greater if $N$ is larger. ROM is also useful when we need to rapidly determine the appropriate size of $N$.

## 5. Conclusions

In this work, we implemented an implicit sampling algorithm for performing Bayesian inverse modeling to determine the permeability distribution of an aquifer based on a synthetic vadose zone hydrological experiment simulated with TOUGH2. Implicit sampling improves sample quality by sampling from the high probability region in the neighborhood of the maximum a posterior and alleviates sample collapse. Compared to Markov chain Monte Carlo methods, the number of forward simulations required is significantly reduced. While the effective sample size of implicit sampling is similar to the acceptance rate of MCMC, it does not require many acceptance-rejection steps, and is embarrassingly parallel once maximum a posterior estimates are obtained. We also demonstrated that ROM improves the efficiency of implicit sampling. In particular, we showed that the sparse Bayesian learning based generalized polynomial chaos expansion, constructed around the MAP for each observation, serves as an accurate and reliable reduced order model that leads to posterior densities that are indistinguishable from that obtained using the full model. In the future, we will test our proposed method for hydrological problems that have significantly more complex posterior distributions.

## Acknowledgment

Pau and Liu were supported by the Director, Office of Science, Office of Biological and Environmental Research of the US Department of Energy under Contract no. DEAC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under the same contract.

## References

Abubakar, A., Habashy, T.M., Li, M., Liu, J., 2009. Inversion algorithms for large-scale geophysical electromagnetic measurements. Inverse Probl. 25 (12), 123012, (⟨http://stacks.iop.org/0266-5611/25/i=12/a=123012⟩).

Andrieu, C., de Freitas, N., Doucet, A., Jordan, M.I., 2003. An introduction to mcmc for machine learning. Mach. Learn. 50 (1), 5–43. http://dx.doi.org/10.1023/A:1020281327116.

Arulampalam, M., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. Trans. Signal. Proc. 50 (2), 174–188. http://dx.doi.org/10.1109/78.978374.

Atkins, E., Morzfeld, M., Chorin, A.J., 2013. Implicit particle methods and their connection with variational data assimilation. Mon. Weather Rev. 141 (6), 1786–1803. http://dx.doi.org/10.1175/MWR-D-12-00145.1.

Babacan, S., Molina, R., Katsaggelos, A., 2010. Bayesian compressive sensing using laplace priors. IEEE Trans. Image Process. 19 (1), 53–63. http://dx.doi.org/10.1109/TIP.2009.2032894.

Carrera, J., Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. Water Resour. Res. 22 (2), 199–210. http://dx.doi.org/10.1029/WR022i002p00199.

Chorin, A.J., Morzfeld, M., Tu, X., 2010. Implicit particle filters for data assimilation. Commun. Appl. Math. Comput. Sci. 5 (2), 221–240, (⟨http://msp.org/camcos/2010/5-2/p03.xhtml⟩).

Chorin, Alexandre J., Morzfeld, Matthias, Tu, Xuemin, 2013. A survey of implicit particle filters for data assimilation. In: Zeng, Yong, Wu, Shu (Eds.), State-Space Models: Applications in Economics and Finance. Springer New York, New York, NY, 63–88 . http://dx.doi.org/10.1007/978-1-4614-7789-1_3, isbn 978-1-4614-7789-1http://dx.doi.org/10.1007/978-1-4614-7789-1_3.

Chorin, A.J., Tu, X., 2009. Implicit sampling for particle filters. Proc. Natl. Acad. Sci. 106 (41), 17249–17254. http://dx.doi.org/10.1073/pnas.0909196106.

Deutsch, C.V., Journel, A.G., 1997. Geostatistical Software Library and User's Guide second edition. Oxford University Press, New York, USA.

Finsterle, S., Kowalsky, M.B., 2008. Joint hydrological-geophysical inversion for soil structure identification. Vadose Zone J. 7 (1), 287–293.

