

Article

Generalized Polynomial Chaos Expansion for Fast and Accurate Uncertainty Quantification in Geomechanical Modelling

Claudia Zoccarato , Laura Gazzola , Massimiliano Ferronato  and Pietro Teatini 

Department of Civil, Environmental and Architectural Engineering, University of Padova, 35131 Padova, Italy; laura.gazzola.1@phd.unipd.it (L.G.); massimiliano.ferronato@unipd.it (M.F.); pietro.teatini@unipd.it (P.T.)

* Correspondence: claudia.zoccarato@unipd.it

Received: 29 April 2020; Accepted: 24 June 2020; Published: 30 June 2020



Abstract: Geomechanical modelling of the processes associated to the exploitation of subsurface resources, such as land subsidence or triggered/induced seismicity, is a common practice of major interest. The prediction reliability depends on different sources of uncertainty, such as the parameterization of the constitutive model characterizing the deep rock behaviour. In this study, we focus on a Sobol’-based sensitivity analysis and uncertainty reduction via assimilation of land deformations. A synthetic test case application on a deep hydrocarbon reservoir is considered, where land settlements are predicted with the aid of a 3-D Finite Element (FE) model. Data assimilation is performed via the Ensemble Smoother (ES) technique and its variation in the form of Multiple Data Assimilation (ES-MDA). However, the ES convergence is guaranteed with a large number of Monte Carlo (MC) simulations, that may be computationally infeasible in large scale and complex systems. For this reason, a surrogate model based on the generalized Polynomial Chaos Expansion (gPCE) is proposed as an approximation of the forward problem. This approach allows to efficiently compute the Sobol’ indices for the sensitivity analysis and greatly reduce the computational cost of the original ES and MDA formulations, also enhancing the accuracy of the overall prediction process.

Keywords: geomechanical modelling; polynomial chaos expansion; bayesian update

1. Introduction

Geomechanical modelling is a scientific and engineering activity of paramount importance to evaluate the safety and predict possible environmental impacts related to the exploitation of subsurface resources. The reliability of the predictions depends on different sources of uncertainty, which are somewhat intrinsically introduced in any modelling process. Uncertainty typically affects the knowledge of the constitutive rock behaviour, the geometry of the depleted formations, the diffusion of the pressure perturbation, just to mention a few important occurrences. In this study, we focus on the reduction of uncertainty affecting the constitutive model parameters and the land subsidence prediction via assimilation of ground surface displacements.

A synthetic test case application dealing with the depletion of a deep hydrocarbon reservoir is considered. Land settlements are predicted with the aid of a 3-D Finite Element (FE) model using a one-way coupled approach [1–3]. The focus is on the calibration of the rock constitutive parameters that mainly control the compaction of the rock formation caused by the hydrocarbon production. Assuming the use of some well-established constitutive models, such as Mohr-Coulomb, Modified Cam-Clay, hypo-elastic, hypo-plastic or visco-elasto-plastic laws [4–7], different approaches can be used to estimate the governing parameters. If the required parameters have a measurable

physical meaning, a major investigation branch concerns either laboratory tests, properly developed for characterization of geological formations [8], or in-situ observations employed to study reservoir deformation with specific measurement apparatus, for example, well-logs equipped with radioactive markers [9]. Another important approach, which is especially related to those parameters that are not directly associated to measurable quantities, deals with inverse models to infer the uncertain model parameters using either deterministic or stochastic techniques. The latter approaches are increasingly employed in geomechanics. For example, in References [10,11] the Ensemble Smoother (ES), that is, a Monte Carlo (MC)-based method, is used to estimate the parameters of a transversely isotropic and isotropic constitutive law by the joint assimilation of horizontal and vertical displacements. Later developments of the ES algorithm in the form of Multiple Data Assimilation (ES-MDA) are used in Reference [12] to constrain the reservoir compaction coefficient and the subsurface basement elastic modulus by assimilation of ascending and descending line-of-sight displacements revealed by Permanent Scatterer Interferometry (PSI). Recently, Reference [13] compares the results obtained from applying the ES and the ES-MDA to estimate the parameters of the modified Cam-Clay and the Vermeer-Neher (VN) visco-elasto-plastic models [3].

The growing interest on the ES as a data assimilation technique is mainly due to its straightforward implementation and the capability of using simultaneously the measurements collected at different times and locations in a 4-D assimilation. However, its convergence is guaranteed only with a large number of MC simulations that may be computationally unfeasible in demanding large-scale problems. Moreover, its effectiveness in constraining the model for a strongly non-linear dependence of the outcome on the uncertain parameters is theoretically questionable. This limitation can be overcome by employing the ES in a multiple assimilation fashion, with the introduction of the ES-MDA technique [14], but this implies a further significant increase of the computational cost for the overall procedure, with the ES algorithm repeatedly applied to different ensembles. It follows the need for fast model predictions that can be achieved by approximating the forward model operators with surrogate solutions, obtained at a highly reduced computational cost.

