Polynomial Chaos Expansion for Global Sensitivity Analysis applied to a model of radionuclide migration in randomly heterogeneous aquifers

Valentina Ciriello [a,*], Vittorio Di Federico [a], Monica Riva [b], Francesco Cadini [c], Jacopo De Sanctis [c], Enrico Zio [c, d], and Alberto Guadagnini [b]

[a] *Dipartimento di Ingegneria Civile, Ambientale e dei Materiali, Università di Bologna, Italy.*

[b] *Dipartimento di Ingegneria Idraulica, Ambientale, Infrastrutture Viarie, Rilevamento, Politecnico di Milano, Milano, Italy.*

[c] *Dipartimento di Energia, Politecnico di Milano, Milano, Italy.*

[d] *Chair on Systems Science and the Energetic challenge,*

European Foundation for New Energy- Électricité de France

Ecole Centrale Paris and Supelec.

[*]Corresponding author: Tel.: +39 051 2093753; Fax: +39 051 2093263.

E-mail: valentina.ciriello3@unibo.it

Abstract

We perform Global Sensitivity Analysis (GSA) through Polynomial Chaos Expansion (PCE) on a contaminant transport model for the assessment of radionuclide concentration at a given control location in a heterogeneous aquifer, following a release from a near surface repository of radioactive waste. The aquifer hydraulic conductivity is modeled as a stationary stochastic process in space. We examine the uncertainty in the first two (ensemble) moments of the peak concentration, as a consequence of incomplete knowledge of (a) the parameters characterizing the variogram of hydraulic conductivity, (b) the partition coefficient associated with the migrating radionuclide, (c) the effective dispersivity at the scale of interest. These quantities are treated as random variables and a variance-based GSA is performed in a numerical Monte Carlo framework. This entails solving groundwater flow and transport processes within an ensemble of hydraulic conductivity realizations generated upon sampling the space of the considered random variables. The Sobol indices are adopted as sensitivity measures to provide an estimate of the role of uncertain parameters on the (ensemble) target moments of the variable of interest. The calculation of the indices is performed by employing PCE as a surrogate model of the migration process to reduce the computational burden. We show that the proposed methodology (a) allows identifying the influence of uncertain parameters on key statistical moments of the peak concentration (b) enables extending the number of Monte Carlo iterations to attain convergence of the (ensemble) target moments and (c) leads to considerable saving of computational time while keeping acceptable accuracy.

Keywords: Performance assessment, radionuclide migration, heterogeneous aquifers, Global Sensitivity Analysis, Sobol indices, Polynomial Chaos Expansion.

1. Introduction

Performance assessment of radioactive waste repositories aims at evaluating the risk of groundwater contamination due to potential release of radionuclides. Modeling the whole chain of processes involved in this analysis is extremely challenging and requires employing highly complex theoretical and numerical models to couple radionuclide migration within the repository and in the groundwater environment. Uncertainty associated with, e.g., incomplete knowledge of initial and boundary conditions, nature and structure of the groundwater system and related key parameters must be added to the list of difficulties (e.g., Tartakovsky 2007; Winter 2010; Volkova et al. 2008 and references therein).

We consider the analysis of the uncertainty associated with the first two (statistical) moments of the peak solute concentration detected at a given location and time. The source of uncertainty is incomplete/imprecise knowledge of the values of the hydrogeological parameters characterizing the system (Rubin 2003; Zhang 2002). For a rational management of the uncertainty analysis, we use Global Sensitivity Analysis (GSA) to obtain information on the relative effects of the uncertain input parameters on the model outputs (Saltelli et al. 2000). In particular, we resort to variance-based methods, which can provide a comprehensive view on the uncertainty and allow identifying the relative and joint contributions of the uncertain input parameters to the uncertainty (variance) of the model outputs (Archer et al. 1997).

Within variance-based GSA, the Sobol indices are widely used as sensitivity metrics (Sobol 1993), because they do not require any assumption of linearity in the interpretive model adopted. Their estimation is traditionally performed by Monte Carlo (MC) sampling (Sobol 2001). The sample size needed to attain statistical convergence of the Monte Carlo estimates can be rather large, depending on the complexity and dimension (number of uncertain input parameters) of the problem (e.g., Ballio and Guadagnini 2004, Zhang et al. 2010, and references therein). This might result in a serious and sometimes unsustainable computational burden in cases where repeated high-resolution simulations of the model are required (Sudret 2008).

Techniques based on advanced sampling strategies can be introduced to reduce the computational cost associated with Monte Carlo simulations. Among these, the Stochastic Finite Element Method (SFEM) (Ghanem and Spanos 1991) is based on a spectral analysis that allows the expansion of the model output into the probabilistic space, called *Polynomial Chaos* (PC) (Wiener 1938). The Polynomial Chaos Expansion (PCE) of the model can be used to build a surrogate model such that the variability of the output is represented in the ensemble of the expansion coefficients (Sudret 2008). Once the

surrogate model has been derived, the calculation of the Sobol indices does not add significant extra computational costs. The formulation of a surrogate model in a polynomial form has the additional advantage of allowing performing Monte Carlo simulations with negligible computational effort, as compared to the original, high-complexity model.

