XVIII International Conference on Water Resources
CMWR 2010
J. Carrera (Ed)
CIMNE, Barcelona 2010

PROBABILISTIC COLLOCATION FOR EFFICIENT UNCERTAINTY ANALYSIS IN GROUNDWATER FLOW

V. Fontaine, M. A. Mamode and T. A. Mara

LPBS - EA 4076, Department of Physics, University of La Réunion,
15 avenue René Cassin, 97715 La Réunion, France
e-mails: {thierry.mara;vincent.fontaine;malik.mamode}@univ-reunion.fr

Key words: Stochastic groundwater flow, probabilistic collocation, polynomial chaos expansion, Karhunen–Loève expansion, global sensitivity analysis.

Summary. Assessment of the effects of uncertainty in model inputs on its output are now widely recognized as important parts of analyses for complex systems. In this paper, we address stochastic groundwater flow problems with random permeability fields. For this purpose, we combine low order mixed finite element method space with polynomial chaos expansions. The so-called Karhunen–Loève expansion is employed to efficiently generate the random fields. A non-intrusive method based on probabilistic collocations is used to compute the polynomial coefficients. The computational cost is reduced by preliminarily *screening* the input random variables. The stochastic error analysis is investigated through numerical experiments.

1 Problem statement

Let $\mathcal{D}$ be a bounded domain in $\mathbb{R}^2$ with boundary Γ. We define the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where Ω is a space of events, $\mathcal{F}$ its σ-algebra and $\mathcal{P}$ its probability measure. Here we consider stochastic groundwater flow problems with random permeability fields defined as follows: Find stochastic functions $u : \mathcal{D} \times \Omega \to \mathbb{R}$ and $\underline{v} : \mathcal{D} \times \Omega \to \mathbb{R}^2$ such that almost surely, the following set of equations holds:

$$\begin{aligned} \underline{v}(\underline{x};\omega) = -\alpha(\underline{x};\omega)\nabla u(\underline{x};\omega) \quad \text{and} \quad \nabla\cdot\underline{v}(\underline{x};\omega) &= f(\underline{x}) \quad \text{in} \quad \mathcal{D}, \\ u(\underline{x};\omega) &= g(\underline{x}) \quad \text{on} \quad \Gamma, \end{aligned} \tag{1}$$

where u is the pressure, $\underline{v}$ the Darcy' velocity and α the given stochastic permeability field.

Different approaches are available to tackle problem (1) that can be classified as either intrusive or non-intrusive. In the first approach, a set of equations is derived that directly accounts for the response uncertainty (e.g. its first moments) as unknowns of the problem. The advantage is that the numerical solution of the stochastic problem can be computed with a few simulation runs. Yet, the number of unknowns drastically increases with the complexity of the relationship between the model response and the random inputs. Their intrusive feature greatly complicates the formulation of the model equations even in the case of linear and stationary input distributions. The stochastic finite element methods [1;2] and moment-based equations [3] are such intrusive approaches. In the non-intrusive approach, the numerical model derived to solve the deterministic problem is used as a black-box. The response uncertainty is assessed through several model runs for different input sets. The price to pay in this case is the rather large number of model runs required to accurately assess the output uncertainty [4].

Here, we solve the deterministic problem with the lowest-order Raviart–Thomas mixed finite element methods. This choice is motivated by their local mass conservation and proper treatment of discontinuous permeability tensor. In order to improve the efficiency of non-intrusive methods, we focus on the probabilistic collocation method suggested in a recent article [5] in conjunction with polynomial chaos (PC) expansions. Given that the computational cost increases drastically with the number of random variables, we propose to preliminarily reduce the input random variables after analyzing the sensitivity indices of the eigenmodes. Then, the stochastic fields (pressure and variance) are accurately estimated with the few relevant eigenmodes. Numerical tests are proposed to assess the performance of the proposed approach.