In geomechanical applications, surrogate models have been recently employed for uncertainty quantification (UQ) and sensitivity analysis (SA). A functional data analysis technique is used to perform a vertical displacements UQ on a synthetic field-scale test by setting the oedometric compressibility as uncertain parameter [15]. Polynomial chaos proxies of the linear poroelastic problem with hydromechanical coupling allows a variance-based SA on Lamé's constants, the Biot-Willis coefficient and the hydraulic mobility [16]. Another non-intrusive approach is used in sequentially coupled reservoir-geomechanical simulations to estimate the caprock safety factor in a probabilistic setting [17]. Global SA and parameter identification are carried out by means of Gaussian processes in a carbon capture and storage test case to model fault poromechanics and induced seismicity in a stochastic framework [18]. A similar approach is used in Reference [19] to investigate the onset of fault activation in deep hydrocarbon reservoirs with surrogates based on polynomial chaos expansion. Global SA are also employed to investigate geomechanical fractured reservoirs [20] and hydraulically fractured wells with reduced order models [21].

In this work, we focus our attention on the generalized polynomial chaos expansion (gPCE) [22–26]. Using a spectral representation in the random space, gPCE-stochastic solutions are expressed by orthogonal polynomials of the input parameters [27]. This approach has already proved successful in several different engineering applications [28,29] with a growing interest in the Earth science community for UQ and SA [30–35]. Moreover, gPCE is suitable for identification problems in the Bayesian inference framework as documented in previous studies [36–40], to cite just a few.

Here we present an approach where we further extend the use of the gPCE as surrogate for regional geomechanical models with the aim of land subsidence prediction due to the exploitation of subsurface resources. Our contribution is threefold: (i) to validate the gPCE surrogate model for a highly non-linear constitutive law such as the VN visco-elasto-plastic model, (ii) to perform a global SA and rank the influence of the model parameters and their combinations, and (iii) based on

sensitivity results, to adapt accordingly the probabilistic framework for the ultimate goal of solving a parameter identification problem via ensemble-based Bayesian updating schemes. The gPCE surrogate is used in combination with the ES (gPCE-ES) and ES-MDA (gPCE-MDA) to solve the inverse problem based on measurements of vertical displacements, by sampling the ensemble members from the gPCE expansion. This is possible at a very low computational cost, because only direct polynomial evaluations are involved instead of running a full ensemble of forward models, provided that the gPCE approximation is sufficiently representative of the actual geomechanical outcome.

The paper is organized as follows. Section 2 includes the description of the governing equations of the forward simulation model, their approximation via gPCE expansion, the derivation of the Sobol' indices and the Bayesian-based updating schemes (ES and ES-MDA). Section 3 presents the numerical test case focusing on the variability range of the significant uncertain parameters. The results are provided in Section 4, where uncertainty propagation via gPCE approximation is first presented, with further assimilation of surface displacements to update the model outcome and provide the best estimate of the model parameters. The result of the gPCE-ES updating is compared with the solution from the classical ES formulation and with the gPCE-MDA approach, showing the computational efficiency and algorithmic effectiveness of the proposed surrogate model for uncertainty quantification and numerical prediction purposes. Finally, some conclusive remarks close the presentation.

2. Mathematical Framework

2.1. Geomechanical Governing Equations and Model Parameterization

The geomechanical model of a producing reservoir computes the stress and displacement fields generated by spatial and temporal pore pressure changes due to fluid withdrawal. The analysis is carried out by solving numerically the governing partial differential equations of 3-D poro-mechanics [41,42] with the aid of the FE method. A one-way coupled approach is used, as it is usually accepted in petroleum engineering. This implies that the pore pressure change in space and time is first computed by a dynamic flow model, and then used as an external source of strength in the equilibrium equations.

Let $\Omega \subset \mathbb{R}^3$ denote a 3-D domain in the x-y-z reference system with boundary Γ . Neumann and Dirichlet boundary conditions are imposed on boundaries Γ_N and Γ_D , respectively, with $\Gamma = \Gamma_N \cup \Gamma_D$ and $\Gamma_N \cap \Gamma_D = \emptyset$. The equilibrium equations governing the consolidation of a porous medium at some instant $t \in I =]0, T[$, with T the simulation period read:

$$\begin{cases} \nabla \cdot \underline{\sigma} - \alpha \nabla p = \mathbf{b} & \text{on } \Omega \\ \mathbf{u} = \hat{\mathbf{u}} & \text{on } \Gamma_D \\ \underline{\sigma} \cdot \mathbf{n} = \mathbf{f} & \text{on } \Gamma_N, \end{cases} \quad (1)$$

where $\underline{\sigma}(\mathbf{x}, t)$ is the effective stress tensor, $\alpha(\mathbf{x})$ the Biot coefficient, $p(\mathbf{x}, t)$ the pore pressure change, $\mathbf{b}(\mathbf{x}, t)$ the external body forces, and $\mathbf{n}(\mathbf{x})$ the outer normal to Γ_N . The system of Equations (1) is solved in terms of displacements $\mathbf{u}(\mathbf{x}, t)$, subject to the prescribed displacements $\hat{\mathbf{u}}(\mathbf{x}, t)$ along Γ_D and the force per unit surface $\mathbf{f}(\mathbf{x}, t)$ along Γ_N . To close the system, a non-linear constitutive relationship is introduced to link the effective stress to the strain tensor $\underline{\epsilon}(\mathbf{x}, t)$:

$$\underline{\sigma} = \mathbf{D}(\underline{\epsilon}), \quad (2)$$

where $\underline{\epsilon}$ is related to \mathbf{u} according to the small strain hypothesis:

$$\underline{\epsilon} = \frac{1}{2} (\nabla + \nabla^T) \mathbf{u}. \quad (3)$$

In this study, the VN [43] visco-elasto-plastic constitutive law is used. Equation (2) is written in differential form as:

$$\dot{\underline{\sigma}} = \mathbf{C} : \left(\dot{\underline{\epsilon}} - \dot{\gamma} \frac{\partial p_c}{\partial \underline{\sigma}} \right), \tag{4}$$

where \mathbf{C} is the standard rank-4 elasticity tensor, $\dot{\gamma}$ is the plastic multiplier rate and p_c is the plastic potential. The elasticity tensor \mathbf{C} depends on Poisson’s ratio ν and a stress-dependent Young’s modulus E :

$$E = (1 - 2\nu) \frac{3\rho_\kappa p}{\lambda^*}, \tag{5}$$

with p the volumetric stress (first stress invariant), λ^* the modified compression index, and ρ_κ the ratio between the modified compression and swelling index. The plastic multiplier rate is defined as:

$$\dot{\gamma} = \frac{\lambda^*}{\rho_\mu \tau^* \frac{\partial p_c}{\partial p}} \left(\frac{p_c}{R p_{c,0} \exp\left(\frac{-\rho_\kappa \epsilon_v^p}{\lambda^* (\rho_\kappa - 1)}\right)} \right)^{\rho_\mu (\rho_\kappa - 1) / \rho_\kappa}, \tag{6}$$

where ρ_μ is the ratio between the modified compression and creep index, τ^* is a reference time value, ϵ_v^p is the volumetric plastic strain, and p_c is set equal to a lumped scalar representation of the stress state in the plane of the first and second stress invariants (p, q) :

$$p_c = p + \frac{c}{\tan \varphi} + \frac{q^2}{M^2 \left(p + \frac{c}{\tan \varphi} \right)}. \tag{7}$$

In this last equation, c is the material cohesion, φ the friction angle, and M the slope of the Critical State Line in the (p, q) plane:

$$M = 3 \sqrt{\left(\frac{1 - K_0}{1 + 2K_0} \right)^2 + \frac{(1 - K_0)(1 - 2\nu)(\rho_\kappa - 1)}{(1 + 2K_0)(1 - 2\nu)\rho_\kappa - (1 - K_0)(1 + \nu)}}, \tag{8}$$

with K_0 the ratio between the vertical and horizontal stress in normal consolidation state, also known as confinement factor. Finally, in Equation (6) R is the geotechnical initial overconsolidation ratio defined as $p_{c,r,0} / p_{c,0}$, where $p_{c,0}$ is a representation of the stress state at initial conditions and $p_{c,r,0}$ is a parameter related to the plastic strain developed by the material before loading. For details on the numerical implementation of Equations (4)–(8) in the forward model, the reader is referred to Reference [3].

The VN visco-elasto-plastic model requires a significant number of independent material parameters, which may vary according to the selected implementation. In particular, for the formulation presented herein we need: ν , λ^* , ρ_κ , ρ_μ , τ^* , R , c , φ , and K_0 . The present analysis focuses on the parameters that mostly characterize the VN constitutive law with respect to more traditional ones, specifically, (1) λ^* , (2) ρ_κ , (3) ρ_μ , and (4) R . Parameter λ^* represents the slope of the normal consolidation profile of volumetric strain vs axial stress in a logarithmic plot, while ρ_κ allows to obtain the slope of the same profile in unloading/reloading conditions. The parameter ρ_μ depends on the slope of the volumetric strain vs. time profile, in natural logarithmic scale. Finally, the ratio R provides the over-consolidation degree of the porous rock and describes the size of the implicit yield surface defined by the VN model in the plane of the first two stress invariants (p, q) . This set of parameters is also selected because of the availability of literature estimates and the independence on the employed unity of measure. All the remaining parameters have been fixed to standard values generally accepted for a wide range of materials ($\nu = 0.3$, $\tau^* = 1$ day, $c = 1$ MPa, $\varphi = 30^\circ$, and $K_0 = \nu / (1 - \nu) = 0.43$).

2.2. gPCE Surrogate Model

Running the forward geomechanical model multiple times for large and complex systems can be a very demanding task, both in terms of CPU and memory requirements. A gPCE approach [22,44] is therefore proposed to approximate the outcome of the deterministic simulator and then compute the propagation of the input uncertainty through the forward model. The main idea of gPCE surrogate models is based on using orthogonal polynomial approximations of the random input to project the stochastic model output. In the following, we provide the basic mathematical framework as derived in Reference [44].

Let us consider the random vector $\mathbf{U} = (U_1, \dots, U_\ell) \in \mathbb{R}^\ell$ written as a function of the random vector \mathbf{Z} of mutually independent random variables $\mathbf{Z} = (Z_1, \dots, Z_n)$ and distribution function $F_Z(z_1, \dots, z_n) = \mathcal{P}(Z_1 \leq z_1, \dots, Z_n \leq z_n)$, where we are considering a stochastic process in the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with space of events Ω , σ -algebra \mathcal{F} and probability measure \mathcal{P} on \mathcal{F} , see for example Reference [27]. As usual, the independence assumption implies $F_Z(\mathbf{z}) = \prod_{i=1}^n F_{Z_i}(z_i)$, where $F_{Z_i}(z_i) = \mathcal{P}(Z_i \leq z_i)$ is the marginal distribution function with $i = 1, \dots, n$. Since any random variable may be represented as a series of polynomials in uncorrelated and independent Gaussian variables [22] and, in its generalized extension, in non-Gaussian measures, gPCE basis functions of a univariate random variable Z_i are defined as the polynomials $\{\phi_k(Z_i)\}_{k=0}^N$ of N th-degree satisfying the orthogonality conditions