In this work, we rely on PCE to analyze the uncertainty affecting the outputs of a numerical model of radionuclide migration in an aquifer, following a release from a near surface repository. The outflow from the repository is modeled within the Monte Carlo framework proposed by Cadini et al. (2012). Radionuclide migration in the aquifer is modeled through an Advection-Dispersion-Reaction-Equation (ADRE). The aquifer hydraulic conductivity constitutes a (second-order stationary) randomly heterogeneous field. In this context, the model outputs of interest are the first two (statistical) moments (i.e., mean and variance) of the peak concentration at a given control location in the aquifer. We study how the incomplete/imprecise knowledge of (a) the correlation scale, λ, of the variogram of the log-conductivity field, (b) the partition coefficient associated with the migrating radionuclide, k_d, and (c) the effective longitudinal dispersivity at the scale of interest, α_L, propagates to the selected (ensemble) moments of the output distribution.

GSA is performed jointly with PCE to compute the Sobol indices associated with the three uncertain parameters (λ, k_d, α_L), which are treated as random variables. The PCE – based surrogate model is then employed to perform an exhaustive set of Monte Carlo (MC) simulations to attain convergence for the target moments of interest. Given the prohibitive computational costs involved in performing a large number of MC simulations on the original flow and transport model, the goodness of PCE-based results is then assessed on the basis of a limited number of simulations, obtained upon sampling the selected random parameter space.

2. Theoretical Background and Methodology

2.1 Variance-based approaches for GSA

In this context, the ANOVA (ANalysis Of VAriance) representation of a model output (Archer et al. 1997) is a useful tool for the definition of the Sobol indices (Sobol 1993; Archer et al. 1997).

Consider a model function $y = f(\mathbf{x})$, y being a target random response of the model at a prescribed space-time location. This response depends on the vector $\mathbf{x}$ of n independent random model parameters defined in the n-dimensional unit hypercube, I^n. If $f(\mathbf{x})$ is integrable, the following representation holds:

$$f(\mathbf{x}) = f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots n}(x_1, x_2, \ldots x_n) \quad (1)$$

$$\int_0^1 f_{i_1,\ldots i_s}(x_{i_1},\ldots x_{i_s})\,dx_k = 0\,, \quad k = i_1,\ldots i_s \quad (2)$$

where $1 \le i_1 < \ldots < i_s \le n$ $(s = 1,\ldots,n)$ are the indices specifying the parameters upon which each term depends and the 2^n summands in (1) are orthogonal functions that can be expressed as integrals of $f(\mathbf{x})$, e.g. $f_0 = \int f(\mathbf{x})d\mathbf{x}$ is the mean of the model, $f_i(x_i) = \int f(\mathbf{x}) \prod_{k \ne i} d\mathbf{x}_k - f_0$ and so on. Therefore condition (2) renders representation (1), which is typically termed ANOVA decomposition, unique.

When $f(\mathbf{x})$ belongs to the space of square-integrable functions, then the total variance, V, of the model due to the uncertainty of its parameters is:

$$V = \int f^2(\mathbf{x})d\mathbf{x} - f_0^2 = \sum_{s=1}^{n} \sum_{i_1<\ldots<i_S}^{n} V_{i_1,\ldots i_S}\,, \qquad V_{i_1,\ldots i_s} = \int f_{i_1,\ldots i_s}^2\, dx_{i_1} \ldots dx_{i_s} \quad (3)$$

$V_{i_1,\ldots i_s}$ being the partial variance, expressing the contribution to V due to the interaction of the set of model parameters $\{x_{i_1},\ldots x_{i_s}\}$. The generic s-order Sobol index $S_{i_1,\ldots i_s}$ is defined as (Sobol 1993):

$$S_{i_1,\ldots i_s} = V_{i_1,\ldots i_s} / V \quad (4)$$

The sum of the indices defined in (4) is unity. The first-order or principal sensitivity indices, S_i, describe the significance of each parameter individually considered. Higher-order indices describe the effects of interactions among parameters. The overall effect of a given parameter x_i is described by the total sensitivity index S_{T_i}, defined as:

$$S_{T_i} = \sum_{\eta_i} S_{i_1,\ldots i_s}\,, \quad \eta_i = \{(i_1,\ldots i_s) : \exists k, 1 \le k \le s, i_k = i\}\,. \quad (5)$$

A complete GSA requires the estimation of 2^n integrals of the kind in (3). This is usually done by Monte Carlo simulation (Sobol 2001), but the computational cost becomes prohibitive when the model is complex and the number of uncertain parameters is large (Sudret 2008).

2.2 Polynomial Chaos Expansion representation of a stochastic model

We focus on the identification of a surrogate model (or metamodel) of a high complexity model (which is hereafter termed full system model) by the Polynomial Chaos Expansion (PCE) technique. This involves the projection of the model equation

into a probabilistic space, termed Polynomial Chaos, to construct an approximation of the model response surface. Wiener (1938) showed that the expansion performed by adopting Hermite Polynomials as a basis converges, in L_2-sense, for any random process characterized by finite second-order moments. While the Hermite basis is suitable for Gaussian processes, different types of orthogonal polynomials are required for optimum convergence rate in the case of non-Gaussian processes (Xiu and Karniadakis 2002).

