Advances in Water Resources 86 (2015) 385–399

Contents lists available at ScienceDirect

Advances in Water Resources

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

Surrogate accelerated sampling of reservoir models with complex structures using sparse polynomial chaos expansion

Hamid Bazargan*, Mike Christie, Ahmed H. Elsheikh, Mohammad Ahmadi

Institute of Petroleum Engineering, Heriot–Watt University, Edinburgh, United Kingdom

ARTICLE INFO

Article history:
Available online 19 October 2015

Keywords:
Bayesian parameter estimation
Subsurface flow model
Polynomial chaos
MCMC

ABSTRACT

Markov Chain Monte Carlo (MCMC) methods are often used to probe the posterior probability distribution in inverse problems. This allows for computation of estimates of uncertain system responses conditioned on given observational data by means of approximate integration. However, MCMC methods suffer from the computational complexities in the case of expensive models as in the case of subsurface flow models. Hence, it is of great interest to develop alterative efficient methods utilizing emulators, that are cheap to evaluate, in order to replace the full physics simulator. In the current work, we develop a technique based on sparse response surfaces to represent the flow response within a subsurface reservoir and thus enable efficient exploration of the posterior probability density function and the conditional expectations given the data.

Polynomial Chaos Expansion (PCE) is a powerful tool to quantify uncertainty in dynamical systems when there is probabilistic uncertainty in the system parameters. In the context of subsurface flow model, it has been shown to be more accurate and efficient compared with traditional experimental design (ED). PCEs have a significant advantage over other response surfaces as the convergence to the true probability distribution when the order of the PCE is increased can be proved for the random variables with finite variances. However, the major drawback of PCE is related to the curse of dimensionality as the number of terms to be estimated grows drastically with the number of the input random variables. This renders the computational cost of classical PCE schemes unaffordable for reservoir simulation purposes when the deterministic finite element model is expensive to evaluate. To address this issue, we propose the reduced-terms polynomial chaos representation which uses an impact factor to only retain the most relevant terms of the PCE decomposition. Accordingly, the reduced-terms polynomial chaos proxy can be used as the pseudo-simulator for efficient sampling of the probability density function of the uncertain variables.

The reduced-terms PCE is evaluated on a two dimensional subsurface flow model with fluvial channels to demonstrate that with a few hundred trial runs of the actual reservoir simulator, it is feasible to construct a polynomial chaos proxy which accurately approximates the posterior distribution of the high permeability zones, in an analytical form. We show that the proxy precision improves with increasing the order of PCE and corresponding increase of the number of initial runs used to estimate the PCE coefficient.

1. Introduction

Traditional calibration methods of subsurface reservoirs (aka. history matching) usually obtain only a single set of parameters of the model, with uncertainty assessment provided by sensitivity calculations around the matched model. However, for example in oil field management, modern techniques have focused on predicting the likely range of field recoveries and consequently providing economic evaluations of different field development strategies. This approach takes into account observational errors in the observed history of the reservoir, and retrieves a set of model parameters whose simulation results lie within the vicinity of observed data, and uses them to estimate ranges in likely recovery factors. Generally, two main approaches for subsurface model calibration exist in the literature, one based on the optimization methods and the other based on the Bayesian inference.

The optimization methods adjust the unknown parameter values through an automated process to obtain reservoir models within the allowed range of a misfit function. Various optimization techniques have been developed in the literature, including Genetic Algorithms

* Corresponding author. Tel.: +989126079582.
E-mail address: hamidbazargan@gmail.com, hamid.bazargan@pet.hw.ac.uk (H. Bazargan).

http://dx.doi.org/10.1016/j.advwatres.2015.09.009
0309-1708/© 2015 Elsevier Ltd. All rights reserved.

[1], Particle Swarm Optimization [2], Neighborhood Algorithm [4], Estimation of Distribution [5], Levenberg–Marquardt [6] and LBFGS [3]. Existing optimization methods can be roughly classified into two general categories: stochastic algorithms and gradient-based methods. Gradient-based algorithms have several inherent limitations, including the need to compute the gradients at each step of the optimization process. A definite advantage of stochastic algorithms is that they are able to easily honor complex geological constraints by preserving multipoint statistics present in the prior geological model; the main drawback of these approaches is their inefficiency, as they require large number of simulations for convergence [9,10]. However, most of these optimization-based algorithms do not provide any statistically valid estimate for the parameters uncertainty without additional calculations. For example, Genetic Algorithms [1] and Particle Swarm Optimization [2] do not correspond to a valid sampling mechanism. The reason for this is that the distribution of parameter values is mainly controlled by the algorithm settings [11]. This needs to be corrected by running a second code to compute probabilities associated with each set of parameters [4]. A review of recent research activities on subsurface flow model calibration can be found in [8,59].