$$\mathbb{E}[\phi_s(Z_i)\phi_r(Z_i)] = \int_{\Sigma_i} \phi_s(z_i)\phi_r(z_i)dF_{Z_i}(z_i) = \gamma_s\delta_{s,r} \quad 0 \leq s, r \leq N, \tag{9}$$

with $\gamma_s = \mathbb{E}[\phi_s^2(Z_i)]$ the normalization factors, $\delta_{s,r}$ the Kronecker delta function and Σ_i is the support of Z_i . In the multivariate case, the gPCE basis functions $\Phi_\alpha(\mathbf{Z})$ of degree up to N are products of the univariate orthogonal polynomials:

$$\Phi_\alpha(\mathbf{Z}) = \phi_{\alpha_1}(Z_1) \dots \phi_{\alpha_n}(Z_n) \quad \text{with} \quad 0 \leq |\alpha| \leq N, \tag{10}$$

where $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$ is a multi-index with $|\alpha| = \alpha_1 + \dots + \alpha_n$. The multivariate basis functions are orthogonal polynomials in $L^2_{dF_Z}$, that is, the space of all mean-square integrable functions of \mathbf{Z} with respect to the inner product based on the measure dF_Z :

$$\mathbb{E}[\Phi_\alpha(\mathbf{Z})\Phi_\zeta(\mathbf{Z})] = \int_{\Sigma} \Phi_\alpha(\mathbf{z})\Phi_\zeta(\mathbf{z})dF_Z(\mathbf{z}) = \gamma_\alpha\delta_{\alpha\zeta}, \tag{11}$$

where Σ is defined by $\Sigma = \Sigma_1 \times \Sigma_2 \dots \times \Sigma_n$. As a consequence, the class of orthogonal polynomials is selected according to the measure F_{Z_i} . The polynomials orthogonal for the standard normal distribution are the Hermite polynomials, which form an ideal basis for the output stochastic domain [27].

In the gPCE context, we aim at finding an approximation $\tilde{\mathbf{U}}_N(\mathbf{Z})$ of the random function $\mathbf{U}(\mathbf{Z}) \in \mathbb{R}^\ell$ in the N -th degree polynomial space generated by the basis functions $\Phi_\alpha(\mathbf{Z})$:

$$\mathbf{U}(\mathbf{Z}) \approx \tilde{\mathbf{U}}_N(\mathbf{Z}) = \sum_{|\alpha| \leq N} \mathbf{c}_\alpha \Phi_\alpha(\mathbf{Z}), \tag{12}$$

where $\mathbf{c}_\alpha \in \mathbb{R}^\ell$ are the coefficients of the expansion. For $\mathbf{U}(\mathbf{Z}) \in L^2_{dF_Z}$, the coefficients \mathbf{c}_α can be computed by defining $\tilde{\mathbf{U}}_N$ as the orthogonal projection of \mathbf{U} onto the polynomial space $\mathcal{Z} = \text{span}\{\Phi_\alpha\}$. By prescribing the orthogonality condition $\mathbf{U} - \tilde{\mathbf{U}}_N \perp \text{span}\{\Phi_\alpha\}$:

$$\int_{\Sigma} [\mathbf{U}(\mathbf{Z}) - \tilde{\mathbf{U}}_N(\mathbf{Z})] \Phi_\alpha dF_Z = \mathbf{0} \tag{13}$$

the coefficients c_α simply read

$$c_\alpha = \frac{1}{\gamma_\alpha} \mathbb{E}[\mathbf{U}(\mathbf{Z})\Phi_\alpha(\mathbf{Z})] = \frac{1}{\gamma_\alpha} \int_{\Sigma} \mathbf{U}(\mathbf{z})\Phi_\alpha(\mathbf{z})dF_{\mathbf{Z}}(\mathbf{z}) \quad |\alpha| \leq N, \tag{14}$$

that is, they can be numerically computed as an integral of the product of Φ_α and \mathbf{U} . The expansion terms of Equation (14) guarantees the optimal approximation of \mathbf{U} in the sense of the norm defined in $L^2_{dF_{\mathbf{Z}}}$.

The coefficients c_α of the approximating gPCE are numerically computed by a non-intrusive approach, where the forward model providing $\mathbf{U}(\mathbf{Z})$ is used in a black-box fashion. We use a pseudo-spectral projection, with the integral term approximated by a high-dimensional quadrature rule:

$$c_\alpha \approx \tilde{c}_\alpha = \sum_{j=1}^{n_q} \mathbf{U}(\mathbf{z}^j)\Phi_\alpha(\mathbf{z}^j)w(\mathbf{z}^j) \tag{15}$$

with \mathbf{z}^j and $w(\mathbf{z}^j)$ the n_q integration nodes and weights, respectively. Since Φ_α is at most of degree N , the integrand function has at most degree $2N$. In the univariate case, this requires the use of a $(n_{q,1} = N + 1)$ -point Gaussian quadrature rule, while in the multivariate case with n random variables the number of points grows up to $n_q = (N + 1)^n$. Using this approximation, the surrogate model needs the evaluation of \mathbf{U} through the the numerical solver \mathcal{S} of the forward model at the n_q integration points \mathbf{z}^j .

