Polynomial Chaos Expansion for Uncertainty Quantification in Closed-Loop Reservoir Management*

Tarek Diaa-Eldeen*, Morten Hovd*, Carl Fredrik Berg†
*Department of Engineering Cybernetics
†Department of Geoscience and Petroleum
Norwegian University of Science and Technology (NTNU)
N-7491 Trondheim, Norway
{tarek.diaa-eldeen, morten.hovd, carl.f.berg}@ntnu.no

***Abstract*—Closed-loop reservoir management (CLRM) is a model-based optimal control procedure that aims at optimizing oil and gas production strategies under both physical and operational constraints and large model uncertainties. Using stochastic simulation in this decision-making process is imperative due to the large uncertainties that impact the model predictions. However, this often involves performing a large number of model evaluations repeatedly to integrate an ensemble of realizations that represent the input uncertainty. In CLRM, this requires excessive computational effort due to the complexity of the nonlinear and high-dimensional reservoir simulation models. In this study, a surrogate modeling technique, namely, polynomial chaos expansion (PCE), is leveraged for efficient and accurate implementation of the stochastic reservoir simulation in CLRM. A PCE for the reservoir dynamics is computed and employed to propagate the uncertainty without the need for additional expensive model evaluations. This can reduce the computational burden in both the forward and inverse problems of the CLRM. Results show that the PCE surrogate model can accurately quantify the uncertainty and evaluate a large number of model realizations at the cost of evaluating a polynomial, compared to the full model evaluations using Monte Carlo simulations.**

***Index Terms*—Closed-loop reservoir management (CLRM), Uncertainty quantification (UQ), Polynomial chaos expansion (PCE), Spectral methods, Stochastic simulation.**

I. Introduction

Uncertainty quantification (UQ) concerns the propagation of input uncertainties and the characterization of their impact on the predictions of computer simulation models. The significance of UQ comes from the fact that, in simulating complex systems, such as reservoir dynamics which are governed by nonlinear partial differential equations (PDEs) with high input uncertainties, it is imperative to take the uncertainties' effect into account from the beginning of the simulation [1]. This is because its impact will accumulate over time to make considerable changes that cannot be accounted for using other techniques such as increasing the numerical resolution or incorporating the uncertainty directly in the outputs. This is of particular importance when the models are used in simulation-based control, such as model predictive control (MPC) and optimal control problems (OCPs) [2]. In such cases, stochastic simulation techniques enable solving the optimization problem based on the characterized probabilistic uncertainties instead of the traditional approaches based on the worst-case analysis. To that end, different methods have been applied. Mainly, Monte Carlo (MC) simulations are exploited for the purpose of the stochastic simulation of dynamical systems by using the deterministic model to simulate a large ensemble of model realizations that represents the uncertainty in the inputs.

Although MC simulation has been considered the main stochastic simulation approach in a wide range of dynamical systems, it exhibits a significant drawback that limits its applicability in complex and high-dimensional systems. That is, it requires a large number of realizations due to its slow convergence; where it follows directly from the central limit theorem (CLT) that, independently of the problem dimensionality, MC converges inversely proportional to the square root of the number of realizations. In other words, it requires a large number of full-model evaluations using the deterministic simulator, which can be intractable in high-dimensional systems where each single simulation run takes considerable time. A sub-optimal solution is to trade the performance accuracy for the computational feasibility by reducing the number of model realizations. However, this leads to systematic underestimation in the propagated uncertainties both in the forward and the inverse problems in the closed-loop reservoir management (CLRM), leading eventually to substantial degradation in the overall control performance.

Other non-sampling approaches for simulating stochastic dynamics have been introduced and applied. For example, the perturbation method [3] is used to represent the random field using Taylor series expansion around the mean value. However, representing the uncertainties as perturbation cannot generally be a valid assumption when the system is characterized by large uncertainties. Similar are other operator-based methods, such as [4], which depend on the manipulation of the stochastic operator in the system equations. Another approach

This research is a part of BRU21 – NTNU Research and Innovation Program on Digital and Automation Solutions for the Oil and Gas Industry (www.ntnu.edu/bru21) and supported by Equinor.

is to derive the moments of the solution directly from the governing stochastic equation [1], which is only applicable in a limited number of cases.

On the other hand, a computationally effective way to handle those different limitations in high-dimensional systems characterized by large uncertainties is to combine sampling with metamodelling. In metamodelling, computationally demanding models, such as finite element models (FEM), are substituted by surrogate models that can be computed efficiently. For stochastic simulation purposes, the surrogate model can approximate the original system either in a strong sense where the response of the surrogate model converges to the true system's response in a proper norm, *e.g.*, mean-square convergence; or in a weak sense where it convergences in the solution statistics, *i.e.*, it only approximates the probability distribution of the original system's response. The polynomial chaos expansion (PCE) [5] has been introduced as a computationally-efficient metamodelling technique for simulating nonlinear stochastic dynamical systems. It is a high-order method that uses the spectral decomposition of the system dynamics in the input random space to represent the stochastic solution. Therefore, it can be efficiently used for UQ purposes in a wide range of complex applications. For Gaussian random variables and second-order random processes, Hermite polynomials [7] have a universal approximation property. For other probability distributions, the generalized polynomial chaos (gPC) [6] generalizes the PCE theory by introducing other polynomial functions.