2 Input uncertainty modelling

In this section, we address the problem of propagating the input uncertainty through the model (Eq. 1). This is efficiently achieved with the Karhunen–Loève (KL) expansion that led Ghanem & Spanos to develop their spectral approach [1]. We denote by $\kappa_M(\underline{x};\omega) = log(\alpha(\underline{x};\omega))$ the approximated log-permeability field. We assume that the log-permeability field is gaussian and characterized by a mean $\bar{\kappa}(\underline{x})$ and a covariance kernel $C_\kappa(\underline{x},\underline{x}')$, bounded, symmetric and positive definite. The truncated KL expansion writes,

$$\kappa_M(\underline{x};\omega) = \bar{\kappa}(\underline{x}) + \sum_{i=1}^{M} \sqrt{\lambda_i}\xi_i(\omega)\varphi_i(\underline{x}), \tag{2}$$

where λ_i and $\varphi_i(\underline{x})$ are called eigenvalues and eigenfunctions of $C_\kappa(\underline{x},\underline{x}')$ respectively. This decomposition is optimal in the sense that the mean square error, integrated over $\mathcal{D}$, is minimized. The deterministic eigenfunctions are orthogonal. $\underline{\xi} = (\xi_1,\ldots,\xi_M)^T$ is a set of independent standard normal random variables. In our applications, we retain the first M eigenmodes that contain at least 95% of the variance. Eigenvalues and eigenfunctions of $C_\kappa(\underline{x},\underline{x}')$ are solution of the following Fredholm equation:

$$\int_{\mathcal{D}} C_k(\underline{x},\underline{x}')\varphi_i(\underline{x}')d\underline{x}' = \lambda_i\varphi_i(\underline{x}). \tag{3}$$

Computational issue for numerically solving Eq. (3) is adressed in Schwab & Todor (2006) [6]. Once the eigenpairs determined, to propagate the random field uncertainty through the model, one needs to generate M independent standard normal random variables (i.e. the ξ_i's in Eq. 2). This can be achieved with Monte Carlo (or Quasi Monte Carlo) samples or specific collocation points.

3 Output uncertainty and sensitivity analyses

Response surface methodologies are suited for output uncertainty representation. They consists in building an approximate of the model response called response surface [7], metamodel or emulator[8]. Generalized polynomial chaos expansions are such particular response surfaces.

Let $F(\underline{\xi})$ be a square integrable function of M independently and normally distributed random variables. Then, it can be spanned onto the Hermite PC as follows,

$$F(\underline{\xi}) = \sum_{k \in \mathbb{N}^M} F_k \psi_k(\underline{\xi}), \tag{4}$$

where $\{\psi_k\}$ denotes a Hilbertian basis of $L^2(\Omega, \mathcal{F}, \mathcal{P})$ defined by the so called Hermite PC. The deterministic PC coefficients $\{F_k\}$ are to be computed. Their knowledge provides a complete characterization of the uncertainty of F. In practice, the PC expansion is truncated to a given polynomial degree (say P). The total number of PC coefficients ($N + 1$) increases drastically with the dimension of the model:

$$N + 1 = \frac{(M + P)!}{M!P!}. \tag{5}$$

Different computational methods are proposed in the literature to efficiently compute the PC coefficients. In particular, the collocation method proposed in a recent paper [5] requires a number of model runs of the same order than ($N + 1$). The latter is employed in the present work.

In the cited article, the author noticed that PC expansion is equivalent to Sobol' ANOVA-decomposition [9] and proposed to use PC expansions to compute the variance-based sensitivity indices. The computation is straightforward once computed the PC coefficients (the interested readers are referred to[5] for more details on the calculaton). In the present work, we investigate the total sensitivity indices defined by,

$$S_{T_i} = \frac{E(V(F(\underline{\xi}) | \underline{\xi}_{-i}))}{V(F(\underline{\xi}))}, \tag{6}$$

where $E(\cdot)$ is the mathematical expectation, $V(\cdot)$ the variance, $V(\cdot|\cdot)$ is the conditional variance and $\underline{\xi}_{-i} = \{\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_M\}$. S_{T_i} measures the total contribution of ξ_i to the variance of $F(\underline{\xi})$ including its marginal as well as its cooperative contributions with the other inputs. Hence, if $S_{T_i} \simeq 0$ then ξ_i can be deemed as a non important input and its value can be arbitrarily fixed within its uncertainty range without modifying $F(\underline{\xi})$. In this work, we propose to further reduce the number of eigenmodes M in the KL expansion (Eq. 2) by retaining only the most important ones. By so doing, we also reduce the number of PC coefficients (Eq. 5) and consequently the computatonal cost when probabilistic collocation is used.

