Polynomial Chaos Expansion for an Efficient Uncertainty and Sensitivity Analysis of Complex Numerical Models

E. Bastug, A. Menafoglio, Tatiana Okhulkova

Polynomial Chaos Expansion for an Efficient Uncertainty and Sensitivity Analysis of Complex Numerical Models

E. Baştuğ
Alcatel-Lucent Chair - SUPÉLEC, Gif-sur-Yvette, France.

A. Menafoglio
MOX-Department of Mathematics "F.Brioschi", Politecnico di Milano, Italy.

T. Okhulkova
Laboratoire MSSMat - ECP, Châtenay-Malabry, France.
INERIS, Verneuil-en-Halatte, France.

ABSTRACT: In this work we address the problem of performing uncertainty and sensitivity analysis of complex physical systems where classical Monte-Carlo methods are too expensive to be applied due to the high computational complexity. We consider the Polynomial Chaos Expansion (PCE) as an efficient way of computing a response surface for a model of gas injection into an incompressible porous media aiming at assessing the sensitivity indices and the main distributional features of the maximal spread of the gas cloud. The necessity of an uncertainty study for such a model arises in case of $CO2$ storage risk assessment and is here performed by jointly using a numerical scheme to solve the system of partial differential equation (PDE) governing the model and the PCE method to efficiently simulate the physical system response by a meta-model. The performances of the PCE method and a standard MC approach are compared through an extended simulation study showing that the computational gain of the PCE approach is remarkable without significant loss in the precision of the estimates.

1 INTRODUCTION

The uncertainty quantification for a complex numerical model is an important goal for which significant challenges are yet to be achieved. In general, the uncertainty might be associated to several elements of the system such as input data measurements, parameter values, model structure, model solution algorithms and even human behavior. Therefore, the uncertainty assessment under a given cost might not be straightforward.

Most of the research efforts have been focused so far on the development of efficient algorithms to be used in different applications, assuming ideal input data measurements with precisely defined computational domains: to this end, for instance, Monte-Carlo approaches can be employed. However, the simulation of a large-scale system turns in computationally very heavy operations: thus the need of making the convergence to the solution faster naturally arises. The classical Monte-Carlo approach is used in such cases as the last resort. Another option is the perturbation method where all the stochastic quantities are expanded around the mean via Taylor series, nevertheless, it is limited to small perturbations and does not provide information on high-order statistics of the response. The expansion of the inverse of the stochastic operator in a Neumann series can be considered, but it is again limited to small fluctuations (Shinozuka and Deodatis 1988).

A more effective approach pioneered by Ghanem and Spanos (1991) in the context of finite element method is the Polynomial Chaos Expansion, which is a method based on a spectral representation of uncertainty. This idea comes from the theory about polynomial chaos by Wiener (1938), (1962), which was firstly applied in the study of turbulence in the 1960s (Meecham and Siegel 1964, Siegel et al. 1965, Meecham and Jeng 1968). Nevertheless, its convergence rate was judged slow (Orszag and Bissonnette 1967, Crow and Canavan 1970, Chorin 1974): hence, it did not receive much attention until the work of Ghanem and Spanos (1991). A rich state of art with many applications can be found in (Sudret and Der Kiureghian 2000, Xiu and Karniadakis 2002, Xiu et al. 2002).

However, the availability of solid theoretical basis as well as efficient numerical codes (OpenTURNS 2012) for performing uncertainty analysis has recently contribute in making possible the sensitivity analysis of very complex systems (Blatman and Sudret 2010, Sudret 2008, Formaggia et al. 2012, Oladyshkin et al. 2011).

The main purpose of this work is to show how Polynomial Chaos Expansion (PCE) can be used for uncertainty and sensitivity analysis. In particular, we study as a specific application a model of gas injection into porous media, presented in Section 2. The theoretical basis of PCE is briefly introduced in Section 3. The validity of this approach is proved by simulation in Section 4: here, we first perform the sensitivity analysis on our model, then we compare the results of the analytical solution from available simplified model and the results of the Monte Carlo simulation.

2 INJECTION OF GAS INTO POROUS MEDIA

The predictive numerical model engaged in this study represents a two-phase flow of immiscible compressible fluids through incompressible porous medium. This model is developed in INERIS in order to estimate the intensity of physical processes in underground compartments in case of CO2 storage in a saline aquifer. The injected gas displaces water (initially in place) and simultaneous flow of the two fluids takes place in the porous medium. The model accounts for gravity, capillarity and fluid compressibility effects, and is formulated in terms of CO2 and water pressures. Empirical laws by van Genuchten (1980) are used to close the system with the relations between capillary pressure, relative permeability and saturation.

The corresponding system of differential equations is solved within the software COMSOL (2012) in 2D-axisymmetric mode. The source (injection well) is concentric to the symmetry axis.

The upper and lower bounds of the reservoir are considered impermeable, therefore a "zero flux" boundary condition is used. On the bound along the injection well the debit of the gas phase is fixed, on the other side the boundary condition fluid pressures are fixed. The main configuration and the variable of interest are shown in Figure 1.

In Figure 1, p_w and p_{nw} [Pa] are water and gas pressures respectively, C_p [Pa^{-1}] the coupling term, θ_w and θ_{nw} the water and gas volume fractions, k_{rw} and k_{rnw}[-] the water and gas relative permeabilities; Q_{nw} [m/s] represents the gas injection rate, K_{int} [m^2] the intrinsic permeability, p_{w0} and p_{nw0} [Pa] the initial water and gas pressures, ρ_w and ρ_{nw} [kg/m^3] the fluid densities, μ_w and μ_{nw} [$Pa \cdot s$] the fluid viscosities and, finally, χ_w and χ_{nw} [Pa] are the fluid compressibilities.