On the application side, implementing model-based control and real-time optimization procedures is hindered by the model's complexity in the case of subsurface reservoir systems. This complexity comes mainly in the form of high-dimensional problem spaces that characterize the subsurface flow dynamics (can be in the order of 10^6), resulting from discretizing a coupled system of PDEs. This high-dimensionality attribute is threefold—it represents a high-dimensional state space, a high-dimensional parameter space, and high input uncertainties which require large numbers of model realizations. The uncertainty in the initial models, for instance, can cause the actual production measurements to be exceedingly distinct from predicted values during the initial production phases before model updates based on production data. This makes it essential to take the uncertainty into account from the start of the process. On the other hand, pure data-driven control approaches are not practical alternatives due to the scarcity of field data. Recently, the technological advancements in computer systems, field sensors, and industrial infrastructures, in addition to the increased computational power, have led to the ability to make real-time decisions based on predictive models whose uncertainties are sequentially updated by assimilating the real-time production data. This optimization framework, known as CLRM, combines online model updating with real-time optimal control to make the production decision-making process a near-continuous controlled process based on the continuously updated uncertainties [8]. It aims at maximizing the revenue, usually represented in terms of the net present value (NPV), from the oil recovery and meeting the increased energy demands, while minimizing costs and satisfying different production and environmental constraints. As illustrated in Figure 1, CLRM starts with highly uncertain prior models that are refined using production data assimilation to produce models with higher accuracy while being amenable to real-time optimization.

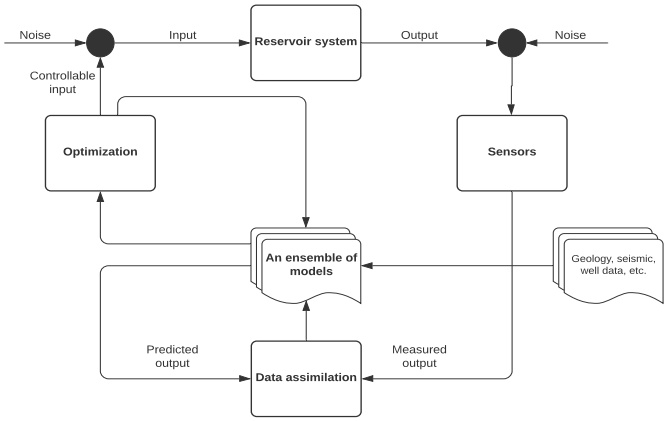

Fig. 1. Closed-loop reservoir management (CLRM).

Implementing the CLRM however requires propagating high-dimensional uncertainties twice in each loop–once in an inverse problem to update the models, and then in a forward problem to predict future performance to solve a real-time OCP. With multiple simulation runs that are required in each problem to represent the different scenarios, implementing CLRM in real-world problems is still challenging. For this reason, PCE can be a practical alternative that quantifies the uncertainty in both the inverse and the forward problems in CLRM at a much lower computational cost. In addition to oil and gas reservoirs, this efficient stochastic simulation of the fluid flow dynamics through porous media is also of significant importance in other subsurface application areas such as CO_2 sequestration, geothermal energy system, groundwater remediation and aquifer management, and bioreactor landfills.

II. High-Dimensional Uncertainty Quantification

Given a probability space defined by the triplet $(\Omega, \mathcal{F}, P)$, where Ω is the sample space, $\mathcal{F} \subset 2^{\Omega}$ is the σ-field of Ω, and $P : \mathcal{F} \to [0, 1]$ is the probability measure. We define the Hilbert measure space $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ of all the random variables on $(\Omega, \mathcal{F}, P)$ that have finite second moments; *i.e.*, $\mathcal{L}^2(\Omega, \mathcal{F}, P) = \{X : \mathbb{E}[X^2] < \infty\}$. A general system of partial differential equations (PDEs) is defined in a time domain $[0, T]$, such that $T > 0$, and a spatial domain $\mathcal{D} \subset R^l$, where $l \in \mathbb{Z}$ as follows

$$\begin{aligned} u_t(x, t, \omega) &= \mathcal{G}(u), & \mathcal{D} \times (0, \mathcal{T}] \times \Omega, \\ \mathcal{B}(u) &= 0, & \partial\mathcal{D} \times [0, \mathcal{T}] \times \Omega, \\ u &= u_0, & \mathcal{D} \times \{t = 0\} \times \Omega, \end{aligned} \tag{1}$$

where $\omega \in \Omega$ is the random inputs, $x \in \mathcal{D}$ are spatial points, $\mathcal{G}$ is a general (nonlinear) differential operator, $\mathcal{B}$ is the boundary condition operator, and u_0 is the initial condition. Simulating the deterministic dynamics of system (1) (one realization of ω) requires solving a high-dimensional system resulting from the spatial discretization of the domain $\mathcal{D}$. Considering the intrinsic stochasticity in addition would require solving this high-dimensional system multiple times at each time step. Metamodeling using PCE offers a computationally efficient solution where a small number of model simulations, N_{DoE}, such that $N_{DoE} << N_{MC}$, where N_{DoE} is the number of nodes in the experimental design and N_{MC} is the number of MC simulations, is used to train a surrogate stochastic model for the system (1), which can then be used to quantify the uncertainty with much smaller sampling errors by arbitrarily increasing the number of model evaluations at approximately no additional cost (the computational cost of polynomial substitution is negligible compared to integrating a system of nonlinear PDEs).