In this framework, one starts by noting that any square-integrable random model response, S, admits the following expansion, or chaos representation (Soize and Ghanem 2004):

$$S = \sum_{j=0}^{\infty} s_j \Psi_j \left(\{\zeta_n\}_{n=1}^{\infty} \right) \tag{6}$$

Here, Ψ_j denotes the j-order multivariate orthogonal polynomial, $\{\zeta_n\}_{n=1}^{\infty}$ is the set of independent random variables whose distribution is linked to the choice of the polynomial basis (Xiu and Karniadakis, 2002), and s_j are the polynomial coefficients.

In various engineering fields one typically considers stochastic models associated with a finite number *M* of input random variables. The PCE of the random model output can be derived by approximating (6) to polynomials of degree not exceeding *p* as

$$S(x_1,...x_M) \cong \sum_{j=0}^{P-1} s_j \Psi_j (\zeta_1,..\zeta_M), \quad P = \frac{(M+p)!}{M!\,p!} \tag{7}$$

where P is the number of (unknown) polynomial coefficients.

The distribution of the input random variables of the model, included in vector $\mathbf{x}$, does not affect the applicability of the method. Note that in cases where this distribution is not interpreted by the one required by the chosen polynomial basis, an isoprobabilistic transformation is required to relate $\mathbf{x}$ and $\boldsymbol{\zeta} = (\zeta_1,...,\zeta_M)$. Correlation amongst random input model parameters can be accommodated in the methodology by applying the Nataf transformation (Nataf 1962), for which the knowledge of the marginal probability density functions of the parameters and the associated correlation matrix is required.

Assessment of the coefficients s_j in (7) can be performed by regression, upon minimization of the variance of a residual defined as the difference between the surrogate model response, $\tilde{S}$, and the exact solution given by the original model (Sudret 2008)

$$\varepsilon = S(\mathbf{x}) - \tilde{S}(\boldsymbol{\zeta}) = S(\mathbf{x}) - \sum_{j=0}^{P-1} s_j \Psi_j (\boldsymbol{\zeta}) \tag{8}$$

Minimization with respect to the vector of the unknown coefficients $\boldsymbol{\varsigma}$ renders

$$\varsigma = Min\left\{E\left[\left(S(\mathbf{x}) - \tilde{S}(\zeta)\right)^2\right]\right\} \tag{9}$$

with $E[\cdot]$ denoting expected value. It is useful to rewrite (9) as

$$\varsigma = \left(\mathbf{\Psi}^T\mathbf{\Psi}\right)^{-1}\mathbf{\Psi}^T\mathbf{S}', \qquad \mathbf{\Psi}_{ij} = \Psi_j\left(\zeta^i\right), \quad i = 1,...N;\ j = 0,...P-1 \tag{10}$$

where N is the number of regression points, $\mathbf{S}'$ is the vector denoting the model response at these points, while the product $\mathbf{\Psi}^T\mathbf{\Psi}$ defines the so-called information matrix.

The choice of the optimum set of regression points is performed following the same criterion adopted in the context of integral estimation by Gaussian quadrature (Huang et al. 2007). Solving (10) requires a minimum of $N = P$ regression points. One typically selects $N > P$ to avoid singularity in the information matrix.

2.3 Polynomial Chaos Expansion and Global Sensitivity Analysis

Polynomial Chaos Expansion can be considered as a powerful tool for Global Sensitivity Analysis because the entire variability of the original model is conserved in the set of PCE coefficients (Ghanem and Spanos 1991). The Sobol indices can be analytically calculated from these coefficients without additional computational cost (Sudret 2008). Manipulating $\tilde{S}$ by appropriate grouping of terms allows isolating the contributions of the different (random) parameters to the system response:

$$\begin{aligned}\tilde{S}(\zeta) = s_0 + \sum_{i=1}^{n}\sum_{\alpha\in\varphi_i} s_\alpha\Psi_\alpha(\zeta_i) + \sum_{1\le i_1<...<i_s\le n}\ \sum_{\alpha\in\varphi_{i_1...i_s}} s_\alpha\Psi_\alpha\left(\zeta_{i_1},...,\zeta_{i_s}\right) + ... \\ + \sum_{\alpha\in\varphi_{1,2,..n}} s_\alpha\Psi_\alpha(\zeta_1,...,\zeta_n)\end{aligned} \tag{11}$$

where φ denotes a general term depending only on the variables specified by the subscript.

In this sense, a PCE is similar to the ANOVA representation of the model. Orthogonality of the polynomial basis allows recognizing that the mean of the model response coincides with the coefficient of the zero-order term, s_0, in (11) while the total variance of the response is

$$V_{\tilde{S}} = Var\left[\sum_{j=0}^{P-1} s_j\Psi_j(\zeta)\right] = \sum_{j=1}^{P-1} s_j^2 E\left[\Psi_j^2(\zeta)\right] \tag{12}$$

The Sobol indices can then be derived as

$$S_{i_1,...i_s} = \frac{\sum_{\alpha\in\varphi_{i_1...i_s}} s_\alpha^2 E\left[\Psi_\alpha^2\right]}{V_{\tilde{S}}} \tag{13}$$

calculation of $E\left[\Psi_\alpha^2\right]$ can be performed, e.g., according to Abramowitz and Stegun (1970).