The numerical solution of the system at time $t = 10$ years is finally shown in Figure 2. The uncertainty analysis as well as the sensitivity analysis which will be described in Section 4 are referred to this solution.

3 POLYNOMIAL CHAOS EXPANSION FOR SENSITIVITY ANALYSIS

3.1 *Polynomial Chaos Expansion*

Let y be the scalar output of a model that depends on the n-dimensional random vector $\boldsymbol{x} = (x_1, ..., x_n)$ collecting n random parameters x_i, valued in $\Omega_i \subseteq \mathbb{R}$, $i = 1, ..., n$:

$$y = f(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega \subseteq \mathbb{R}^n, \tag{1}$$

where Ω denotes the cartesian product $\Omega_1 \times \cdots \times \Omega_n$.

Suppose $x_1, ..., x_n$ to be statistically independent and the response y to be a second-order integrable variable: the PCE of y (Maître and Knio 2010, Ghanem and Spanos 1991, Xiu and Karniadakis 2002) is the spectral decomposition of f over a set of polynomials $\Psi_i(\boldsymbol{x})$, which are orthogonal with respect to the probability measure on Γ (i.e. the law of $\boldsymbol{x}$), namely (Sudret 2008):

$$y = \sum_{k=0}^{\infty} \beta_k \Psi_k(\boldsymbol{x}). \tag{2}$$

The PCE was first introduced for Gaussian input parameters involving Hermite polynomials (e.g., Wiener (1938)). In the presence of non-Gaussian distributed variables, two strategies can be adopted in order to preserve the optimal exponential convergence rate: a) transform the input parameters to obtain Gaussian random entries b) consider other families of polynomials (Xiu and Karniadakis 2002, Soize and Ghanem 2005).

In this work, for every input parameter x_i, $i = 1, ..., n$, a log-normal distribution $\rho_{\Gamma_i}(x_i)$ is considered. In this case, the simplest way to reach the exponential rate of convergence is to use the Hermite basis, by transforming each input parameter x_i into a standard normal random variable ξ_i:

$$\xi_i = \Phi^{-1}(F_{x_i}(x_i)), \tag{3}$$

where Φ and F denotes the cumulative density function (CDF) of a standard Gaussian variable and of the variable x_i respectively. Given the vector of transformed (independent) input parameters $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)$, the PCE can be computed as:

$$y = \widetilde{f}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} \beta_k \Psi_k(\boldsymbol{\xi}). \tag{4}$$

For computation, expansion (4) needs to be truncated, obtaining finally an approximation of the response as:

$$y \approx \sum_{k=0}^{P-1} \beta_k \Psi_k(\boldsymbol{\xi}), \tag{5}$$

Figure 1: Specification of the problem: the vertical section of the domain and the system of differential equations entered in Comsol Multiphysics.

$r = 0$

Zero flux: $n \cdot \nabla p_w = 0, n \cdot \nabla p_{nw} = 0.$

$Q_{nw} = 1 Mt/year,$
$Q_w = 0.$

$$\rho_w \left(C_p \frac{\partial p_{nw}}{\partial t} + \left(-C_p + \frac{\theta_w}{K_w} \right) \frac{\partial p_w}{\partial t} \right) + \nabla \left(-\frac{K_{int} k_{rw} \rho_w}{\mu_w} \nabla (p_w - \rho_w g z) \right) = 0,$$
$$\rho_{nw} \left(C_p \frac{\partial p_w}{\partial t} - \left(C_p + \frac{\theta_{nw}}{K_{nw}} \right) \frac{\partial p_{nw}}{\partial t} \right) + \nabla \left(-\frac{K_{int} k_{rnw} \rho_{nw}}{\mu_{nw}} \nabla (p_{nw} - \rho_{nw} g z) \right) = 0.$$

Hydrostatic pressure
$p_w = p_{w0},$
$p_{nw} = p_{nw0}$

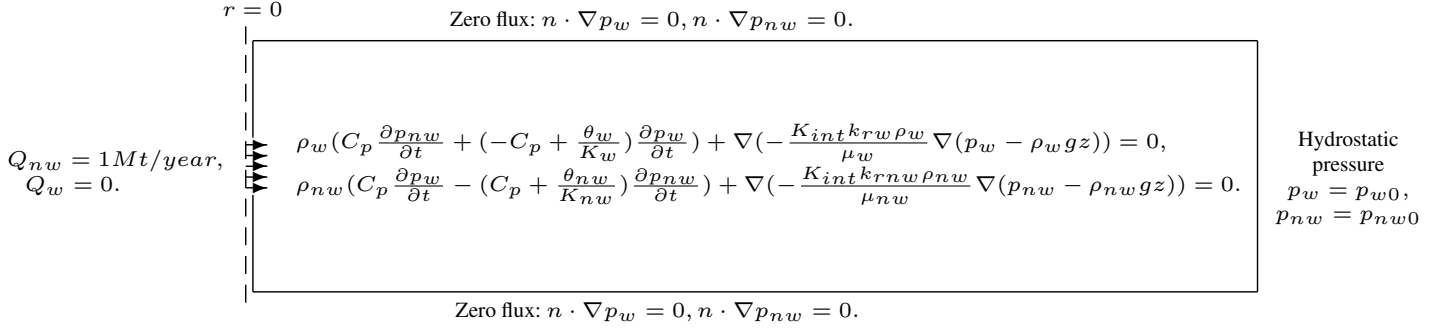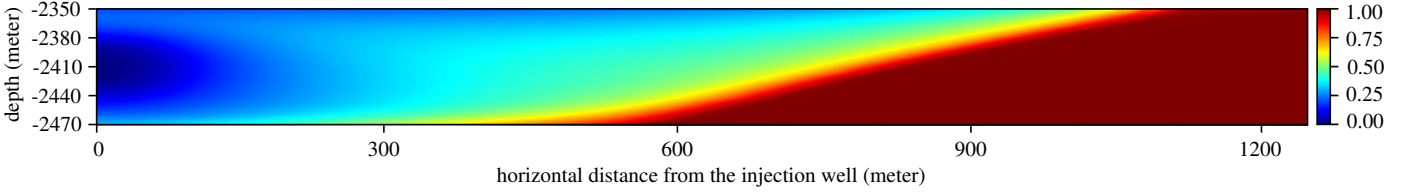Zero flux: $n \cdot \nabla p_w = 0, n \cdot \nabla p_{nw} = 0.$