A. Random Field Discertization

Although the infinite-dimensional probability space in (1) will be reduced to a finite-dimensional space after discretizing the PDE in simulation, it is still required to further reduce the discretized probability space in order to be amenable to computing since the PDE discretization schemes often produce a high-dimensional finite space. In reservoir models, the random quantities, such as the space-varying parameters, are random fields that are functions of space. Following the notation in [1], the space-varying parameters are modeled as stochastic processes X_x defined on the continuous domain $\mathcal{D}$ and the space Ω: $(X_x(\omega), x \in \mathcal{D}, \omega \in \Omega)$. Therefore, for a vector of standard random variables (called the *germ*) $\Xi = (\xi_1, \xi_2, ..., \xi_d), d \geq 1$, where ξ_i's are finite and mutually independent, a diffeomorphism $T : \Xi \rightarrow X_x$ is required such that $X_x = T(\Xi)$. In other words, the resulting distribution in the standard random vector, Ξ, is equivalent to the original distribution, $X_x \sim T(\Xi)$. The need for a random variable with mutually independent components is desired for practical purposes, as most of the numerical techniques require the probability parameterization to be characterized by mutual independence in addition to being finite [1]. Different functions can be used in practice to realize this transformation for a given *germ*. In this study, Karhunen–Loève expansion (KLE) is applied such that it implements the transformation from a finite set of mutually independent random variables, Ξ, that are required for the numerical implementation of the PCE, to the original random fields used in the simulation, X_x. KLE, also known as proper orthogonal decomposition (POD), is a spectral method that uses the spectral decomposition of the covariance function for a stochastic process to approximate it using a finite set of uncorrelated random variables. It is an optimal expansion in mean squares sense [5]. Given a bounded symmetric positive definite covariance function $C(x, s)$ for the stochastic process X_x, the KLE is expressed as

$$\begin{aligned} X_x(\omega) = X(x,\omega) &= \langle X(x) \rangle + X'(x, \omega) \\ &= \mu_X(x) + \sum_{i=0}^{\infty} \sqrt{\lambda_i} \xi_i(\omega) \psi_i(x), \end{aligned} \tag{2}$$

where $\mu_X(x) = \langle X(x) \rangle$ is the mean of the stochastic process X_x, such that $\mu_X(x) = \mathbb{E}[X(x, \omega)] = \int_\Omega X(x, \omega) dP(\omega)$, and $X'(x, \omega)$ is the fluctuation. The $\{\xi_i(\omega)\}$ is a sequence of mutually uncorrelated zero-mean unit-variance Gaussian random variables. That is, $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}[\xi_i \xi_j] = \delta_{ij}$. λ_i and ψ_i are the eigenvalues and the corresponding orthogonal eigenfunctions of the covariance function $C(x, s) = \langle X'(x, \omega) X'(s, \omega) \rangle$, respectively, and are obtained by solving the following Fredholm integral equation

$$\int_{\mathcal{D}} C(x, s) \psi_i(s) ds = \lambda_i \psi_i(x). \tag{3}$$

KLE using the infinite series is exact for Gaussian random fields. However, a finite expansion is required in practice to discretize the random field, X_x. Therefore, the infinite series is truncated at a specific term $i = M$ that satisfies a certain energy content in the expansion; ***i.e.***,

$$\sum_{i=1}^{M} \lambda_i / \sum_{i=1}^{\infty} \lambda_i \geq \delta_e, \tag{4}$$

where δ_e is a cutoff threshold value. For a certain threshold, the truncation limit, M, is inversely proportional to the correlation length of the random field.

KLE is one of the most used techniques for random field discretization, as it preserves the variability of the original distribution while using a small number of independent random variables [9]. Hence, it handles both requirements of dimensionality reduction and random space reparameterization at the same time. Therefore, KLE will represent the transformation T such that $X_x(\omega) = T(\Xi) \equiv KLE(\Xi)$.

B. Polynomial Chaos Expansion (PCE)

On the other hand, the distribution of the system's response represented by the numerical simulator $Y = f(X_x(\omega))$ cannot be decomposed using KLE since its covariance function cannot be determined due to its dependence on uncertain random fields, $X_x(\omega)$. Therefore, PCE is leveraged to approximate the numerical simulator Y in terms of the finite independent random variables, Ξ, by expanding the function f in a polynomial series with polynomial basis Φ, that constitutes a set of orthogonal functions with respect to the distribution of the *germ*, Ξ.

$$\begin{aligned} Y(x, t, \omega) &= f(X_x(\omega)) = f(X(x, \omega)) \\ &= f(T(\Xi)) = \sum_{J \in \mathbb{N}_0^M} \alpha_J(x, t) \Phi_J(\Xi), \end{aligned} \tag{5}$$

where $J = (j_1, j_2, ..., j_M) \in \mathbb{N}_0^M$ is a multi-index with $|J| = j_1 + j_2 + ... + j_M$. The mode strength $\alpha_J(x, t)$ are

the polynomial coefficients, which are deterministic functions of space and time such that

$$\alpha_J = \frac{\langle f, \Phi_J \rangle}{\langle \Phi_J, \Phi_J \rangle}, \tag{6}$$

and the mode function $\Phi_J(\Xi)$, where Ξ is a vector, are multivariate polynomials that constitute the tensor product of the polynomial bases for each ξ_i. That is, the N-th degree polynomial in an M-variate random vector Ξ is given by

$$\Phi_J(\Xi) = \prod_{i=1}^{M} \phi_{j_i}(\xi_i),\ 0 \leq |J| \leq N, \tag{7}$$

and the orthogonality is defined by the inner product

$$\langle \phi_n, \phi_m \rangle := \int_{\xi \in \Xi} \phi_n(\xi)\phi_m(\xi) dP_\Xi(\xi) = \gamma_n \delta_{nm}$$

where $\gamma_n = \mathbb{E}[\phi_n^2(\xi)]$ are the normalization factors, and δ_{nm} is the Kronecker delta function. Since orthogonality is defined in terms of the probability function of the *germ*, $dP_\Xi(\xi)$, there exists a relation between the type of the orthogonal polynomial basis and the distribution of the *germ*. The orthogonality in the case of standard Gaussian distributed random variables defines a Hermite polynomial $\{H_J(\Xi)\}$ as defined in (7), such that

$$\phi_{j_i}(\xi_i) = h_{j_i}(\xi_i) = (-1)^{j_i} exp\left(\frac{1}{2}\xi_i^2\right) \frac{d^{j_i}}{d\xi_i^{j_i}} exp\left(-\frac{1}{2}\xi_i^2\right). \tag{8}$$

Hermite-type PCE can approximate any functional in $\mathcal{L}_2$ and converges in the $\mathcal{L}_2$ sense with a convergence rate that depends on the smoothness of the original function [10].