Finsterle, S., Najita, J., 1998. Robust estimation of hydrogeologic model parameters. Water Resour. Res. 34 (11), 2939–2947. http://dx.doi.org/10.1029/98WR02174.

Foreman-Mackey, D., Hogg, D.W., Lang, D., Goodman, J., 2013. Emcee: the mcmc hammer. Publ. Astron. Soc. Pac. 125 (925), 306–312, (⟨http://www.jstor.org/stable/10.1086/670067⟩).

Gómez-Hernánez, J.J., Sahuquillo, A., Capilla, J., 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data – i. Theory. J. Hydrol. 203 (1–4), 162–174. http://dx.doi.org/10.1016/S0022-1694(97)00098-X.

Goodman, J., Weare, J., 2010. Ensemble samplers with affine invariance. Commun. Appl. Math. Comput. Sci. 5 (1), 65–80.

Kabanikhin, S., 2008. Definitions and examples of inverse and ill-posed problems. J. Inverse Ill-Posed Probl. 16 (4), 317–357.
Liu, J.S., Chen, R., 1998. Sequential monte carlo methods for dynamic systems. J. Am. Stat. Assoc. 93 (443), 1032–1044. http://dx.doi.org/10.1080/01621459.1998.10473765.
Liu, Y., Bisht, G., Subin, Z.M., Riley, W.J., Pau, G.S.H., 2016a. A hybrid reduced-order model of fine-resolution hydrologic simulations at a polygonal Tundra site. Vadose Zone J. 15, 2.
Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resour. Res. 43 (7), w07401. http://dx.doi.org/10.1029/2006WR005756, (n/a-n/a).
Liu, Y., Hussaini, M.Y., Ökten, G., 2016b. Accurate construction of high dimensional model representation with applications to uncertainty quantification. Reliab. Eng. Syst. Saf. 152, 281–295. http://dx.doi.org/10.1016/j.ress.2016.03.021.
Liu, Y., Jimenez, E., Hussaini, M.Y., Ökten, G., Goodrick, S., 2015a. Parametric uncertainty quantification in the Rothermel model with randomised quasi-Monte Carlo methods. Int. J. Wildland Fire 24 (3), 307–316.
Liu, Y., Pau, G.S.H., Finsterle, S., 2015b. Bayesian parameter inversion with implicit sampling for a vadose zone hydrological model. In: Blanco-Martin, L., Doughty, C., Finsterle, S., Reagan, M., Rutqvist, J., Valladao, C., Zheng, L., (Eds.), TOUGH Symposium 2015. Lawrence Berkeley National Laboratory, Berkeley, California. ⟨http://esd1.lbl.gov/files/research/projects/tough/events/symposia/toughsymposium15/Proceedings_TOUGHSymposium2015.pdf⟩.
Lu, Z., Zhang, D., 2003. On importance sampling monte carlo approach to uncertainty analysis for flow and transport in porous media. Adv. Water Resour. 26 (11), 1177–1188, (⟨http://www.sciencedirect.com/science/article/pii/S0309170803001064⟩).
Martin, J., Wilcox, L.C., Burstedde, C., Ghattas, O., 2012. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. SIAM J. Sci. Comput. 34 (3), A1460–A1487. http://dx.doi.org/10.1137/110845598.
Mondal, A., Efendiev, Y., Mallick, B., Datta-Gupta, A., 2010. Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump markov chain Monte Carlo methods. Adv. Water Resour. 33 (3), 241–256. http://dx.doi.org/10.1016/j.advwatres.2009.10.010.
Morzfeld, M., Chorin, A.J., 2012. Implicit particle filtering for models with partial noise, and an application to geomagnetic data assimilation. Nonlinear Process. Geophys. 19 (3), 365–382. http://dx.doi.org/10.5194/npg-19-365-2012.
Morzfeld, M., Tu, X., Atkins, E., Chorin, A.J., 2012. A random map implementation of implicit filters. J. Comput. Phys. 231 (4), 2049–2066. http://dx.doi.org/10.1016/j.jcp.2011.11.022.