3. Application to a model of radionuclide migration in a randomly heterogeneous aquifer

We exemplify our approach by considering an environmental problem related to the performance assessment of a radioactive waste repository. We use a Monte Carlo simulation model to describe radionuclide release at the repository scale. This model of release of radionuclides, i.e., ^{239}Pu, from the repository is linked to a groundwater flow and transport numerical model to simulate radionuclide migration within a heterogeneous aquifer.

The aquifer hydraulic conductivity is modeled as a second-order stationary stochastic process in space. We take the first two (statistical) moments (i.e., mean and variance) of the peak concentration detected at a given control location in the aquifer, as the target model responses. Uncertainty in these variables is considered to be a consequence of incomplete knowledge of (a) the correlation scale of the variogram of the log-conductivity field (b) the partition coefficient associated with the migrating radionuclide, and (c) the effective dispersivity at the scale of interest.

3.1 Repository representation and modeling of radionuclide release history

The conceptual repository design considered in the performance assessment illustrated in this study has been proposed by ENEA (Marseguerra et al. 2001a, b) and has similarities with the currently operative disposal facility of El Cabril in Spain (Zuolaga 2006).

We model the repository as a one-dimensional (along the vertical direction) system (Cadini et al. 2012). The major containment structures of the disposal facility are the waste packages, the modules or containers, the cells and the disposal units. These constitute a multiple-barrier system designed to limit water infiltration and subsequent radionuclide migration. Figure 1a depicts a typical waste package consisting in a steel drum containing the radioactive waste and immobilized in a concrete matrix. The diameter and the height of the waste package have been set respectively to 0.791 m and 1.1 m, for a total volumetric capacity of around 400 l. Figure 1b shows a cross-section of the containment module adopted in this study, i.e., a concrete box-shaped structure which contains 6 waste packages and is sealed with a concrete top cover. The empty spaces between the packages are filled by bentonite. The external length of the module is 3.05 m, with a width and height of 2.09 m and 1.7 m, respectively. The corresponding internal dimensions are 2.75 m, 1.79 m and 1.37 m. The modules are arranged in $5 \times 6 \times 8$ arrays within concrete structure cells built below the natural ground level.

Figure 2 depicts the modules arrangement and the typical repository placement at a given site. The disposal unit is a concrete structure embedding a row of 6 to 10 cells. The disposal facility comprises several units, which are typically arranged into parallel rows. Each unit can be modeled as an independent system which can be built and operated without interfering with the remaining units.

In agreement with typical engineering scenarios we consider that (Marseguerra et al. 2001a, b): (i) the modules are identical; (ii) the mass transport occurs chiefly along the vertical direction; and (iii) lateral diffusive spreading is symmetric. Under these assumptions, estimating the probability of radionuclide release into the groundwater system below the repository can be reduced to the one-dimensional problem of estimating the release from a column of five identical vertically stacked modules, i.e., the repository column may be envisioned as a one-dimensional array of compartments, each corresponding to a module.

The radionuclides transition across the compartments is described stochastically. Under the assumption that solute displacement can be modeled as a Markovian process, the transition rates can be identified from the classical advection/dispersion equation. Non-Fickian transport can be modeled according to existing conceptual schemes (Berkowitz et al. 2006 and references therein) where the relevant transport parameters could be estimated by detailed data analysis at the temporal and spatial scales at which the processes of interest occur.

For the purpose of our example we adopt the following criteria, which can be considered as conservative in a performance assessment protocol: (i) the protection offered by the concrete cell roof and ceiling and the backfill layers fails; (ii) the whole column, which is formed by 5 modules, is saturated and a constant water head of 0.15 m is applied at the top of the highest module, i.e., the water head at the top of the column is $h(z = 5 \times 1.7 \text{ m}) = 8.65$ m; (iii) the water head at the bottom of the column is zero; (iv) each module is subject to constant head gradient $\Delta h/\Delta z = 1.018$, where $\Delta h = 8.65$ m and $\Delta z = 5 \times 1.7 \text{ m} = 8.5$ m is the column height; (v) the ^{239}Pu radioactive decay and the subsequent generation of other radionuclides from the decay chains are neglected within the repository; (vi) the migration of ^{239}Pu occurs at linear isothermal equilibrium.

The numerical code MASCOT (Marseguerra and Zio 2001; Marseguerra et al. 2003; Cadini et al. 2012) has been adopted to compute the probability density function of the release of ^{239}Pu from the modules. Details of the computations and the resulting temporal dynamics of the radionuclide release history are presented in Cadini et al. (2012).

3.2 Radionuclide migration in the groundwater system

For simplicity and for the purpose of our illustration we disregard the radionuclide transfer time within the partially saturated zone and analyze only contaminant residence time within the fully saturated medium. This assumption may be regarded as conservative because it leads to overestimating the radionuclide concentration detected downstream of the repository. This can also be considered as a viable working assumption in the presence of shallow reservoirs. The effect of processes occurring within the partially saturated region may require an additional analysis, which is outside the scope of this work.