In practice, a finite series is used by truncating (5) to the N-th degree approximation

$$Y^N(x,t,\omega) = \sum_{J \in \mathcal{A}^M} \alpha_J(x)\Phi_J(\Xi). \tag{9}$$

where $\mathcal{A} \subset \mathbb{N}^M$ is a subset of multi-indices selected according to a truncation scheme such that

$$\mathcal{A}^M = \{J \in \mathbb{N}^M : |J| \leq N\}, \text{and} \quad \text{card}\, \mathcal{A}^M \equiv C = \binom{M+N}{N},$$

which indicates an exponential growth in the expansion terms with increasing the polynomial degree, N.

C. Non-intrusive Projection

Determining the *germ*, Ξ, directly determines the type of the polynomial, $\Phi(\Xi)$. The remaining is to compute the coefficients $\alpha_J(x)$ to completely construct the PCE surrogate model. The coefficients are computed such that they minimize the distance (error) between the expansion and the original function. This can be done using projection as shown in (6) to exploit the orthogonality property of the basis functions. However, calculating the numerator in (6) usually requires solving the integration problem of the inner product numerically, which is nontrivial; particularly, in the reservoir simulation case when the function f requires solving another high-dimensional numerical integration problem. This makes computing the PCE coefficients using intrusive approaches such as stochastic Galerkin projection intractable in the reservoir case. Non-intrusive approaches, on the other hand, are generally based on collocation methods and interpolation, where the residue of the governing stochastic equations is brought to zero at some discrete points in the random input domain (the collocation nodes). This can be implemented exactly by solving the integration in (6) using a quadrature rule, or approximately using regression analysis. Although the quadrature methods are more accurate as they have an exact solution at some points in the random domain, their applicability is also limited to low random dimensions due to the curse of dimensionality problem in the case of multivariate random variables, as they are applied through the tensor-product of one-dimensional nodes. For example, applying a multivariate Gaussian quadrature rule is done through the tensor-product of the univariate integration rules (tensor product collocation), which makes the number of collocation nodes (*i.e.*, the required full-scale model evaluations) increase drastically with the number of random variables. Although some high-order stochastic collocation methods have been developed [11] based on sparse grids collocation techniques such as Smolyak sparse grids [12], they are still implemented using tensor-product construction–on a subset of the full tensor grids in this case–and their computational cost still cannot be arbitrarily reduced due to the restrictions in the experimental design. That is, the number (and weights) of the collocation nodes are uniquely determined from the applied quadrature rule (deterministic sampling). Therefore, sparse regression analysis using regularized least-squares is used in this study as it gives more freedom in the experimental design and the selection of the model evaluation points.

With a proper ordering of the multi-index, J, to represent it by a single index, *e.g.*, the graded lexicographic order, the truncated PCE (9) can be represented as a regression model

$$\begin{aligned} Y^{(i)}(x,t,\omega) = f(T(\Xi^{(i)})) &= \sum_{j=0}^{C-1} \alpha_j(x,t)\Phi_j(\Xi^{(i)}) + \epsilon_C \\ &= \alpha^T \mathbf{\Phi}(\mathbf{\Xi^{(i)}}) + \epsilon_C \end{aligned} \tag{10}$$

where $\Xi^{(i)}$ are the design points (collocation nodes), such that $i = \{0, 1, ..., Q\}$. C is the cardinality of $\mathcal{A}^M$, ϵ_C is the truncation error, $\alpha^T(x,t) = \{\alpha_0, \alpha_1, ..., \alpha_{C-1}\}$, and $\mathbf{\Phi}(\mathbf{\Xi}) = \{\Phi_0(\Xi), \Phi_1(\Xi), .., \Phi_{C-1}(\Xi)\}$. The sparse least angle regression (LARS) algorithm [13] is applied in this work to solve the regression problem (10) and compute the PCE coefficient. The LARS algorithm is implemented by regularizing the least-squares minimizer for (10) as follows

$$\hat{\alpha} = \underset{\alpha \in \mathbb{R}^{|\mathcal{A}|}}{\text{argmin}}\ \mathbb{E}\left[\left(\alpha^T \mathbf{\Phi}(\mathbf{\Xi}) - f(T(\Xi))\right)^2\right] + \lambda ||\alpha||_1 \tag{11}$$

where the regularization term $||\alpha||_1 = \sum_{J \in \mathcal{A}} |\alpha_J|$ penalizes high-rank solutions, resulting in a sparse regression technique which is suitable for high-dimensional problems.