In our application, we denote with \mathbf{U} the vector of state variables, that is, displacements and stresses on the points and times of interest, provided by the solution of problem (1) via a FE numerical solver \mathcal{S} . The random variables \mathbf{Z} are the material parameters defined by the visco-elasto-plastic VN constitutive law, that is, λ^* , ρ_κ , ρ_μ , and R , with n at most equal to 4.

2.3. Sobol' Indices in the gPCE Framework

The objective of a preliminary SA is to quantify the influence of the uncertain input parameters and their combinations on the model output, so as to retain only the most significant contributions. This is important to (i) focus the design of laboratory/field experiments on the collection of measurements of the most important parameters, and (ii) reduce the number of calibration parameters for an easier solution of the inverse problem, especially in real-world applications. In this work, we propose the use of the variance-based SA according to the framework developed in Reference [45]. A variance decomposition of the model output into partial variances of the model input (and combination of more than one parameter), allows to estimate the so-called Sobol' indices, which provide a measure of the relative importance of each uncertain parameter to the model outcome. In the sequel, we provide a brief description of the ideas lying behind Sobol' indices. For completeness, the reader is referred to the work of Reference [45] or to the application developed in Reference [19].

For the sake of simplicity, let us consider the model outcome U , that is, the quantity of interest, as a one-dimensional variable function f of the n -variate random vector $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ of mutually independent components defined over an n -dimensional hypercube Σ^n . Assuming f to be a square-integrable function, the functional decomposition of $U = f(\mathbf{Z})$ reads

$$f = f_0 + \sum_{i=1}^n f_i + \sum_{i=1}^n \sum_{j>i}^n f_{i,j} + \dots + f_{1,2,\dots,n}, \tag{16}$$

where f_0 is a constant representing the mean value of f , and $f_i, f_{i,j}, \dots, f_{1,2,\dots,n}$ are the uncorrelated random effects associated to the factors in their indices, for example, $f_i = f_i(Z_i)$ are the main effects due to the factors Z_i and $f_{i,j} = f_{i,j}(Z_i, Z_j)$ are the effects related to interactions between the factors Z_i and Z_j with $j > i$. It has been proved in Reference [45] that a unique expansion of Equation (16) exists for any function $f(\mathbf{Z})$ integrable in Σ^n under the hypothesis of zero mean of all expansion

terms with respect to each variable. The expansion (16) for the quantity of interest U is used to derive the associated variance $V(U)$. In fact, squaring Equation (16) and using the orthogonality condition holding for the expansion terms [45], it can be demonstrated that $V(U)$ reads:

$$V(U) = \sum_{i=1}^n V_i + \sum_{i=1}^n \sum_{j>i}^n V_{i,j} + \dots + V_{1,2,\dots,n}, \tag{17}$$

where $V_i, V_j, \dots, V_{1,2,\dots,n}$ are the partial variances of $f_i, f_j, \dots, f_{1,2,\dots,n}$, respectively. Generally, the partial variance V_{i_1,\dots,i_s} in (17) reads:

$$V_{i_1,\dots,i_s} = \int_{\Sigma^s} f_{i_1,\dots,i_s}^2(Z_{i_1}, \dots, Z_{i_s}) d(Z_{i_1}, \dots, Z_{i_s}) \tag{18}$$

which can be written in terms of conditional expectations as:

$$\begin{aligned} V_i &= V(\mathbb{E}(U|Z_i)) \\ V_{i,j} &= V(\mathbb{E}(U|Z_i, Z_j)) - V_i - V_j \\ V_{i,j,k} &= V(\mathbb{E}(U|Z_i, Z_j, Z_k)) - V_i - V_j - V_k - V_{i,j} - V_{i,k} - V_{j,k} \\ &\vdots \\ V_{1,\dots,n} &= V(U) - \sum_{i=1}^n V_i - \sum_{i=1}^n \sum_{j>i}^n V_{i,j} - \dots - \sum_{1 \leq i_1 < \dots < i_{n-1} \leq n} V_{i_1,\dots,i_{n-1}}. \end{aligned} \tag{19}$$

Equations (19) provide practical relationships for the computation of partial variances for any indices combination. The Sobol' first and higher-order indices are defined as:

$$S_i = \frac{V_i(U)}{V(U)} \tag{20}$$

$$S_{1,\dots,s} = \frac{V_{1,\dots,s}(U)}{V(U)}, \tag{21}$$

where S_i measures the relative importance of a single factor Z_i on the total model variance and the higher-order indices $S_{1,\dots,s}$ represent a measure of the combined model sensitivity to the group of factors Z_1, \dots, Z_s . The total effects $S_{T,i}$ are also computed to quantify the contribution of the i -th factor to the total output variation as [46]:

$$S_{T,i} = 1 - \frac{V(\mathbb{E}(U|Z_{\sim i}))}{V(U)}, \tag{22}$$

where $V(\mathbb{E}(U|Z_{\sim i}))/V(U)$ include first and higher-order interactions of all factors except Z_i .