Figure 2: Numerical solution of the system for $t = 10$ years. The color mapping shows the water saturation level.

where $P = \binom{n+p}{p}$, if (5) is the truncated expansion on a basis of n-dimensional Hermite polynomials of degree not exceeding p.

The P coefficients β_k, $k = 0, ..., P-1$ appearing in (5) are assumed deterministic and fully characterize the randomness of the response y. In applications, these coefficients are unknown and thus need to be properly estimated. A number of methods are presented in the literature (Sudret 2008, Crestaux et al. 2009); among them, in this work, the regression method is used. It consists in estimating β_k by least-squares on the basis of an experimental design $\{y^j, j = 1, ..., N\}$. The experimental design is built by collecting N exact solutions y^j, $j = 1, ..., N$, which are derived solving the complete mathematical model for N realizations $\boldsymbol{x}^j$, $j = 1, ..., N$, of the random input $\boldsymbol{x}$. The least-squares estimate of $\boldsymbol{\beta} = (\beta_k) \in \mathbb{R}^P$ is found by solving the minimization problem:

$$\boldsymbol{\beta} = \operatorname*{argmin}_{\beta_0, ..., \beta_{P-1} \in \mathbb{R}} \frac{1}{N} \sum_{j=1}^{N} \left\{ y^j - \sum_{k=0}^{P-1} \beta_k \Psi_k(\boldsymbol{\xi}^j) \right\}^2, \quad (6)$$

being $\boldsymbol{\xi}^j = (\Phi^{-1}(F_{x_i}(x_i)))$. If the design matrix $\boldsymbol{\Psi} = (\Psi_j(\boldsymbol{\xi}^i)) \in \mathbb{R}^{N,P}$ is of full rank, problem (6) admits the unique solution $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \boldsymbol{y}, \quad (7)$$

where $\boldsymbol{y} = (y^1, ..., y^N)$. Techniques for an efficient computation of coefficients estimates (7) has been presented in the literature (e.g., Sudret (2008)) and are exploited in the simulations here presented.

The utility of PCE for the uncertainty analysis of the response is two-fold:

- Response mean and variance can be directly obtained from (5) as:

$$\mathbb{E}[y] = \beta_0; \quad \operatorname{Var}(y) \approx \sum_{k=0}^{P-1} \beta_k^2 \|\psi_k\|^2.$$

Analogously, the sensitivity Sobol' indices can be computed directly from coefficients β_k, $k = 0, ..., P-1$ (see Subsection 3.2)

- The approximation (5) can be used as metamodel for Monte-Carlo simulations (i.e., the response is simulated through the approximate model instead of solving the PDE system). This allows to perform a much more efficient assessment of the CDF as the approximated model is easier to simulate compared to the full model (1).

3.2 Sensitivity Analysis through Sobol Indices

Sobol' indices (Sobol 1993) will be used in Section 4 to perform the sensitivity analysis on the response. The total variance of the response can be decomposed into the effect of each random parameters x_j and the joint effects of all the subsets $x_{i_1}, ..., x_{i_s}$ of the input parameters $x_1, ..., x_n$. As stated in (Sobol 1993, Sudret 2008, Formaggia et al. 2012), this variance decomposition follows from the Sobol' decomposition of the

function $\widetilde{f}(\boldsymbol{\xi})$ which appears in (4):

$$\widetilde{f}(\boldsymbol{\xi}) = \widetilde{f}_0 + \sum_{i=1}^{N} \widetilde{f}_i(\xi_i) + \sum_{i,j=1}^{N} \widetilde{f}_{i,j}(\xi_i, \xi_j) + \tag{8}$$

$$+ \cdots + \widetilde{f}_{1,2,\ldots,n}(\xi_1, \xi_2, \ldots, \xi_n),$$

where

$$\widetilde{f}_0 = \int_{\Gamma} \widetilde{f}(\boldsymbol{\xi}) \rho_{\Gamma}(\boldsymbol{\xi}) d\boldsymbol{\xi},$$

$$\widetilde{f}_{\boldsymbol{i}^{(s)}}(\boldsymbol{\xi}_{\boldsymbol{i}^{(s)}}) = \int_{\Gamma_{-\boldsymbol{i}^{(s)}}} \widetilde{f}(\boldsymbol{\xi}) \rho_{\Gamma_{-\boldsymbol{i}^{(s)}}}(\boldsymbol{\xi}_{-\boldsymbol{i}^{(s)}}) d\boldsymbol{\xi}_{-\boldsymbol{i}^{(s)}} - \sum_{\mathcal{I} \subseteq \boldsymbol{i}^{(s)}} \widetilde{f}_{\mathcal{I}}$$