4 Optimal probabilistic collocation

Probabilistic collocation is a method to propagate input uncertainty through the model. Contrarily to Monte Carlo simulations, it only relies on specific design points $\{\underline{\xi}^j\}$ in the input space. The number of points N_s should at least equal the number of coefficients in the truncated PCE (i.e. $(N+1)$ see Eq. 5). If a P-th order PC is investigated, the collocation points are chosen as combination of the roots of the polynomial of order $(P+1)$. In the aforementioned article [5], it is proposed to select the combinations the closest to the origin, so that the rank of the Fisher matrix $(\underline{\underline{\Psi}}^T\underline{\underline{\Psi}})^{-1}$ equals $(N+1)$. Then, the PC coefficients are obtained by regression,

$$\underline{\tilde{F}} = (\underline{\underline{\Psi}}^T\underline{\underline{\Psi}})^{-1}\underline{\underline{\Psi}}^T\underline{Y}, \tag{7}$$

where $(\underline{Y})_{|i} = F(\underline{\xi}^i)$ and $(\underline{\underline{\Psi}})_{|ij} = \underline{\psi}_j(\underline{\xi}^i)$ with $i = 1, \ldots, N_s$ and $j = 0, \ldots, P$, $\underline{\tilde{F}}$ is the vector of estimated PC coefficients. The determination of the optimal design points may be time consuming but is independent of the model of interest F and $N_s \simeq N+1$. Consequently, the experimental design is derived once for all and could serve for the analysis of other models.

5 Numerical experiments

The deterministic problem is solved with the mixed hybrid finite elements (MHFE) method by using the lowest-order Raviart–Thomas (RT_0) space [10]. Let $\mathcal{D}$ be the unit square $[0,1]^2$ and we denote by $\mathcal{T}_h$ a partition of $\mathcal{D}$ into triangles. In our experiments, a fixed mesh is employed which is composed of 22190 triangles. The boundary conditions are defined as follows: $g(0,y) = 1$ and $g(1,y) = 0$ with no-flow elsewhere. The source term is null in the physical domain. A preconditioned conjugate gradient iterative solver with the Eisenstat [11] procedure is used for the resolution of the sparse linear system. The error is computed in the discrete L^2-norm ,

$$\|u\|_0 = \left[\sum_{A\in\mathcal{T}_h} |A| u(x_A)^2\right]^{1/2}, \tag{8}$$

where x_A is the center of A and $|A|$ its measure.

We choose the covariance kernel so that only a few eigenmodes are preponderant ($M = 4$). For this purpose, we assume Gaussian random log-permeability fields characterized by a mean, $\bar{\kappa}(\underline{x}) = 0$ and a covariance function of the form $C_\kappa(\underline{x}, \underline{x}') = \exp(-|\underline{x} - \underline{x}'|/10)$.

The algorithm to perform the uncertainty and sensitivity analyses is as follows,

- Compute and select the M first eigenvalues and eigenfunctions of $C_\kappa(\underline{x}, \underline{x}')$: λ_i and $\varphi_i(\underline{x})$, $i = 1, \ldots, M$,
- Choose the PC order P and generate a sample of size N_s of M independent standard normal random variables (the $\underline{\xi}^i$'s) with the probabilistic collocation design,
- For each collocation point, compute $\kappa_M(\underline{x}, \underline{x}')$ the log-permeability field with Karhunen-Loève expansion (Eq. 2) and deduce the permeability field $\alpha(\underline{x};\omega) = exp(\kappa_M(\underline{x};\omega))$,

- For each permeability field so generated, perform the simulation run by solving the *deterministic* PDE associated (Eq. 1) and save the outcomes of interest,

- Once the simulation runs achieved, for each mesh and for each outcome, compute the PC coefficients and deduce the expected and variance fields of the outcomes as well as the total sensitivity fields of each random variable $\{\xi_1, \ldots, \xi_M\}$.

Order	Coll. points	Pressure error		Velocity error	
P	N	Mean	Variance	Mean	Variance
2	81	1.88E-07	9.13E-07	1.01E-02	1.03E-00
4	625	4.08E-09	3.41E-08	1.60E-03	8.22E-02
6	2401	1.29E-09	5.15E-09	8.00E-04	1.24E-02

Table 1: Stochastic convergence results in the discrete L^2-norm for the full collocation (FC).

Order	Coll. points	Pressure error		Velocity error	
P	N	Mean	Variance	Mean	Variance
2	15	1.88E-07	1.19E-06	1.04E-02	1.19E-00
4	70	1.72E-09	4.52E-09	3.00E-04	1.09E-01
6	210	6.91E-10	1.59E-09	1.00E-04	5.40E-03

Table 2: Stochastic convergence results in the discrete L^2-norm for the optimal collocation (OC).