The computed PCE can then be used as a surrogate model in UQ and stochastic simulation. In addition, the statistical

properties of the stochastic response of the system can be inferred directly from the properties of orthogonal polynomials with no more computations. For example, the first two statistical moments–which fully characterize the uncertainty in the Gaussian case–are found in closed form as follows

$$\mu^Y(x,t) = \mathbb{E}\big[Y(x,t,\omega)\big] \ = \alpha_{j_0}(x,t), \tag{12}$$

$$\begin{aligned}\Sigma^Y(x,t) &= \mathbb{E}\big[(Y(x,t,\omega) - \mu^Y(x,t))^2\big] \\ &= \sum_{0\leq|J|\leq N} \big[\gamma_J \alpha_J^2(x,t)\big]\end{aligned} \tag{13}$$

III. Closed-Loop Reservoir Management

CLRM consists of two main components, data assimilation and decision-making; and both depend on uncertainty propagation. In the data assimilation part, the model uncertainties are propagated forward in time to be updated when new information is available from the field measurements. In the decision-making step, the updated uncertain models are used to predict future production in order to implement optimal production strategies. PCE is an efficient tool to propagate uncertainty in both steps.

A. Flow Dynamics

The dynamics of the subsurface flow through porous media are considered for an isothermal, heterogeneous, two-phase (oil and water), black-oil reservoir model. In the isothermal case, the governing equations are modeled by combining a mass-balance equation (conservation of mass), for each phase, and the extension of Darcy's law to multiple phases together with the equations of state that describe the fluid properties as a function of pressure. The following system of nonlinear PDEs models the reservoir dynamics, where the gravity forces, inertial effects, and capillary pressure are neglected

$$\frac{\partial}{\partial t}(\phi\rho_w s_w) = \nabla\cdot\left(\overrightarrow{K}\frac{k_{rw}}{\mu_w}\rho_w\nabla p\right) + q_w, \tag{14}$$

$$\frac{\partial}{\partial t}(\phi\rho_o(1-s_w)) = \nabla\cdot\left(\overrightarrow{K}\frac{k_{ro}}{\mu_o}\rho_o\nabla p\right) + q_o, \tag{15}$$

where ϕ is the porosity, ρ is the fluid density, and s_i is the i-th phase saturation, where the subscript $i \in \{w, o\}$ indicates water or oil, respectively. p represents the fluid pressure, which is equivalent for the water phase and the oil phase given that the capillary pressure is neglected. $\overrightarrow{K}$ is the absolute (intrinsic) permeability tensor, k_r is the relative permeability, μ is the dynamic viscosity, and q is the flow rate (source term).

Using the definition of the i-th phase compressibility, c_i, and rock compressibility, c_r,

$$c_i = \frac{1}{\rho_i}\frac{\partial\rho_i}{\partial p}, \quad c_r = \frac{1}{\phi}\frac{\partial\phi}{\partial p}, \tag{16}$$

(14) and (15) can be written as follows

$$\begin{aligned}&\phi\rho_w\left(s_w(c_w+c_r)\frac{\partial p}{\partial t} + \frac{\partial s_w}{\partial t}\right) \\ &\quad = \nabla\cdot\left(\overrightarrow{K}\frac{k_{rw}}{\mu_w}\rho_w\nabla p\right) + q_w,\end{aligned} \tag{17}$$

$$\begin{aligned}&\phi\rho_o\left((1-s_w)(c_o+c_r)\frac{\partial p}{\partial t} - \frac{\partial s_w}{\partial t}\right) \\ &\quad = \nabla\cdot\left(\overrightarrow{K}\frac{k_{ro}}{\mu_o}\rho_o\nabla p\right) + q_o.\end{aligned} \tag{18}$$

In addition, a Peaceman nonlinear model [14] that describes the well dynamics is used to specify the pressure in at least one grid block to define the boundary conditions.

B. Deterministic and Stochastic Simulation

The system is deterministically simulated using a fully-implicit approach that combines a finite difference (FD) discretization scheme with a Newton-based iterative algorithm to solve the nonlinear system of PDEs. This results in the following state and output finite difference equations, respectively,

$$g_k(u_{(i)}, u_k, r_k, \theta) = e_k, \tag{19}$$

$$y_k(u_k, r_k, \theta) = 0. \tag{20}$$

where u is the state vector, r is the input vector, θ is the parameter vector, the subscript $\square_k$ is for the time steps in the FD scheme, while $\square_{(i)}$ is for the iterative solver that solves for the state at the next time step, u_{k+1}, and e is a vector of residuals that is being minimized over the iterations.

Although the reservoir models are subject to various sources of uncertainty, both in time- and space-varying variables, it is typically the uncertainty in the space-varying parameters that dominates other sources. Uncertainty in the time-varying variables (system states) comes primarily from discretization errors and unmodeled physics. However, characterizing the space-varying physical properties (parameter fields) is highly uncertain due to the spatial heterogeneity of the subsurface and the scarcity of information sources. Therefore, the parameters θ will represent the random inputs, ω, in (1).

IV. Numerical Study

In this study, a heterogeneous 2D two-phase synthetic reservoir in a waterflooding production process is used to demonstrate the PCE implementation for CLRM purposes. Four production wells in the four corners of the spatial domain and one injection well in the middle are used to recover the subsurface oil. The boundary conditions are set by specifying the control inputs as the water injection rate in the injectors and the production bottom hole pressure (BHP) in the producers, and no-flow boundaries otherwise. The observed outputs are specified as the flux out of the producers and the BHP in the injector. The coordinate system is assumed to be aligned with the geological layers of the reservoir, and therefore, the permeability tensor is a diagonal matrix; *i.e.*, $\overrightarrow{K} = \left[\begin{smallmatrix} k_x & 0 \\ 0 & k_y \end{smallmatrix}\right]$. The porosity field is assumed to be uniform over the whole domain. Each spatial dimension is discretized to 21 grid