denoting with $\boldsymbol{i}^{(s)}$ the multi-index $\{i_1, \ldots, i_s\}$, with $\boldsymbol{\xi}_{\boldsymbol{i}^{(s)}}$ the collection $\{\xi_{i_1}, \ldots, \xi_{i_s}\}$, with $\rho_{\Gamma} = \prod_{i=1}^{n} \rho_{\Gamma_i}$ the multivariate gaussian density of $\boldsymbol{\xi}$, $\Gamma_{-\mathcal{I}} = \times_{i \in \{1,\ldots,N\} \setminus \mathcal{I}} \Gamma_i$ and $\rho_{\Gamma_{-\mathcal{I}}} = \prod_{i \in \{1,\ldots,N\} \setminus \mathcal{I}} \rho_{\Gamma_i}$.

Notice that in the considered case $\Gamma_i = \mathbb{R}$, $i = 1, \ldots, n$, but what follows remains valid for any measurable $\Gamma_i \subseteq \mathbb{R}$. In particular, decomposition (8) is unique whenever $\widetilde{f}$ is integrable over Γ and each term $\widetilde{f}_{\boldsymbol{i}^{(s)}}$ in (8) is orthogonal with respect to the others.

To define the Sobol' indices, call V_f, $V_{\widetilde{f}}$ the variance of f, $\widetilde{f}$ respectively. The Sobol' index relative to the mixed effect $\boldsymbol{\xi}_{\boldsymbol{i}^{(s)}}$ (or to $\boldsymbol{x}_{\boldsymbol{i}^{(s)}}$) is defined as:

$$S_{\boldsymbol{i}^{(s)}} = \frac{1}{V_{\widetilde{f}}} \int_{\Gamma_{\boldsymbol{i}^{(s)}}} \widetilde{f}_{\boldsymbol{i}^{(s)}}^2(\boldsymbol{\xi}_{\boldsymbol{i}^{(s)}}) \rho_{\Gamma_{\boldsymbol{i}^{(s)}}} d\boldsymbol{\xi}_{\boldsymbol{i}^{(s)}}, \tag{9}$$

where $\Gamma_{\boldsymbol{i}^{(s)}} = \Gamma_{i_1} \times \cdots \times \Gamma_{i_s}$. Hence, $S_{\boldsymbol{i}^{(s)}}$ represents the proportion of the total variance explained by $\boldsymbol{\xi}_{\boldsymbol{i}^{(s)}}$ (or $\boldsymbol{x}_{\boldsymbol{i}^{(s)}}$) and thus the Sobol' indices sum to 1.

Concerning the total effect of a parameter $\xi_i(x_i)$ (or of x_i) on the total variability, it can be computed directly from Sobol' indices as:

$$S_i^T = \sum_{\boldsymbol{i}^{(s)} : i \in \boldsymbol{i}^{(s)}} S_{\boldsymbol{i}^{(s)}}. \tag{10}$$

The computational effort needed to compute the sensitivity indices (9) and (10) consists mainly in the cost of calculating (numerically) $2^n - 1$ integrals. To this end, Monte-Carlo quadrature schemes can be used but, due to its low rate of convergence, the computational cost might become unaffordable. However, this problem can be faced and solved by introducing the PCE of the response y, since Sobol' indices can be directly computed from coefficients β_k, $k = 0, \ldots, P-1$. Indeed, expansion (5) can be reordered so that it is equivalent to (8):

$$\widetilde{f}(\boldsymbol{\xi}) = \sum_{\boldsymbol{i}^{(s)} \subseteq \{1,\ldots,n\}} \sum_{k \in K_{\boldsymbol{i}^{(s)}}} \beta_k \Psi_k(\boldsymbol{\xi}), \tag{11}$$

where $K = \{0, \ldots, P-1\}$, $K_{\boldsymbol{i}^{(s)}} = \{k \in K | \Psi_k(\boldsymbol{\xi}) = \Psi_k(\boldsymbol{\xi} = \boldsymbol{\xi}_{\boldsymbol{i}^{(s)}})\}$ (Crestaux et al. 2009).

From the equivalence between (4) and (11) the bijective relation existing between Sobol indices and coefficients β_k, $k = 1, \ldots, n$ can be derived. Indeed, the *PC-based Sobol' indices* can be obtained as:

$$S_{\boldsymbol{i}^{(s)}} = \frac{\sum_{k \in K_{\boldsymbol{i}^{(s)}}} \beta_k^2 \langle \Psi_k, \Psi_k \rangle}{\sum_{k=0}^{P-1} \beta_k^2 \langle \Psi_k, \Psi_k \rangle},$$

being $\boldsymbol{i}^{(s)} \subseteq \{1, \ldots, n\}$, while the total indices can be computed from (10).

4 SIMULATION AND RESULTS

After a preliminary study on the analytical model, the following four input parameters of the model have been considered: the maximal relative gas permeability kr_{max}[-], the reservoir thickness H[m], the reservoir total porosity ϕ [-] and the irreducible water saturation (or maximal gas saturation Snw_{max}[-]). These four parameters form the random vector $\boldsymbol{x}$ appearing in (1): $\boldsymbol{x} = (kr_{max}, H, \phi, Snw_{max})$. In this study the log-normal probability density function is attributed to each parameter.