5.1 Assessment of the stochastic convergence

In the first numerical experiment, we compare the output uncertainties obtained with two PCE expansions computed with a full tensor product grid of probabilistic collocation points (the computational cost is $N_s = (P+1)^M$) and the probabilistic collocation method of Sudret[5] ($N_s \simeq \frac{(P+M)!}{P!M!}$). In the sequel, the first collocation technique is referred to as the full collocation (FC) while the second one is referred to as the optimal collocation (OC). We focus on the ability of OC to provide accurate estimates of the two first moments (mean and variance) as compared to FC. For this purpose, we perform different computational experiments to estimate PCE's of increasing order with the two methods. Because, with the regression-based approach (Eq. 7), the errors decrease monotonically for even and odd P respectively, we only investigate even PC order ($P = 2, 4, 6, 8$). We study the stochastic convergence of the two PCE's estimates toward the results of the 8-th order PC computed with FC.

Inspection of the stochastic fields obtained with the 8-th order PC estimated with FC ($N = 6561$) has shown that: (i) the pressure variance is largest in a vertical strip in the middle of the domain, away from the Dirichlet boundary edges (see Figure 1), (ii) the velocity variance is smallest along the Neumann edges and is affected by the direction of the flow. For this reason, we exclude the vertical component velocity in the analysis of convergence.

The stochastic errors are reported in Tables 1 & 2. We can note the exponential convergence as confirmed by other authors[12;4]. The results with OC are particularly accurate as regard to the computational cost. A 6-th order PCE seems sufficient to properly capture the output uncertainty. So, for the sensitivity analysis, we consider the 6-th order PC computed with OC.

5.2 The most relevant eigenmodes

We compute the sensitivity maps of each eigenmode for each outcome. As aforementioned in § 3, their computation is straightforward once estimated the PC coefficients. For the pressure and velocitiy fields, we investigate the eigenmodes that mainly contribute to the variance. For this purpose, we set that an eigenmode is important if its total sensitivity indices over the domain is greater than 5%. We find that for the pressure only mode 3 is relevant and for the velocity modes 1 and 2. This means that, the stochastic outcomes can be characterized by a 6-th order PCE with 2 random variables. This leads to a collocation design of size $N_s \simeq \frac{(2+6)!}{2!6!} = 28$ (actually 31). As a comparison, Figure 1 depict the stochastic fields of the 6-th order PC computed with the four eigenmodes ($N_s = 210$) and three eigenmodes ($N_s = 31$). As expected, the results are very close.

5.3 Discussion

The numerical experiments show that the eigenmodes that mainly contribute to the variance of the pressure field differ from those that drive the velocity variance. This is explained by the fact that streamlines and isopressure curves are orthogonal as well as the eigenfunctions. Consequently, one can use global sensitivity analysis to point out the most relevant eigenmodes for each outcome. This allows to reduce the stochastic dimension of the problem and consequently the computational cost. In the case of covariance function that implies much more eigenmodes in the Karhunen-Loève expansion it is recommended to first investigate a second-order PCE in order to detect the most relevant eigenmodes. Then, with the few important eigenmodes, one can investigate higher-order PCE that should provide more accurate stochastic fields.

REFERENCES

[1] R. Ghanem and P.D. Spanos, *Stochastic finite elements: A spectral approach*, Springer, New York, (1991).

[2] H.G. Matthies and A. Keese, Galerkin methods for linear nonlinear elliptic stochastic partial differential equations, *Comput. Meth. Appl. Mech. Eng.*, **194**, 1295–1331 (2005).

[3] D. Zhang and Z. Lu, An efficient, high-order perturbation approach for flow in random porous media via Karhunen–Loève and polynomial expansions, *J. Comput. Phys.*, **194**, 773–794 (2004).