Morzfeld, M., Tu, X., Wilkening, J., Chorin, A.J., 2015. Parameter estimation by implicit sampling. Commun. Appl. Math. Comput. Sci. 10 (2), 205–225, (⟨http://msp.org/camcos/2015/10-2/p04.xhtml⟩).
Neto, F.D.M., da Silva Neto, A.J., 2012. An Introduction to Inverse Problems with Applications. Springer Publishing Company, Incorporated.
Pau, G.S.H., Zhang, Y., Finsterle, S., Wainwright, H., Birkholzer, J., 2014. Reduced order modeling in itough2. Comput. Geosci. 65, 118–126. http://dx.doi.org/10.1016/j.cageo.2013.08.008.
Pruess, K., Oldenburg, C., Moridis, G., 2012. Tough2 user's guide, version 2.1. Report lbnl–43134, Lawrence Berkeley National Laboratory. ⟨http://esd1.lbl.gov/files/research/projects/tough/documentation/TOUGH2_V2_Users_Guide.pdf⟩).
Rabitz, H., Aliş, Ö.F., Shorter, J., Shim, K., 1999. Efficient input-output model representations. Comput. Phys. Commun. 117, 11–20.
RamaRao, B.S., LaVenue, A.M., de Marsily, G., Marietta, M.G., 1995. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments. Water Resour. Res. 31 (3), 475–493. http://dx.doi.org/10.1029/94WR02258.
Ramm, A.G., 2005. Inverse Problems: Mathematical and Analytical Techniques with Applications to Engineering. Mathematical and Analytical Techniques with Applications to Engineering. Springer, New York. ⟨http://opac.inria.fr/record=b1101679⟩.
Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48 (7), w07401.
Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. Physics 1 (5), 318. http://dx.doi.org/10.1063/1.1745010.
Sargsyan, K., Safta, C., Najm, H.N., Debusschere, B.J., Ricciuto, D., Thornton, P., 2014. Dimensionality reduction for complex models via Bayesian compressive sensing. Int. J. Uncertain. Quantif. 4 (1), 63–93. http://dx.doi.org/10.1615/Int.J.UncertaintyQuantification.2013006821.
Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-gaussian errors. Water Resour. Res. 46, 10. http://dx.doi.org/10.1029/2009WR008933, (n/a–n/a).
Tarantola, A., 2004. Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res., 211–244. http://dx.doi.org/10.1162/15324430152748236.
Tipping, M.E., Faul, A.C., 2003. Fast marginal likelihood maximisation for sparse Bayesian models. In: Bishop, C.M., Frey, B.J. (Eds.), Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL. ⟨http://www.miketipping.com/papers/met-fastsbl.pdf⟩.
van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils1. Soil Sci. Soc. Am. J. 44 (5), 892. http://dx.doi.org/10.2136/sssaj1980.03615995004400050002x.
Vogel, C.R., 2002. Computational Methods for Inverse Problems. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int. J. Nonlinear Sci. Numer. Simul. 10 (3), 273–290.
Xiu, D., 2010. Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton University Press, Princeton, N.J, (⟨http://opac.inria.fr/record=b1131521⟩).
Xiu, D., Karniadakis, G.E., 2002. The wiener-askey polynomial chaos for stochastic differential equations. SIAM J. Sci. Comput. 24 (2), 619–644. http://dx.doi.org/10.1137/S1064827501387826.
Zeng, X., Ye, M., Burkardt, J., Wu, J., Wang, D., Zhu, X., 2016. Evaluating two sparse grid surrogates and two adaptation criteria for groundwater Bayesian uncertainty quantification. J. Hydrol. 535, 120–134, (⟨http://www.sciencedirect.com/science/article/pii/S0022169416000767⟩).