Groundwater flow and contaminant transport are modeled within a two-dimensional system. The (natural) log-transformed hydraulic conductivity, $Y(\boldsymbol{x})$ ($\boldsymbol{x}$ denoting the space coordinates vector), is modeled as a second-order stationary spatial random function. For our example, the parameters of the variogram of Y have been selected as representative of a field case study, which we do not specifically report for confidentiality reasons. We note, however, that the particular choice of these values does not affect the generality of the methodology. Log-conductivity is characterized by an isotropic variogram of the exponential type, with sill $\sigma^2 = 1.21$. For the purpose of our illustrative example, we set the variogram sill and consider its correlation scale as an uncertain parameter (see Section 4) because of its poor identifiability due to typical horizontal spacing of available field-scale measuring locations. Monte Carlo realizations of $Y(\boldsymbol{x})$ have been performed by employing the sequential Gaussian scheme implemented in the code GCOSIM3D (Gómez-Hernández and Journel 1993).

We consider a two-dimensional domain of uniform lateral side equal to 2000 m. As an example, a selected realization of the log-conductivity distribution is depicted in Figure 3 together with the repository projection (R), with sides equal to 50 m and 80 m, and the target control point (W), located 960 m downstream of the repository fence line.

The domain is discretized into square cells with uniform side of 10 m, ensuring that there are at least five log-conductivity generation points per correlation scale (see Section 4 for additional details). Each of the 8 × 5 cells located under the repository projection area receives the release of a cluster of 4 × 3 columns of 5 modules. These cells are modeled through a recharge boundary condition so that a time-dependent influx solute mass is injected in the porous medium according to a suitable discretization in time of the Monte Carlo-based outflow from the repository. As in Cadini et al. (2012), we set the incoming water flow [m^3/y] from the repository at a constant value equal to $\Phi_{in} = q_d S, \quad q_d = 21.2$ [m/y] being the water Darcy flux at the bottom of the 5 modules column and S [m^2] being the area of the source cells. The associated radionuclide concentration [Bq/m^3] released to the aquifer is then:

$$C_{in}(t) = A_0 \frac{pdf_{out}(t)}{\Phi_{in}} \quad (14)$$

where A_0 = 1.6 × 10^6 [Bq] is the total activity of ^{239}Pu (which we assumed to be uniformly distributed) in the repository at a reference time t = 0 and $pdf_{out}(t)$ [y^{-1}] is the release probability density function from the four compartment domain (i.e., the five module column). The adopted ^{239}Pu activity level corresponds to the Italian inventory (Enea 2000) and justifies the assumption of disregarding solubility-limited release. In our example, the concentration of ^{239}Pu within the repository is $C_{rep}^{Pu239} \cong \frac{\lambda_r A_0}{N_A V_{rep}} = 2.96 \cdot 10^{-14} < C_{sl}^{Pu239} = 2.30 \cdot 10^{-4}$ [mol/m^3], where $\lambda_r = 0.28761 \cdot 10^{-4}$ [y^{-1}] is the ^{239}Pu constant decay, N_A is the Avogadro constant, V_{rep} is the total volume of the repository and C_{sl}^{Pu239} is the solubility limit of ^{239}Pu. Additional details are presented in Cadini et al. (2012).

Base groundwater flow in the aquifer is driven by a constant hydraulic head drop between the East and West boundaries, resulting in a unit average head gradient. No-flow conditions are assigned to the North and South boundaries.

Simulations of the steady state flow problem for each conductivity realization are performed with the widely used and thoroughly tested finite difference code MODFLOW2000 (McDonald and Harbaugh 1988). Radionuclide migration in the groundwater system is then modeled by means of the classical Advection-Dispersion Equation (ADE), where the partition coefficient, k_d, governing sorption of the contaminant onto the host solid matrix and the effective longitudinal dispersivity, α_L (for simplicity, transverse dispersivity is assumed to be equal to 0.1 α_L), are considered to be random variables, as described in Section 4. A uniform effective porosity of 0.15 is considered.

4. Global Sensitivity Analysis of the (ensemble) moments of radionuclide peak concentration

The three random parameters selected for our demonstration are assumed to be uniformly distributed within the intervals reported in Table 1. The ranges of variability of λ and α_L are compatible with the selected domain size, and consistent with the lack of a sufficiently large number of closely spaced Y measuring points. The degree of variability of k_d has been chosen according to ENEA (1997) and Nair and Krishnamoorthy (1999).

The model response, i.e., the radionuclide peak concentration, c_p, at the control point is then, in turn, a random variable. As introduced in Section 3, we perform our analysis in a numerical Monte Carlo framework according to the following steps: (a) a set

of $N_f = 100$ Y fields are generated by GCOSIM for given values of the random parameters sampled within the intervals presented in Table 1; (b) groundwater flow and transport are solved and (ensemble) mean, $\langle c_p \rangle$, and standard deviation, σ_{c_p}, of the peak concentration are computed; (c) steps (a) and (b) are repeated for different sampling points in the random parameters space; and (d) GSA is performed to discriminate the relative contribution of the random parameters to uncertainty of $\langle c_p \rangle$ and σ_{c_p}. Note that due to the random nature of $Y(\boldsymbol{x})$, we propose to perform GSA on the (ensemble) moments of c_p rather than on its actual value calculated at the selected control location for each random realization. Conceptually, this is equivalent to performing a GSA of the results stemming from the solution of transport equations satisfied by the ensemble moments of the evolving concentrations (e.g., Guadagnini and Neuman (2001) and Morales-Casique et al. (2006 a,b) for conservative solutes).