TABLE I
RESERVOIR MODEL PARAMETERS

Variable	Symbol	Value	SI Unit
L_x, L_y	Reservoir dimensions	1000	m
$n_x \times n_y$	Number of grid-blocks	21×21	–
h	Grid-block height	10	m
μ_w	Water viscosity	4.0×10^{-4}	Pa.s
μ_o	Oil viscosity	9.0×10^{-4}	Pa.s
$s_{w,0}$	Initial water saturation	0.2	–
$p_{o,0}$	Initial reservoir pressure	3.5×10^7	Pa
r_w	Wellbore radius	0.1	m
q_{inj}	Water injection rate	1.7×10^{-3}	sm^3/s
BHP_{prod}	Production bottom-hole pressure	3×10^7	Pa

blocks, resulting in a total of 441 grid blocks in the space domain. In general, each grid block is characterized by four states, namely the phase saturations and pressures, and two highly uncertain parameters representing the permeability in both directions. In our model, the permeability field is assumed to be isotropic; therefore, $k_x = k_y$ in each grid block; and since capillary pressure is set to zero there is only one pressure variable. A simulation time domain of 30 years is discretized using a $6-$month discretization timestep, and observations are collected at the wells every 30-th month (5 simulation timesteps). Table I summarizes the main properties of the reservoir and the waterflooding experiment used in this study. The MATLAB Reservoir Simulation Toolbox (MRST) [15] and the UQLab [16] are used to implement the experiments.

A. Model Reparameterization

Since the independent uncertain variable is assumed to be the directional permeability, which is spatially correlated along the domain, the model is first reparameterized to be represented in terms of independent low-dimensional random variables. In this case, a large ensemble matrix of 1000 realizations is used to represent the parameter's variability to obtain the KLE using the sample covariance matrix instead of the covariance function which is not available in closed form. The sample covariance matrix, $\hat{P}$, is computed as follows

$$\hat{P} = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (e_i - \bar{e})(e_i - \bar{e})^T, \tag{21}$$

where N_e is the number of ensemble members, e_i, and $\bar{e}$ is the ensemble mean. The KLE for the distribution of the permeability field is obtained using the first 16 principal components that represent the variability in the ensemble. Figure 2 shows those modes in which the distribution of the field is represented. It was found that 16 principal components retain more than 75% of the variance energy as shown in Figure 3. That is, $\delta_e = 0.75$ in (4) results in $M = 16$.

B. Design of Experiments (DoE)

The PCE coefficients are computed using LARS sparse regression algorithm. The design of experiments (DoE) step concerns finding a set of samples from the random input space and the corresponding model evaluations so that they are employed to solve the regression problem in (10). $N_{DoE} = 200$ model realizations are sampled from the input random space using Latin hypercube sampling (LHS) and the corresponding model responses are evaluated for this purpose. This constitutes the total cost of computing PCE, since the full-scale simulator has to be run N_{DoE} times to solve the regression problem. Hermite polynomials of order $N = 6$ are selected to form the polynomial basis, and (11) is solved to compute the coefficients for five PCEs that approximate the five system outputs over the simulation period. It was found that $N = 6$ preserves a proper balance between the computational cost and the approximation accuracy for the surrogate model. That is, increasing the order over 6 introduces a significant increase in the computational time of learning the surrogate model, while offering negligible improvements in the approximation accuracy.

C. Results and Validation

After computing the different PCE surrogate models, they are used as predictors to forecast the response of new points outside the experimental design. Figures 4 - 8 compare the responses of the actual simulator and the PCE models in a cross-validation experiment using 30 MC samples. With no loss of generalization, only 30 samples are used in validation in order to clarify the convergence and to enable implementing this number of MC simulations using the actual simulator for the sake of comparison. However, the results generalize to any number of samples. The responses illustrate that the obtained PCE can accurately approximate the stochastic simulation of the actual system. Nevertheless, while the traditional MC simulation requires running the full-scale simulator N_{MC} multiple times, the computational cost of the PCE model is only equivalent to computing a polynomial N_{MC} multiple times. This constitutes a reduction in the computational time from 0.9913 second in the full simulation case to $3.0669e-04$ second in the PCE case, for the single reservoir realization (averaged over 1000 realizations). The most expensive step in the PCE, however, is in computing its coefficients, which requires $N_{DoE} = 200$ simulation runs. That means, at the cost of 200 model evaluations in this example, a PCE model is tuned and it can then be used to evaluate an arbitrarily large number of model evaluations at almost no additional computational cost. This makes it most adequate for problems that require large numbers of repetitive simulations, which is typically the case in CLRM.

The figures show some rigidity (low resolution) in the response of the PCE that does not resemble the smooth response of the actual system around the changes around water breakthrough (between 100 and 150 months). This can mainly be attributed to the polynomial degree, $N = 6$, and can be smoothed out by increasing the order of the polynomial basis.