The response variable which is of interest for the sensitivity analysis is the natural logarithm of the maximal spread of the gas cloud, $\ln(r_{max})$. Once a set of input parameters $\boldsymbol{x}^0 = (kr_{max}^0, H^0, \phi^0, Snw_{max}^0)$ is fixed, the corresponding $\ln(r_{max}^0)$ can be calculated by numerically solving the system of partial differential equation governing the model (Figure 1). The solution includes the saturation 2D profile for any time step: $\ln(r_{max}^0)$ is calculated as the maximum horizontal extension of the 0.99 saturation contour. The numerical solution of the system for a given set of parameters requires 5.67 minutes on a personal computer (4 CPU Xeon E31245 3.30GHz, RAM 16GB).

The experimental design In order to determine the least-square estimate of vector $\boldsymbol{\beta}$ as in (6), an experimental design $\{\ln(r_{max}^j), j = 1, ..., N\}$ of size $N = 100$ has been considered. The number of replications N has been tuned by balancing the computational cost and the precision in least square estimate: the higher N is, the higher either the precision in estimating $\boldsymbol{\beta}$ or the computational cost are. In particular, the computational cost spent for generating an experimental design of size N on a personal computer with the technical characteristic specified before is about $5.67 \cdot N$ minutes ($O(N)$), which corresponds to 9 hours and 27 minutes in the considered scenario.

In order to justify the choice of N, a preliminary study has been performed on the analytical model corresponding to the simplified physical problem presented in (Nordbotten and Celia 2006), for which the exact expression of the response distribution can be computed. In Figure 3 the empirical CDF of the output samples on which the meta-model is built (solid green line) is compared to the exact CDF calculated

Figure 3: Comparison of the meta-model CDF (red line) with the exact CDF (dashed blue line) and with the empirical CDF of the samples used for the meta-model construction (green line). The figures correspond to the analytical solution with varying number of the samples for PCE decomposition, from left to right: 50 samples (first panel), 100 samples (second panel), 1000 samples (third panel).

(a) $N = 50$ (b) $N = 100$ (c) $N = 1000$
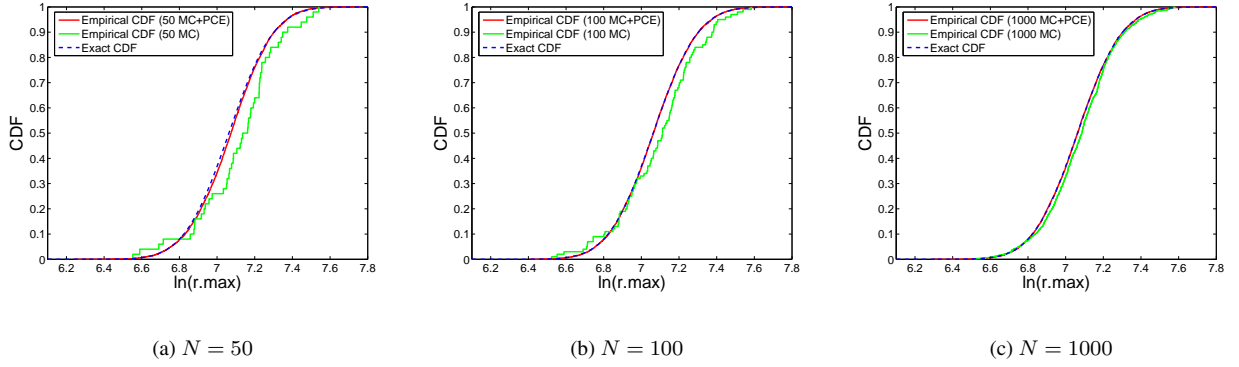
analytically (dashed blue line), for $N = 50, 100, 1000$ (subfigures (a), (b) and (c) respectively). Even the meta-model built out of 50 samples reproduces the response variability quite closely; for $N = 100$ the approximation is very accurate without showing significant gains with respect to $N = 1000$. Therefore, a reasonable value for the parameter N is $N = 100$: this choice seems to balance the most the precision in the CDF estimate and the corresponding computational cost.

Supported by these results, the PCE on the gas injection model –detailed in Section 2– has been performed on the basis of an experimental design of size $N = 100$ (Figure 4). In Figure 4 the $N = 100$ realizations of the (log-transformed) response r_{max} are represented against the parameters kr_{max}, H, ϕ and Snw_{max}, superimposing the corresponding least squares lines. The slope of the regression line is positive in the case of parameters kr_{max}, while it is negative for H, ϕ and Snw_{max}. This result can be expected from the point of view of the physical model: the higher the relative permeability of the gas is, the further it can penetrate, that is why the first correlation is positive. Concerning the other three parameters, the correlation is negative. Indeed, the increase of the reservoir height is not favorable for the horizontal spread. Increasing the porosity and maximal gas saturation leads to a bigger fraction of gas staying in place after the front passage, which means that for the same injected volume the maximal horizontal extension would be smaller.

PCE expansion and sensitivity analysis The coefficients to be included in PCE has been selected following the sparse PCE approach proposed in (Blatman and Sudret 2010, Blatman 2009). Having fixed the maximum degree p of Hermite polynomials to $p = 4$, the 15 coefficients β_0, β_1, β_2, β_3, β_4, β_5, β_7, β_9, β_{15}, β_{18}, β_{20}, β_{38}, β_{40}, β_{48}, β_{58}, proved to be significant for the expansion. Only the first 5 coefficients are reported in Table 1, since the other coefficients proved unstable when replicating the simulations. Noticed that coefficients β_1, β_2, β_3, β_4 reflects the behav-

Table 1: Estimates of the decomposition coefficients

Parameter	Estimate
β_0	7.2157
β_1	0.0387
β_2	-0.0370
β_3	-0.0485
β_4	-0.0475

ior previously noticed by inspecting Figure 4.