The procedure illustrated is rather cumbersome when considering the solution of the full system model, because of the large number of simulations required, so that a GSA might become impractical. Therefore, we adopt the PCE technique presented in Section 2 and derive expansions of order p = 2, 3 and 4, for both $\langle c_p \rangle$ and σ_{c_p}. We resort to the Legendre Chaos space, because the uncertain input parameters are associated with uniform distributions.

The calibration of the coefficients of the surrogate models requires $N_R = 10$, 38 and 78 (respectively for p = 2, 3, 4) sampling points in the space of the three selected uncertain parameters. In our example, this corresponds to $N_{MC} = 1000$, 3800, 7800 runs of the full model of groundwater flow and transport. Calculation of the Sobol indices is then performed with negligible additional computational requirements.

Figure 4 reports the Total Sensitivity Indices, S_T (left), and variances, V (right), of $\langle c_p \rangle$ versus the degree of polynomial expansion, p. Figure 5 reports the corresponding results for σ_{c_p}.

We start by noting that S_T and V are not dramatically influenced by the degree of polynomial expansion selected for both moments. The good agreement obtained between Total and Principal Sensitivity Indices (not shown) implies that the effects of parameters interactions can be neglected in this example.

Figure 4 reveals that k_d and α_L are the parameters which are most influential to $\langle c_p \rangle$, regardless of the degree of expansion adopted. On the other hand, the log-conductivity correlation scale, λ, and (to a lesser degree) the dispersivity, α_L, strongly

influence σ_{c_p}, while k_d does not have a significant impact for the specific values adopted in the case study. The uncertainty associated with the mean peak concentration is thus related mostly to the spatial structure of heterogeneity and to the strength of the dispersion phenomena, and less to the considered geochemical scenario.

The calibrated surrogate models allow extending with negligible computational cost the number of Monte Carlo simulation runs required for computing mean and standard deviation of $\langle c_p \rangle$ and σ_{c_p}, as illustrated in Section 2.2. Figures 6 and 7 respectively depict the dependence of the mean and the standard deviation of $\langle c_p \rangle$ and σ_{c_p} on the number of Monte Carlo runs performed with the calibrated surrogate models. The high number ($\approx 10^4$) of simulations required to attain convergence denotes the complexity of the case study and supports the adoption of a surrogate model to assess the uncertainty associated with the model response at reasonable computational costs.

The reliability of the results obtained through the PCE-based surrogate model has been analyzed by comparison against a number of full model runs performed by uniform sampling of N_s = 100 points in the random parameters space, corresponding to a total of 10^4 random realizations of $Y(\boldsymbol{x})$. The limited amount of sampling points selected is due to the excessive computational cost associated with the full model run (about 4 min for each simulation on a standard computer with a 3.16 GHz processor).

Figure 8 reports the relative fraction, $\mathscr{F}$(%), of the mean concentration values, $\langle c_p \rangle_l^{SM}$ (l = 1, 2, …, N_s), calculated with the PCE at different orders (p = 2, 3, 4) and comprised within intervals of width $w = \pm \left(\sigma_{c_p}^{FM}\right)_l$, $\pm 2 \left(\sigma_{c_p}^{FM}\right)_l$, and $\pm 3 \left(\sigma_{c_p}^{FM}\right)_l$ centered around $\langle c_p \rangle_l^{FM}$, $\langle c_p \rangle_l^{FM}$ and $\left(\sigma_{c_p}^{FM}\right)_l$ respectively being the mean and standard deviation of the peak concentration computed by means of the full system model. The latter is based on a standard Monte Carlo solution of radionuclide migration within NMC = 100 log-conductivity realizations for each $1 \le l \le N_s$. It can be seen that at least 40% of the values calculated with the surrogate models of different orders are comprised within the intervals of width $\pm \sigma_{c_p}^{FM}$, while about 75% of the results are included within intervals not exceeding $\pm 2\sigma_{c_p}^{FM}$. According to this criterion, Figure 8 suggests that the best results for our example appear to be provided by the PCE of order p = 2.

To complement these results, Table 2 reports the mean and standard deviation of $\langle c_p \rangle$ calculated on the basis of the N_s =100 sampling points in the random parameters space for each model (standard Monte Carlo and surrogate models of different order).

Table 3 reports the corresponding results for σ_{c_p}. The limited number of simulations does not allow to attain convergence of the target moments. However, it is possible to observe that the PCE of order p = 4 provides the best approximation of both the mean and standard deviation of $\langle c_p \rangle$ calculated with the full model. In other words, the Total Sensitivity Indices for $\langle c_p \rangle$ calculated with the PCE of order p = 4 are candidates to provide the best indications for a GSA, as one might expect. Finally, it can be noted that the PCE of order p = 3 best approximates the mean and standard deviation of σ_{c_p} calculated with the full model on the basis of the simulations performed.