Sobol' indices can be easily obtained by MC sampling. However, the computational cost of the evaluations can be prohibitive for large scale problems and the use of gPCE approximation is straightforward for the purpose, being the Sobol' indices readily evaluated from the coefficients of the gPCE [33,47,48]. The idea is to replace the functional decomposition of Equation (16) with the N -th-degree polynomial expansion of Equation (12) by defining [48]:

$$f_{i_1,\dots,i_s}(Z_{i_1}, \dots, Z_{i_s}) \simeq \sum_{|\beta| \leq N} c_\beta \Phi_\beta(Z_{i_1}, \dots, Z_{i_s}) \tag{23}$$

where the multi-index $\beta = (\beta_1, \dots, \beta_s) \in \mathbb{N}_0^s$ satisfies $|\beta| = \beta_1 + \dots + \beta_s$ and $\Phi_\beta(Z_{i_1}, \dots, Z_{i_s})$ are the s -variate N -th-degree gPC basis functions.

2.4. gPCE-ES and gPCE-MDA

In this section, we provide the mathematical framework for the Bayesian inverse modelling approach based on gPCE. We aim at estimating the state variable and model parameter distributions where few information is available on such inputs. This is pursued by establishing a Bayesian inverse modelling approach with the model outcome constrained by spatio-temporal observations of measurable quantities.

Let us assume that $\mathbf{d} \in \mathbb{R}^m$ is the vector of noisy empirical measurements of the true observable vector $\mathbf{d}^t \in \mathbb{R}^m$:

$$\mathbf{d} = \mathbf{d}^t + \boldsymbol{\epsilon}, \tag{24}$$

where $\mathbf{d}^t = \mathcal{H}(\mathbf{u}^t)$ with \mathcal{H} a measurement operator mapping the true state variables $\mathbf{u}^t \in \mathbb{R}^\ell$ into the true observable vector. The vector $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is the additive observational error with probability distribution function $\rho(\boldsymbol{\epsilon}) = \prod_{i=1}^m \rho(e_i)$, that is, assuming e_i with $i = 1, \dots, m$ mutually independent and identically distributed variables. Of course, \mathbf{u}^t is not available because of the uncertainties of the mathematical modelling process. Through inverse modelling we aim at providing the best estimate of the “true” state/parameters variables conditioned by (i) available mathematical models, (ii) prior parameters knowledge, and (iii) measurement data. Here, we employ the ES technique and its iterative version in form of ES-MDA [14,49,50]. The linear unbiased smoother estimate can be written in a matrix form as:

$$\mathbf{X}^u = \mathbf{X}^f + K(D - H\mathbf{X}^f), \tag{25}$$

where $\mathbf{X}^f = (U^f, P^p)^T \in \mathbb{R}^{(\ell+n) \times n_{MC}}$ is the augmented matrix of the forecast ensemble states \mathbf{u}^f and prior parameters \mathbf{p}^p , with U^f obtained as the collection of the outcome of n_{MC} MC simulations and P^p the corresponding prior parameters for every simulations assuming no model errors. The augmented matrix $\mathbf{X}^u = (U^u, P^u)^T \in \mathbb{R}^{(\ell+n) \times n_{MC}}$ is the updated solution by correction of \mathbf{X}^f with assimilated measurements. Matrix $D \in \mathbb{R}^{m \times n_{MC}}$ contains the perturbed measurements matrix with $\boldsymbol{\epsilon} \sim N(0, R)$ and $R \in \mathbb{R}^{m \times m}$ the error covariance matrix, while H is a linearization of the measurement operator \mathcal{H} . The innovation $(D - H\mathbf{X}^f)$ is weighted over the so-called Kalman gain K calculated as:

$$K = C^f H^T (H C^f H^T + R)^{-1}, \tag{26}$$

where C^f is the covariance matrix of the forecast augmented state \mathbf{X}^f . In the form of gPCE-ES, the MC simulation of the forward operator are approximated by straightforward polynomial evaluations to cast the matrix \mathbf{X}^f by using the surrogate solution provided by Equation (12). Thus, a large number of MC samples can be adopted for the forecast ensemble and reduce sampling errors [27].

It is sometimes difficult to achieve a proper constraining of the model through the available data with the single-update of ES, especially when the forward model outcome have a strongly non-linear dependence on the uncertain parameters [14,51]. For this reason, Reference [50] introduced the MDA technique, that was later combined with ES [14], in order to improve the results of a single assimilation for history-matching problems. The ES-MDA is widely recognized as a tool to refine the conditioning of the model especially when non-linearities are significant, even if there is no rigorous proof about its convergence [52]. The idea behind ES-MDA is to repeat the assimilation of the same measurement data for n_{MDA} times. In order to avoid an over confidence in the available data, the covariance matrix R is inflated by a coefficient $\alpha_{MDA} \geq 1$. The choice of the coefficient α_{MDA} for every assimilation step is arbitrary as long as the sum of their inverses is unitary:

$$\sum_{i=1}^{n_{MDA}} \frac{1}{\alpha_{MDA}(i)} = 1. \tag{27}$$

The condition (27) ensures the equivalence between ES and ES-MDA for Gaussian-linear problems [14].

3. Numerical Results

The sequence of steps resulting from the mathematical framework defined in the previous section is summarized in Algorithm 1. After the initialization, the gPCE surrogate model is built taking into account only the most significant parameters, whose selection is based on the computation of the total Sobol' index S_T . We retain in the final model only the parameters with S_T larger than a user-specified threshold ε_S , which can be conveniently adjusted so as to provide a good balance between model representativeness and cost. Whenever a parameter is excluded from the uncertainty analysis, the gPCE model has to be recomputed. The most expensive task, that is, the n_q full model S runs for a variable number of sweeps, is carried out at this stage, which must be therefore carefully designed. Once the surrogate model is available, the uncertainty quantification step can be completed by means of the classical ES or ES-MDA algorithms, where a very large number of MC realizations can be now computed at a low computational cost.