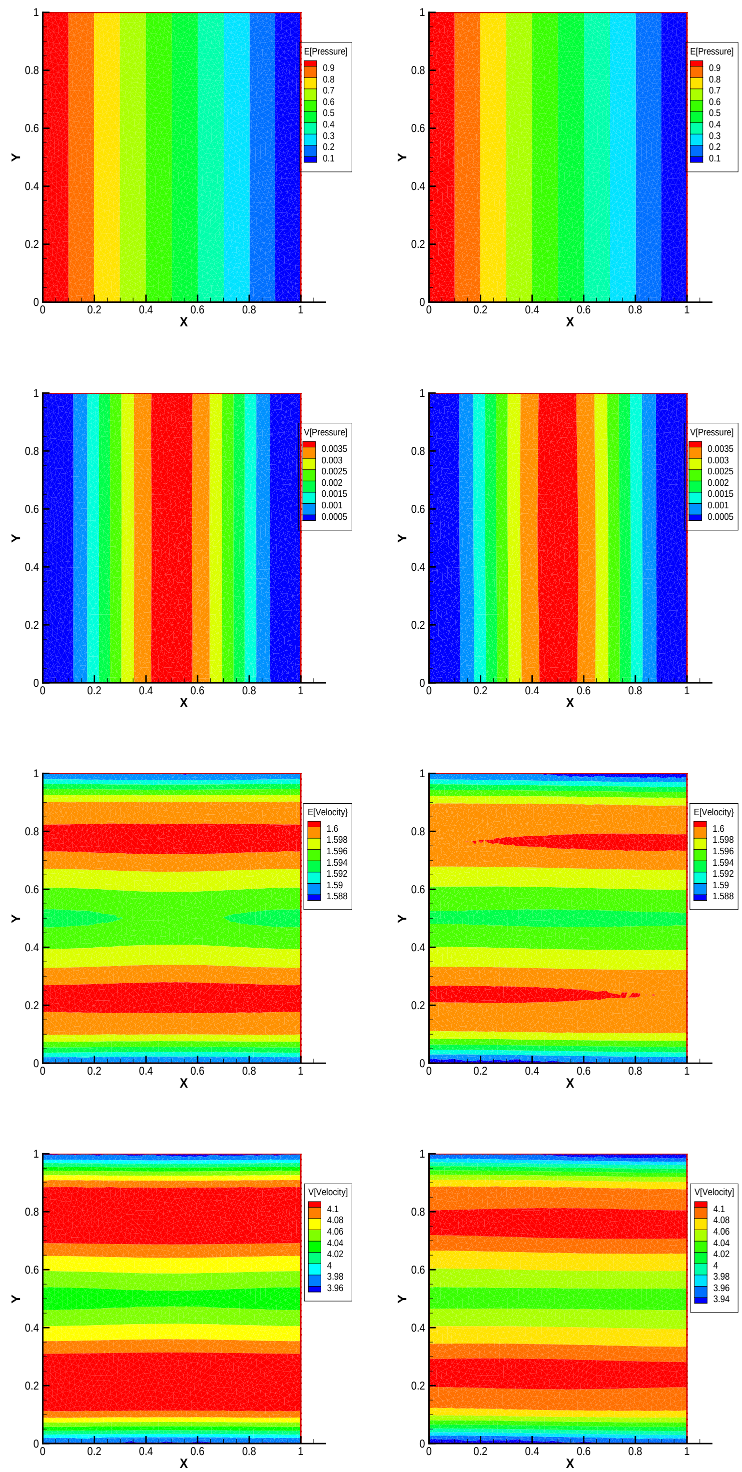

Figure 1: Expected and variance of the pressure and velocity fields computed with a PCE of order $P = 6$ and dimension $M = 4$ (left) and a PCE with $P = 6$ and $M = 2$ (right). The latter is built with the most relevant eigenmodes.

[4] B. Ganis, H. Klie, M.F. Wheeler, T. Widley, I. Yotov and D. Zhang, Stochastic collocation and mixed finite elements for flow in porous media, *J. Comput. Phys.*, **197**, 3547–3559 (2008).

[5] B. Sudret, Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.*, **93**, 964–979 (2008).

[6] C. Schwab and R.A. Todor, Karhunen–Loève approximation of random fields by generalized fast multipole methods, *J. Comput. Phys.*, **217**, 100–122 (2006).

[7] G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for experimenters. An introduction to Design, Data analysis and Model building*, Wiley, New York, (1978).

[8] J.E. Oakley and A. O'Hagan, Probabilistic sensitivity analysis of complex models: a Bayesian approach, *J. Roy. Statis. Soc. B*, **66**, 751–769 (2004).

[9] I.M. Sobol, Sensitivity analysis for non linear mathematical models. *Math. Mod. Comput. Exp.*, **1**, 407–414 (1993).

[10] F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, Springer-Verlag, New York, Vol. 15 (1991).

[11] S.C. Eisenstat, Efficient implementation of a class of preconditioned conjugate gradient methods, *SIAM J. Sci. Statist. Comput.*, **2**, 1– 4 (1981).

[12] I. Babuska, F. Nobile and R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, **45**, 1005–1034 (2007).