Afterward, the estimates of the first two moments of the response probability density function (PDF) have been derived by using (5), obtaining:

$$\mathbb{E}[\ln(r_{max})] \approx 7.2157$$

$$\text{sd}[\ln(r_{max})] \approx 0.0871.$$

These estimates are coherent with a standard Monte Carlo simulation: indeed, 1000 Monte Carlo simulation provides an estimate of 7.2197 for the mean (error: 0.05%), and of 0.0863 for the standard deviation (error: 0.9%)).

In order to assess the uncertainty propagation through the system, Sobol' indices have been computed by using (11) and (10); the results are reported in Table 2. According to Sobol' total indices (Table 2, second column), the parameters which seem to influence the most the response variability are the reservoir total porosity Φ and the maximal gas saturation Snw_{max}, whose variability explains about 31.0% and 29.8%, respectively, of the output variability. The remaining variability is explained in decreasing order of importance by maximal relative gas permeability kr_{max} (20.2%) and reservoir thickness H (18.2%). Very similar conclusions can be drawn by considering the Sobol' indices reported in the first column of Table 2, meaning that the influence of the input parameters on the output is mainly due to single effects, while mixed effects play a minor role.

To access the uncertainty in the estimated statistics, the simulation has been replicated with the same settings starting from different seeds. The results are

Figure 4: Experimental design. From left to right: plot of r_{max} against kr_{max} (first panel), H (second panel), ϕ (third panel) and Snw_{max} (fourth panel).
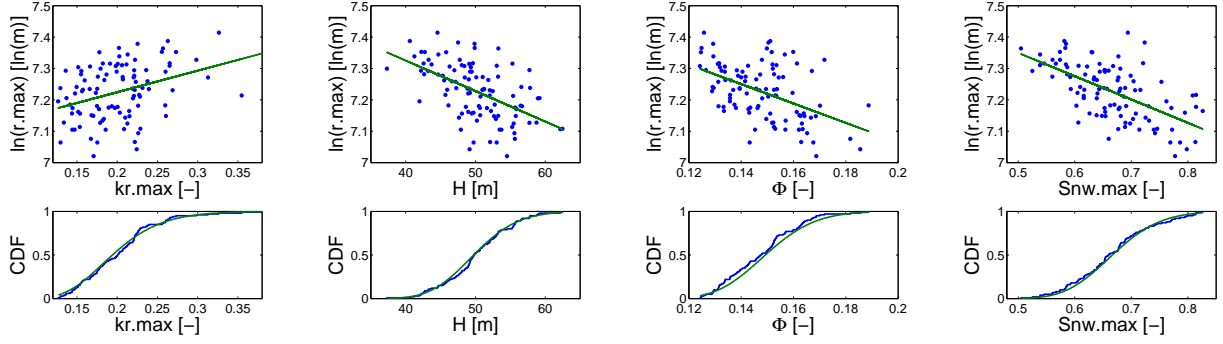

Table 2: Sobol' indices

Parameter	Sobol' index	Sobol' total index
kr_{max}	0.2019	0.2098
H	0.1823	0.1884
Φ	0.3102	0.3177
Snw_{max}	0.2977	0.2990

Table 3: Mean, standard deviation and coefficient of variation relative to the estimated indices.

	Mean	Sd	C.V.
$\mathbb{E}[\ln(r_{max})]$	7.2151	0.0020	2.8e-4
$\mathrm{sd}[\ln(r_{max})]$	0.0861	8.5e-4	0.0099
$S_{kr_{max}}$	0.2026	0.0087	0.0431
S_{H}	0.1640	0.0179	0.1091
S_{Φ}	0.3119	0.0104	0.0335
$S_{Snw_{max}}$	0.3149	0.0214	0.0681
$S_{kr_{max}}^{T}$	0.2079	0.0082	0.0395
S_{H}^{T}	0.1701	0.0236	0.1384
S_{Φ}^{T}	0.3159	0.0100	0.0318
$S_{Snw_{max}}^{T}$	0.3168	0.0192	0.0608

reported in Table 3. It can be noticed that the response mean and standard deviation seem precisely estimated. Instead, the parameters affected by the most significant variability are the Sobol' indices relative to the reservoir thickness H and to the maximal gas saturation Snw_{max} –either the indices relative to the single effects or the total indices–. Notice also that the coefficients β_k selected by the sparse PCE approach varied by replicating the simulation, with only the first five coefficients steadily selected. However, both the variability in the PCE estimates and in the selected coefficient do not seem to significantly affect the conclusions previously drawn about the uncertainty propagation through the system.

Assessing the response distribution function Although Sobol' indices provide an important tool for sensitivity analysis, in this study the whole CDF is
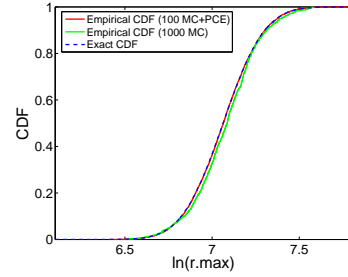

Figure 5: Comparison of the meta-model CDF obtained with an experimental design of size $N = 100$ (red line) with the exact CDF (dashed blue line) and with the empirical CDF obtained with 1000 standard MC-simulations (green line)
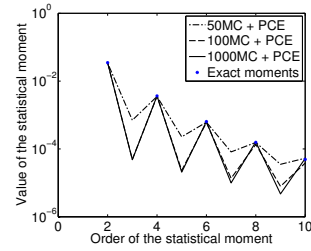

Figure 6: Comparison of the statistical moments obtained through a meta-model constructed on $N = 50, 100, 1000$ experiments(black lines), with the exact values of the statistical moments (blue stars).

of interest. Generally, CDF can be estimated in three ways: (a) analytical calculation (b) standard MC simulation (c) meta-model MC simulation –i.e., generation of an experimental design of size N by a standard MC-simulation on the complete model, computation of the β coefficients in approximation (5) and finally MC simulation on the meta-model exploiting the same approximation–. The main drawback of method (a) is that an analytic form of response CDF is rarely available; on the other hand the computational cost of a standard MC simulation on a complex model as the one considered here increases fast with the number of samples required (more than 94 hours for 1000 samples). Notice in particular that a preliminary MC simulation is needed when using method (c), but a fair size of the experimental design is typically much smaller than the number of simulations required to assess the complete CDF with method (b).