Algorithm 1 gPCE-ES and gPCE-MDA for the forward model S .

```

1: % Initialization
2: Define  $n$  uncertain parameters in the vector  $p$ 
3: Set  $N, \varepsilon_S, n_{MC}, n_{MDA}$  and  $\alpha_{MDA}$ 
4: Input  $d$  and  $H$ 
5: % gPCE model and sensitivity analysis
6: Check = True
7: while Check do
8:   Check = False
9:   Set the gPCE basis functions  $\Phi_\alpha$ 
10:   $n_q = (N + 1)^n$ 
11:   $c_\alpha = \mathbf{0}$ 
12:  for  $j = 1, \dots, n_q$  do
13:    Set the nodes  $z^j \in \Sigma$  and the weights  $w(z^j)$ 
14:     $U^j = S(z^j)$ 
15:     $c_\alpha \leftarrow c_\alpha + U^j \Phi_\alpha(z^j) w(z^j)$ 
16:  end for
17:  Set  $s$  as the total number of Sobol' indices
18:  for  $i = 1, \dots, s$  do
19:    Compute  $S_i = V_i / V$ 
20:  end for
21:  for  $k = 1, \dots, n$  do
22:    Compute  $S_{T,k}$ 
23:    if  $S_{T,k} < \varepsilon_S$  then
24:      Remove the  $k$ -th parameter from  $p$ 
25:       $n \leftarrow n - 1$ 
26:      Check = True
27:    end if
28:  end for
29: end while
30: % Constrain the model via Data Assimilation
31: Set the prior parameter set  $p^p$ 
32: for  $i = 1, \dots, n_{MDA}$  do
33:   Compute  $D$  and  $R$  using  $\alpha_{MDA}$ 
34:   for  $j = 1, \dots, n_{MC}$  do
35:      $u^f = \sum_{|\alpha|} c_\alpha \Phi_\alpha(p^p)$ 
36:   end for
37:   Compute  $C^f$ 
38:    $K = C^f H^T (H C^f H^T + R)^{-1}$ 
39:    $X^u = X^f + K(D - H X^f)$ 
40: end for
41:  $u^u = \sum_{|\alpha|} c_\alpha \Phi_\alpha(p^p)$ 

```

3.1. Model Set-Up and Random Parameters

The proposed approach is tested in a realistic geomechanical analysis concerning the development of a production program from a synthetic hydrocarbon reservoir. For the sake of realism, the model is built by introducing some of the typical features of the off-shore reservoirs buried in the Northern Adriatic basin, Italy [9,10,53].

The model domain covers an area of 50 km × 50 km and extends down to 5 km depth. It is discretized into the 3-D FE mesh shown in Figure 1a consisting of 71,734 nodes and 410,030 tetrahedrons. The rock in the reservoir and the hydraulically connected aquifer has a strongly non-linear visco-elasto-plastic behaviour described by the VN model, while the overburden and the underburden, that is the portions of the domain that are not filled by color in Figure 1a, are linear elastic. Homogeneous essential conditions apply on the outer and the bottom boundaries, with the top of the domain modeled as a traction-free surface. We assume the hydrocarbon production program from the reservoir to last for 3 years, then start again after a 2-year stop and be completed after 7 years, with the monitoring continuing for 3 additional years after the production stop. The average pressure variation occurring within the reservoir in time is shown in Figure 1b. The pressure distribution in space, also within the connected aquifer, is predicted by means of a dynamic reservoir flow model. The reader can refer to Reference [13] for more details.

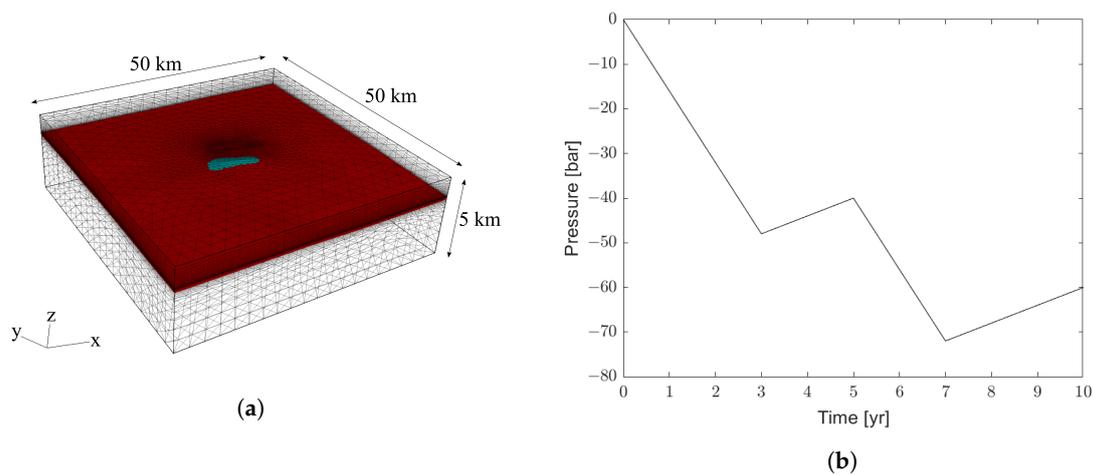


Figure 1. (a) View of the three-dimensional finite element grid of the regional geomechanical model. The hydrocarbon reservoir is blue, while the connected aquifer is red. The vertical axis is exaggerated by a factor 3. (b) Average pore pressure variation in time within the reservoir.