Although the results can be conveniently improved using different modifications, analyzing the statistical performance shows that the current approximation is sufficiently adequate for purposes of uncertainty quantification in CLRM. Nonethe-

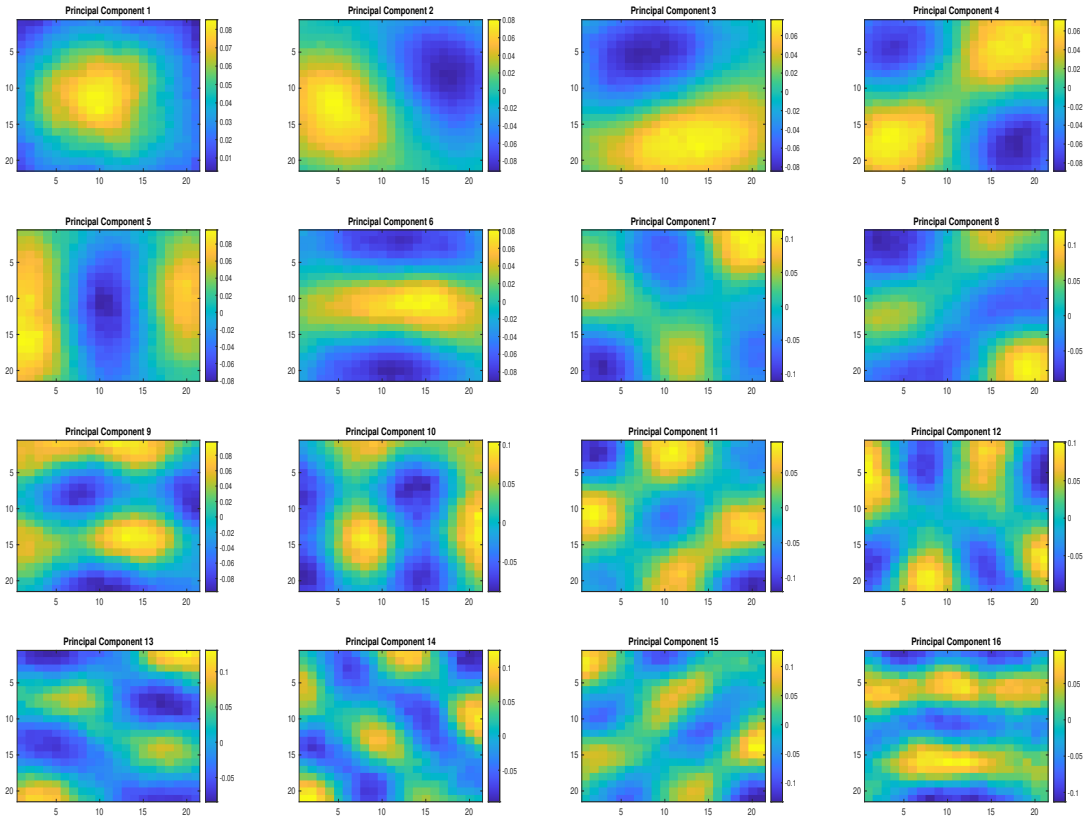

Fig. 2. Principal components used in the KLE of the permeability field distribution.

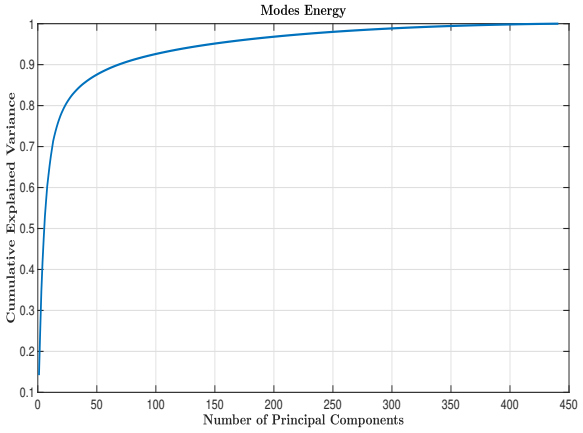

Fig. 3. Energy associated with the eigenvalues of the sample covariance.

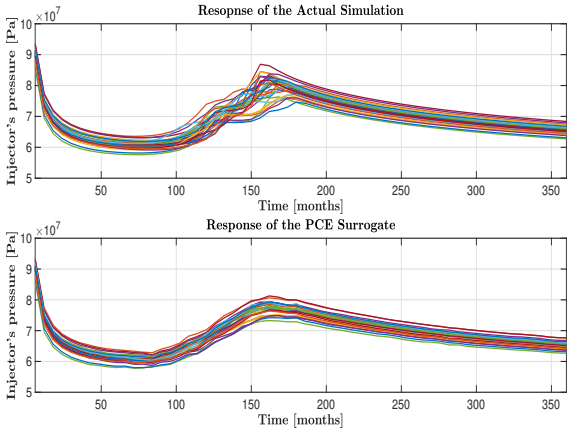

Fig. 4. Response of the actual and PCE models for the injector's pressure.

less, the approximation accuracy can be improved at the cost of increasing the computational cost by increasing the order of the polynomial, N, increasing the energy represented in the KLE, M, and/or increasing the number of samples used in the DoE or using a quadrature rule instead.

V. Conclusions

In this study, PCE surrogate model for the reservoir system has been developed for the purpose of uncertainty propa-

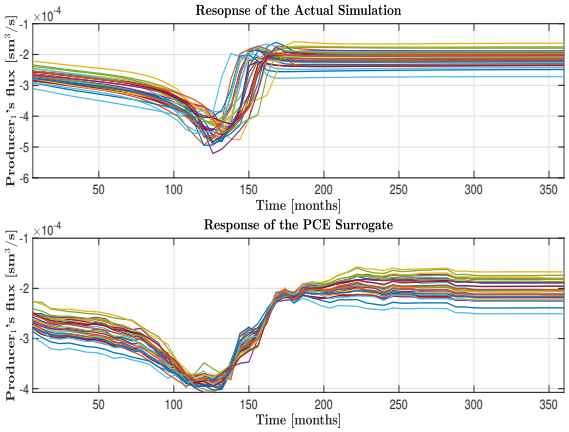

Fig. 5. Response of the actual and PCE models for the first producer's rate.