To evaluate the performance of methods (b) and (c) when the analytical form of the CDF is available, the simplified gas-injection model has been again considered. In Figure 5 the following CDFs are compared: 1) exact CDF –method (a), dashed blue line–; 2) empirical CDF obtained by 1000 standard MC-simulations –method (b), green line–; 3) empirical CDF obtained by 10000 PCE-based meta-model simulations having

an experimental design of size 100 –method (c), red line–. The parameters of the latter simulation have been chosen according to the feasibility of an analogous simulation on the complete gas injection model: indeed, the computational cost of the MC-simulation on the complete model soon becomes unaffordable as the sample size increases and the simulation of 1000 samples already requires more than 94 hours[1].

The empirical CDFs obtained with the two methods are pretty similar to the exact CDF, but more accurate and smooth results are obtained for PCE-based simulations, being based in fact on a rich meta-model MC sample. Figure 6 represents the statistical moments of the response. Notice that the meta-model reproduces exactly the even statistical moments even when it is build on a small size of the experimental design. Theoretically being zero, the odd statistical moments seem to be better approximated for meta-models based on a larger size of the experimental design. Observe that the time consumption required for computing the meta-model estimate is essentially due to the computation of the experimental design. Therefore the very low additional cost required for the PCE step turns in a much more accurate CDF estimate.

Supported by these results, methods (b) and (c) have been applied on the full numerical model of gas-injection detailed in Section 2. The response empirical CDF computed by 1000 standard Monte Carlo simulations (blue dotted line) and with a PCE-based meta-model simulation (red line) are presented in the left panel of Figure 4, superimposed to the results of the replicated simulations (grey lines). It has to be noticed that the results are fairly stable among the replicates, as the lines seems perfectly superimposed. This supports the conclusion previously drawn about the stability of the results with respect to the variability in the selected coefficients and their estimates. Moreover, all the empirical CDF are very close to the MC results obtained from the full model.

In order to appreciate the most the existing differences among the obtained estimates, the right panel of Figure 4 reports the kernel density PDF estimates computed as before from the MC sample, the meta-model sample and the replicates. Replicates have been built either with the same parameters (darker lines), or by fixing the maximum degree of Hermite polynomials to $p = 5$. Although some differences are now appreciable, all the meta-model curves appear very similar each other and very close to the MC estimates. The main difference is present in the right tail of the distribution, where a slight underestimation occurs. However, the meta-model PDF estimates provides a very good approximation of the MC PDF estimates obtained in less than 10 hours against the 94 hours required by the full MC-simulation.

Finally, Figure 8 shows the statistical moments of the resulting response distributions. All the replicates

Figure 7: Response CDF (left panel) and kernel density estimated PDF (right panel) for: 1000 standard Monte Carlo simulation (dotted blue line), meta-model simulation through PCE (solid red line), replicated meta-model simulation (dark grey lines).
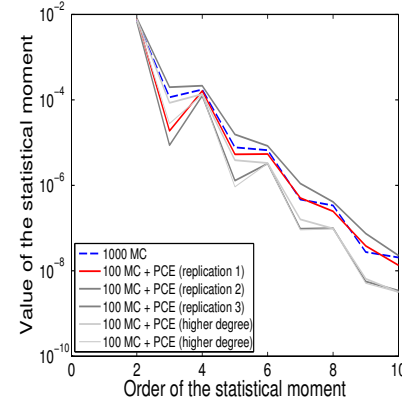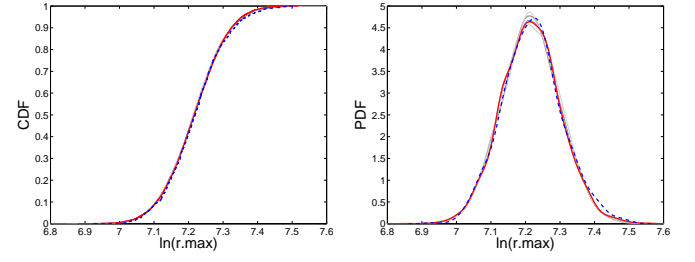

Figure 8: Comparison of the statistical moments for: 1000 standard Monte Carlo simulation (dotted blue line), meta-model simulation through PCE (solid red line), replicated meta-model simulations when the maximum degree of the Hermite polynomials is $p = 4$ (light grey lines) and $p = 5$ (solid grey lines).

seem to provide similar results, that are close to the MC estimate. No significant gains are reached if increasing the maximum degree of the Hermite polynomials. The first two moments of the distribution, in particular the mean, are very precisely estimated, as already noticed when analyzed Table 3. Increasing uncertainty is observed when increasing the order of the estimated moments. As in the simplified model, the uncertainty in the estimates is higher for even order moments, while odd moments are more accurately estimated.