5. Conclusions

In this work we proposed an approach for performing a Global Sensitivity Analysis (GSA) of a high-complexity theoretical and numerical model descriptive of the potential release of radionuclides from a near surface radioactive waste repository and their subsequent migration in the groundwater system. We considered uncertainty stemming from incomplete knowledge of the variogram and transport parameters (i.e., the correlation length of the variogram of log-conductivity, the partition coefficient associated with the migrating radionuclide and the effective dispersivity at the scale of interest) and, due to the random nature of the hydraulic conductivity field. We identified as target system responses the first two (ensemble) moments of the peak concentration at a given control point. GSA has been performed through the Polynomial Chaos Expansion (PCE) technique, leading to the following key results: (a) the analysis of the Sobol indices has revealed that the (ensemble) mean of the peak concentration is strongly influenced by the uncertainty in the partition coefficient and the longitudinal dispersivity, and the effects of these parameters shadow the impact of the spatial coherence of the log-conductivity field at the scale analyzed; (b) on the other hand, the log-conductivity correlation scale is the most influential factor affecting the uncertainty of the standard deviation of the peak concentration in our example; and (c) the PCE surrogate models allow extending, with negligible computational cost and acceptable accuracy, the number of Monte Carlo iterations to attain convergence of the selected target moments.

Our results support the relevance of adopting the proposed model reduction technique for complex numerical models. This methodology allows performing in-depth analyses which would be otherwise unfeasible, thus severely limiting our capability to represent the relevant processes involved in a target environmental scenario.

Acknowledgments

V. Ciriello acknowledges partial support from *Marco Polo Program* 2011 of the University of Bologna. F. Cadini, J. De Sanctis and E. Zio acknowledge the support from Agenzia Nazionale per le Nuove Tecnologie, l'Energia e lo Sviluppo Economico Sostenibile (ENEA).

References

Abramowitz M, Stegun IA (1970) Handbook of mathematical functions. Dover Publications, New York.

Archer GEB, Saltelli A, and Sobol IM (1997) Sensitivity measures, ANOVA like techniques and the use of bootstrap. J Stat Comput Simulation 58:99-120.

Ballio F, Guadagnini A (2004) Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. Water Resour Res 40 W04603.

Berkowitz B, Cortis A, Dentz M, Scher H (2006) Modeling non-Fickian transport in geological formations as a continuous time random walk. Rev of Geophys 44 RG2003.

Cadini F, De Sanctis J, Cherubini A, Zio E, Riva M, Guadagnini A (2012) An integrated simulation framework for the performance assessment of radioactive waste repositories. Annals of Nuclear Energy 39:1-8.

ENEA (1997) Internal Report. Chapman, N. A..

ENEA (2000) Inventario nazionale dei rifiuti radioattivi - Task Force per il sito nazionale di deposito dei materiali radioattivi. 3rd Ed (in Italian).

Ghanem RG, Spanos PD (1991) Stochastic finite elements-a spectral approach. Springer, Berlin.

Gómez-Hernández JJ, Journel AG (1993) Joint sequential simulation of multi-Gaussian field. Geostatitics Troia ’92, 1:85-94. Ed Soares.

Guadagnini A, Neuman SP (2001) Recursive conditional moment equations for advective transport in randomly heterogeneous velocity fields. Transp Porous Med 42:37-67.

Huang S, Sankaran M, Ramesh R (2007) Collocation-based stochastic finite element analysis for random field problems. Probabilistic Engineering Mechanics 22:194-205.

Marseguerra M, Zio E, Patelli E, Giacobbo F, Ventura G, Mingrone G (2003) Monte Carlo simulation of contaminant release from a radioactive waste deposit. Math Comput Simul 62:421-430.

Marseguerra M, Patelli E, Zio E (2001) Groundwater contaminant transport in presence of colloids I. A stochastic nonlinear model and parameter identification. Annals of Nuclear Energy 28:777-803.

Marseguerra M, Patelli E, Zio E (2001) Groundwater contaminant transport in presence of colloids II. Sensitivity and uncertainty analysis on literature case studies. Annals of Nuclear Energy 28:1799-1807.

Marseguerra M, Zio E (2001) Genetic algorithms for estimating effective parameters in a lumped reactor model for reactivity predictions. Nuclear Science and Engineering 139:96-104.

McDonald MG, Harbaugh AW (1988) A Modular Three-Dimensional Finite-Difference Groundwater Flow Model. Man. 83-875, U.S. Geol. Surv. Reston, VA.

Morales-Casique E, Neuman SP, Guadagnini A (2006) Nonlocal and localized analyses of nonreactive solute transport in bounded randomly heterogeneous porous media: Computational analysis. Adv Water Resour 29:1399-1418.

Morales-Casique E, Neuman SP, Guadagnini A (2006) Nonlocal and localized analyses of nonreactive solute transport in bounded randomly heterogeneous porous media: Theoretical framework. Adv Water Resour 29:1238-1255.

Nataf A (1962) Détermination des distributions dont les marges sont données. C R Acad Sci 225:42-3.

Nair RN, Krishnamoorthy TM (1999) Probabilistic safety assessment model for near surface radioactive waste disposal facilities. Environmental Modelling & Software 14:447–460.

Rubin Y (2003) Applied Stochastic Hydrogeology. Oxford Univ. Press, New York.

Saltelli A, Chan K, Scott EM (2000) Sensitivity analysis. Wiley, New York.

Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. Math Modeling Comput 1:407-414.

Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simulation 55:271-280.

Soize C, Ghanem R (2004) Physical systems with random uncertainties: Chaos representations with arbitrary probability measures. J Sci Comput 26(2):395-410.

Sudret B (2008) Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Syst Safety 93:964-979.

Tartakovsky DM (2007) Probabilistic risk analysis in subsurface hydrology. Geophys Res Lett 34.

Volkova E, Iooss B, Van Dorpe F (2008) Global sensitivity analysis for a numerical model of radionuclide migration from the RRC "Kurchatov Institute" radwaste disposal site. Stoch Environ Res Risk Assess 22:17-31.

Wiener N (1938) The homogeneous chaos. Am J Math 60:897-936.

Winter CL, (2010) Normalized Mahalanobis distance for comparing process-based stochastic models. Stoch Environ Res Risk Assess 24:917–923.

Winter CL, Tartakovsky DM (2002) Groundwater flow in heterogeneous composite aquifers. Water Resour Res 38(8):1148.

Xiu D, Karniadakis GE (2002) The Wiener-Askey polynomial chaos for stochastic differential equations. J Sci Comput 24(2):619-644.

Zhang D, Shi L, Chang H, Yang J (2010) A comparative study of numerical approaches to risk assessment of contaminant transport. Stoch Environ Res Risk Assess 24:971-984.

Zhang D (2002) Stochastic methods for flow in porous media: copying with uncertainties. Academic Press, San Diego.

Zuloaga P (2006) New Developments in LLW Management in Spain. ENRESA. <http://www.euronuclear.org/events/topseal/presentations/PP-Session-IIIZuloaga.pdf>.

Figure Captions

Fig. 1 Conceptual design of: (a) a waste package, (b) a containment module (ENEA 1987).

Fig. 2 Sketch of the 5 × 6 × 8 array of modules considered in a repository cell (ENEA 1987; Marseguerra et al. 2001a, b).

Fig. 3 Sketch of the adopted two-dimensional groundwater flow domain, including the repository projection (R) and the selected control point (W), for a selected realization of the log-conductivity field.

Fig. 4 Total Sensitivity Indices (S_T (Ω), $\Omega = \lambda$, α_L, k_d), Total Variance (V) and Partial Variances (V (Ω), $\Omega = \lambda$, α_L, k_d) calculated for $\langle c_p \rangle$ and p=2, 3, 4.

Fig. 5 Total Sensitivity Indices (S_T (Ω), $\Omega = \lambda$, α_L, k_d), Total Variance (V) and Partial Variances (V (Ω), $\Omega = \lambda$, α_L, k_d) calculated for σ_{c_p} and p=2, 3, 4.

Fig. 6 Dependence of the (a) mean and (b) standard deviation of $\langle c_p \rangle$ on the number of Monte Carlo iterations performed with the calibrated surrogate models.

Fig. 7 Dependence of the (a) mean and (b) standard deviation of σ_{c_p} on the number of Monte Carlo iterations performed with the calibrated surrogate models.

Fig. 8 Relative fraction, $\mathcal{F}$(%), of the mean concentration values, $\langle c_p \rangle_l^{SM}$ (l = 1, 2, …, N_s) calculated with the PCE at different orders (p = 2, 3, 4) which are comprised within intervals of width $w = \pm \left(\sigma_{c_p}^{FM}\right)_l$, $\pm 2 \left(\sigma_{c_p}^{FM}\right)_l$, and $\pm 3 \left(\sigma_{c_p}^{FM}\right)_l$ centered around $\langle c_p \rangle_l^{FM}$; $\langle c_p \rangle_l^{FM}$ and $\left(\sigma_{c_p}^{FM}\right)_l$ respectively are the mean and standard deviation of the peak concentration computed through the full system model on the basis of a standard Monte Carlo analysis of radionuclide migration within NMC = 100 log-conductivity realizations for each l.

581 **Table 1** Intervals of variability of the selected uniformly distributed random model parameters.
582

Random Variable	Distribution
Partition Coefficient, k_d	$U\left(1\frac{l}{g};3\frac{l}{g}\right)$
Longitudinal Dispersivity, α_L	$U(50m;70m)$
Correlation length of log-conductivity, λ	$U(40m;100m)$

583

584 **Table 2** Values of the mean and standard deviation of $\langle c_p \rangle$ calculated with the full model and the
585 surrogate models on the basis of 100 sampling points in the random parameter space.
586

Model	Mean of $\langle c_p \rangle$	Standard Deviation of $\langle c_p \rangle$
Full system model	2.738E-06	3.241E-07
Surrogate model p = 2	2.407E-06	7.175E-08
Surrogate model p = 3	3.190E-06	1.887E-07
Surrogate model p = 4	2.538E-06	3.462E-07

587

588 **Table 3** Values of the mean and standard deviation of σ_{c_p} calculated with the full system model
589 and the surrogate models on the basis of 100 sampling points in the random parameter space.
590

Model	Mean of σ_{c_p}	Standard Deviation of σ_{c_p}
Full system model	4.061E-07	8.169E-08
Surrogate model p = 2	4.708E-07	3.310E-08
Surrogate model p = 3	4.278E-07	5.719E-08
Surrogate model p = 4	4.530E-07	1.321E-07

591
592