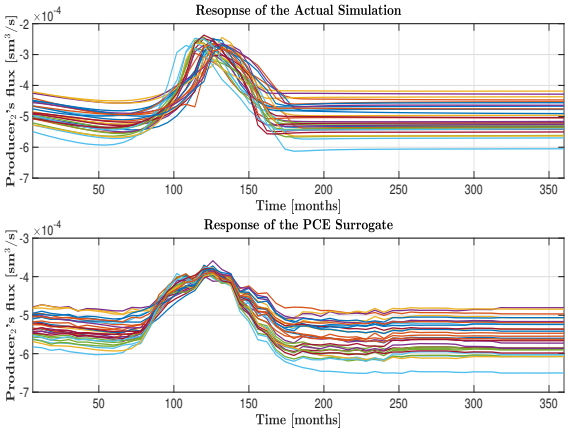

Fig. 6. Response of the actual and PCE models for the second producer's rate.

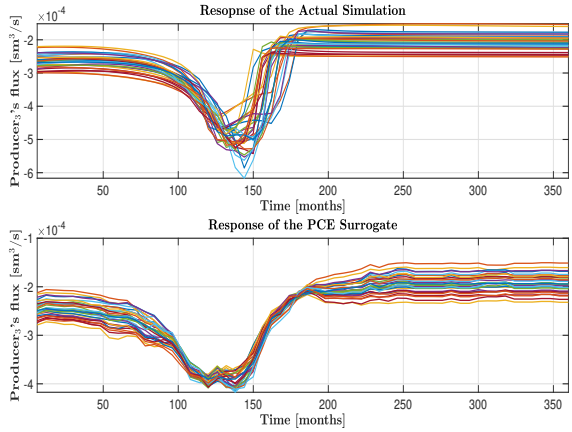

Fig. 7. Response of the actual and PCE models for the third producer's rate.

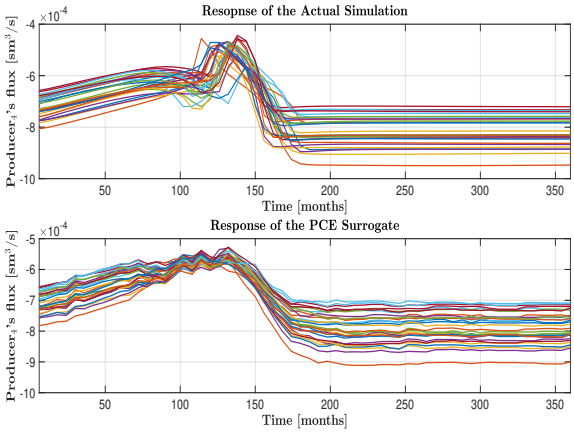

Fig. 8. Response of the actual and PCE models for the fourth producer's rate.

gation and quantification in CLRM. KLE has been used to reparameterize the input random fields. The numerical results show that system responses can be approximated to acceptable accuracy at a moderate computational cost. That, on the other hand, enables enhancing the propagation accuracy and reducing the sampling error by increasing the number of model realizations that can be evaluated using the PCE compared to MC simulation methods.

REFERENCES

[1] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton University Press, 2010.

[2] E. Bradford, and L. Imsland, "Output feedback stochastic nonlinear model predictive control for batch processes," Comput. Chem. Eng. 126, 2019.

[3] M. Kleiber and D. H. Tran, The Stochastic Finite Element Method: Basic Perturbation Technique and Computer Implementation. Wiley, 1993.

[4] F. Yamazaki, M. Shinozuka, and G. Dasgupta, "Neumann expansion for stochastic finite element analysis," J ENG MECH-ASCE, 1988, 114.

[5] R. Ghanem, and P. Spanos, Stochastic Finite Element: a Spectral Approach. Springer, New York, 2003.

[6] D. Xiu and G. Karniadakis, "The Wiener–Askey polynomial chaos for stochastic differential equations," SIAM J. Sci. Comput., 2002, 24.

[7] N. Wiener, The Homogeneous Chaos. American Journal of Mathematics, 60(4), 1938, 897–936.

[8] T. Diaa-Eldeen, C. F. Berg, and M. Hovd, "System-theoretic ensemble generation in ensemble-based history matching. ECMOR 2022, 2022(1).

[9] Y. M. Marzouk and H. N. Najm, "Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems," Journal of Computational Physics, v. 228(6), 2009, pp. 1862-1902.

[10] K. -K. K. Kim, D. E. Shen, Z. K. Nagy and R. D. Braatz, "Wiener's Polynomial Chaos for the Analysis and Control of Nonlinear Dynamical Systems with Probabilistic Uncertainties [Historical Perspectives]," in IEEE Control Systems Magazine, vol. 33, no. 5, pp. 58-67, Oct. 2013.

[11] D. Xiu and J. S. Hesthaven, "High-order collocation methods for differential equations with random inputs," SIAM J. Sci. Comput., 2005, 27(3), 1118-1139.

[12] S. A. Smolyak, "Quadrature and interpolation formulas for tensor products of certain classes of functions," Soviet Mth. Dokl., 1963, 4:240.

[13] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," Ann. Statist. 32(2), 407-499, (April 2004).

[14] D. W. Peaceman, "Interpretation of well-block pressures in numerical reservoir simulation," SPE J. 18 (1978): 183–194.

[15] K. Lie, An Introduction to Reservoir Simulation Using MATLAB/GNU Octave: User Guide for the MATLAB Reservoir Simulation Toolbox (MRST). Cambridge: Cambridge University Press, 2019.

[16] S. Marelli, and B. Sudret, UQLab: A Framework for Uncertainty Quantification in Matlab, 2014.