5 CONCLUSION

In this work, the Polynomial Chaos Expansion method as an efficient technique for performing uncertainty and sensitivity analysis of complex numerical models has been investigated. In particular, the potentiality of this approach has been explored in terms of computational efficiency when dealing with a model of gas injection into porous media. The regularity of the response for this particular physical model is favorable for the use of the described method.

The crucial points encountered while performing PCE on a real system have been underlined. For instance, one need to pay close attention when making the choice of the size of the experimental design used for estimating the expansion coefficients and the basis functions form. The minimal size is dictated by the number of coefficients in the decomposition, further validation of the meta-model quality can only

---

[1]4 CPU Xeon E31245 3.30GHz, RAM 16GB

be done by experience in every particular case. Here, the size of the experimental design has been fixed by exploiting the information coming from the application of the same technique to a simplified model, but other considerations –e.g., in terms of variance of the estimate– can be made.

A great deal of attention has been focused on the computational gain of the PCE approach, which is very significant for the considered application: while obtaining the response PDF by 1000 standard Monte Carlo simulation takes more than 94 hours, with a PCE approach the computational cost is reduced to less than 10 hours.

However, the robustness of the procedure with respect to the size of the experimental design would be worth being further investigated. Indeed, by repeating the simulations with a different size of the experimental design, the coefficients estimate might significantly vary, with possible effects on the Sobol' indices and on the consequent sensitivity analysis. Although this does not seem to be the case, in order to avoid this problem, a direct control on the estimates variance should be considered, in particular if no 'a priori' information is available on the response distribution.

A further development regarding the considered physical model, which might be of industrial and physical importance, is to consider a space-time varying response, such as the water and gas pressure, while analyzing the propagation of the uncertainty from the input parameters to the space-time random field described by the response. In such a complex setting, the use of PCE could make the computational cost of performing uncertainty and sensitivity analysis affordable, which is a very important goal to be reached for this kind of complex numerical systems.

6 ACKNOWLEDGMENT

REFERENCES

Blatman, G. (2009). *Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis.* Ph.D Thesis, University Blaise Pascal, Clermont-Ferrand.

Blatman, G. & B. Sudret (2010). Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering and System Safety 95*(11), 1216 – 1229.

Chorin, A. J. (1974). Gaussian fields and random flow. *Journal of Fluid Mechanics 63*, 21–32.

COMSOL (2012).

Crestaux, T., O. L. Maître, & J. Martinez (2009). Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering and System Safety 94*, 1161–1172.

Crow, S. C. & G. H. Canavan (1970). Relationship between a Wiener-Hermite expansion and an energy cascade. *Journal of Fluid Mechanics 41*, 387–403.

Formaggia, L., A. Guadagnini, I. Imperiali, V. Lever, G. Porta, M. Riva, A. Scotti, & L. Tamellini (2012). Global sensitivity analysis through polynomial chaos expansion of a basin-scale geochemical compaction model. *Computational Geosciences*, 1–18. 10.1007/s10596-012-9311-5.

Ghanem, R. & P. Spanos (1991). *Stochastic finite elements: a spectral approach.* Springer-Verlag, New York.

Maître, O. L. & O. Knio (2010). *Spectral methods for uncertainty quantification.* Scientific Computation. Springer, New York. With applications to computational fluid dynamics.

Meecham, W. C. & D.-T. Jeng (1968). Use of the Wiener-Hermite expansion for nearly normal turbulence. *Journal of Fluid Mechanics 32*, 225–249.

Meecham, W. C. & A. Siegel (1964, August). Wiener-Hermite Expansion in Model Turbulence at Large Reynolds Numbers. *Physics of Fluids 7*, 1178–1190.

Nordbotten, J. M. & M. A. Celia (2006, August). Similarity solutions for fluid injection into confined aquifers. *Journal of Fluid Mechanics 561*, 307–327.

Oladyshkin, S., H. Class, R. Helmig, & W. Nowak (2011, June). An integrative approach to robust design and probabilistic risk assessment for CO_2 storage in geological formations. *Computational Geosciences 15*(3), 565–577.

OpenTURNS (2012).

Orszag, S. A. & L. R. Bissonnette (1967, December). Dynamical Properties of Truncated Wiener-Hermite Expansions. *Physics of Fluids 10*, 2603–2613.

Shinozuka, M. & G. Deodatis (1988). Response variability of stochastic finite element systems. *Journal of Engineering Mechanics 114*(3), 499–519.

Siegel, A., T. Imamura, & W. C. Meecham (1965, May). Wiener-Hermite Expansion in Model Turbulence in the Late Decay Stage. *Journal of Mathematical Physics 6*, 707–721.

Sobol, I. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Modelinf comput. Experiment 1*(4), 407–414.

Soize, C. & R. Ghanem (2005, February). Physical systems with random uncertainties: Chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput. 26*(2), 395–410.

Sudret, B. (2008, July). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety 93*(7), 964–979.

Sudret, B. & A. Der Kiureghian (2000). Stochastic finite elements and reliability: a state-of-the-art report. Technical report, University of California, Berkeley. Technical report no. UCB/SEMM-2000/08.

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal 44*(5), 892–898.

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

Wiener, N. (1962). *Nonlinear Problems in Random Theory.* MIT.

Xiu, D. & G. Karniadakis (2002). The wiener-askey polynomial chaos for stochastic differential equations. *Journal of Scientific Computing 26*.

Xiu, D., D. Lucor, C.-H. Su, & G. E. Karniadakis (2002). Stochastic modeling of flow-structure interactions using generalized polynomial chaos. *Journal of Fluids Engineering 124*(1), 51–59.