Approaches based on the Bayesian inference, on the other hand, aim at estimating the posterior probability for the reservoir properties [12]. Existing Bayesian inference methods broadly entails algorithms based on particle filters such as the Ensemble Kalman Filter (EnKF) [13,14], the sequential Monte Carlo methods [7] and the Markov Chain Monte Carlo (MCMC) approaches [15,16]. MCMC methods are often used to probe the posterior probability distribution in the Bayesian inference inverse problems. Many MCMC methods move toward the target distribution in relatively small steps, with no tendency for the steps to proceed in the same direction [17]. Among these methods are the Gibbs sampling method, the Metropolis–Hasting algorithms and the slice sampling algorithm [17]. These methods are easy to implement and analyze, but unfortunately it can take a long time for the random walker to explore all of the space. The walker will often double back and cover ground already covered [17]. The difficult problem is to determine how many steps are needed to converge to the stationary distribution within an acceptable error. A good chain will have a rapid mixing at which the stationary distribution is reached quickly starting from an arbitrary position. Variants of MCMC techniques have been developed in the literature to increase the convergence rate to the target distribution, but they are usually hard to implement [17]. Among these methods are Langevin MCMC [18], Hamiltonian Monte Carlo [19] and combinations of evolutionary algorithms with MCMC [20,21].

Oliver et al. [15] utilized MCMC in the context of reservoir simulation where MCMC methods were used for conditioning a permeability field to pressure data. Efendiev et al. [66] proposed to use a two-stage approach MCMC for conditioning permeability fields. More recently, Emerick and Reynolds [22] proposed to use MCMC to improve the sampling obtained by the ENKF method. However, typical uses of MCMC methods need more than 10^5 steps to sample from the target distribution with a reasonable error [23]. For subsurface reservoir studies, the required large number of simulations need for convergence is practically infeasible. Hence, the main disadvantage of these approaches is their inefficiency. Therefore, it can be extremely time-consuming if high resolution models are used. This is particularly of concern in closed-loop reservoir management, which requires continuous real-time use of model calibration and uncertainty quantification algorithms [24,25]. Thus, there is a significant need for an efficient proxy (or surrogate) model that can predict simulation results with a reasonable accuracy.

The choice of the polynomial chaos expansions as a proxy model was pioneered by Ghanem and Spanos [26] and has been applied to various engineering problems [27–32,68]. The polynomial chaos proxy has an advantage over all other surrogate models that it systematically guarantees the convergence in probability and also in distribution to the output random variable of interest with finite variance, i.e. cumulative oil production. However, for high-dimensional problems, the number of the polynomial chaos terms increases drastically as the order of the polynomial chaos expansions increases and a large number of the full reservoir simulation runs may be required to compute high-order polynomial chaos expansions. Hence, for the efficient use of the polynomial chaos proxy, the size of the problem has to be effectively reduced.

Dimensionality reduction techniques, which have been applied in many application areas including reservoir simulation, represent promising means for constructing efficient surrogate models. Many of these techniques entail the projection of the high resolution description of reservoir into a low-dimensional subspace, which significantly reduces the number of unknowns. Karhunen–Loeve representation was first introduced by Chen et al. [33] within a reservoir engineering context to reduce the dimension of geological parameters. The basic approach was later used by, among others, Oliver [34], Reynolds et al. [35], and Sarma [25] to approximate the high resolution geological model with a much lower dimensional space. Marzouk and Najm [50] applied the polynomial chaos expansion along with the MCMC algorithm in the Bayesian framework for a low-dimensional problem and proved the efficiency of the algorithm. However, for the practical implementation of the polynomial chaos proxy in the Bayesian framework, the number of the terms is still required to be further reduced. More recently, efforts have been made to construct solution-adaptive uncertainty propagation techniques that exploit any structures in the solution to decrease the computational cost. Among them are the multi-scale model reduction of [31] and the sparse decomposition of [51–53] for the stochastic Galerkin technique, anisotropic and adaptive sparse grids of [46,54,55] for the stochastic collocation scheme, and low-rank solution approximations of [56–58,69].

More recently in [49], the correlation between samples was used to justify the use of sparse promoting regularization (i.e. constraining the ℓ_1 norm) to generate sparse PCE representation of PDEs. Similarly, [61] studied a high-dimensional stochastic collocation method where the polynomial coefficients were obtained by solving a constrained minimization problem with ℓ_1 regularization term. Sparse promoting regularization can be iteratively solved using the orthogonal matching pursuit (OMP) [60,62] as used in [49] or using the least angle regression (LARS) algorithm [63] as used in [64].

In this work, we propose a heuristic method for the sparse representation of polynomial chaos expansion and its application as a proxy substitute for the full reservoir simulator when applied with the MCMC method to efficiently sample from the posterior probability density function of reservoir random parameters. We use the Karhunen–Loeve expansion to decompose the geological parameters into a lower dimension. The polynomial chaos proxy is then trained with the reduced-order parameters of the reservoir model. The Bayesian inference provides a mathematical formulation for the posterior distribution of reservoir parameters. Instead of running the full reservoir simulation, we use the polynomial chaos proxy for the Bayesian inference. Then, we apply the MCMC method to sample from the posterior distribution.

The organization of this paper is as follows; Section 2 presents the framework of the sparse polynomial chaos proxy. It includes the review of the Karhunen-Loeve expansion as a dimensionality reduction technique, and the derivation of a sparse formula for the polynomial chaos expansion, followed by a review of the Bayesian inverse framework and the MCMC method to sample from the analytical approximation of the posterior distribution. In Section 3, the proposed algorithm is applied for calibration of an analytical one dimensional elliptic stochastic PDE and a history matching problem of a two dimensional example of fluvial channels. Finally, the conclusions of our work are drawn in Section 4.