Model uncertainty is connected to the four parameters of the material constitutive law mentioned in Section 2.1. Based on physical considerations and a possible prior knowledge, if available, a range of variability is associated to such parameters. In particular, we set:

$$\begin{cases} 1. \ln(\lambda^*) \sim \mathcal{U}(-5.60; -4.28) \\ 2. \rho_\kappa \sim \mathcal{U}(5; 20) \\ 3. \rho_\mu \sim \mathcal{U}(40; 80) \\ 4. R \sim \mathcal{U}(1.2; 1.5) \end{cases} \quad (28)$$

The range for λ^* has been chosen recalling the physical relationship between λ^* and the vertical compressibility c_m , which has been derived in References [9,53] for the Northern Adriatic basin. For the other parameters typical values reported in the literature, for example in Reference [3], have been taken into account. Hence, we initialize the gPCE algorithm by defining a vector of $n = 4$ random parameters $\mathbf{p} = (\lambda^*, \rho_\kappa, \rho_\mu, R)$. Note that since we employ uniform distributions for the random variables, the Legendre polynomials are used as gPCE basis functions. All the other

modelling assumptions, including initial and boundary conditions, geometry and pressure distribution, are considered as deterministic.

As quantities of interest, collected in the vector \mathbf{U} given as output from the full forward geomechanical model \mathcal{S} , we consider the time series of the vertical displacement at a single-point of the top surface showed in Figure 2a, corresponding to land subsidence in time. Measurements of this quantity are supposed to be collected every three months for ten years, for example, by a GPS station. The true observations are collected in the vector \mathbf{d}^t . They are synthetically obtained by running the forward model with what is assumed to be the true material parameter set (continuous lines in Figure 2b). Then, the observation vector \mathbf{d} is obtained by adding an error ϵ sampled from a zero-mean Gaussian noise with standard deviation $\sigma_\epsilon = 0.01$ m (dots and squares in Figure 2b).

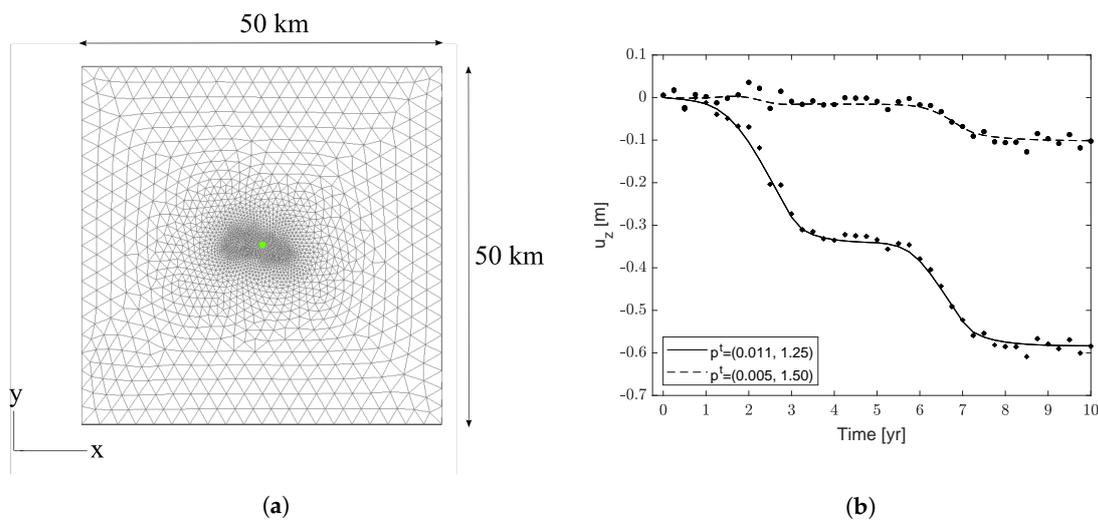


Figure 2. (a) 2D view of the model top with the green point indicating the location where the measurements are supposed to be recorded. (b) Synthetic sequence of vertical displacements u_z used in the assimilation process for two different parameter sets p^t . The continuous and dashed lines are the outcomes of the geomechanical model, while dots and squares indicate the assimilated observations.

3.2. gPCE Surrogate Model

Since we use uniform distributions for the input parameters, the gPCE basis functions are multivariate Legendre polynomials, which are orthogonal with respect to the uniform measure. The first objective of this analysis is to test the quality of gPCE as a surrogate of the exact outcome of the forward model. We provide results for different degrees of the polynomial truncation, $N = 1, 2, 3$. A full tensor approach is used for $N = 1$ and $N = 2$, leading to a total number n_q of forward model runs equal to 16 and 81, respectively. Since with $N = 3$ we would have needed $n_q = 256$ full model runs, a sparse grid approach based on Smolyak’s coarse tensorization [54] is used, thus reducing n_q to 137. It should be noticed that the use of such sparse grid numerical integration might lead to unacceptable errors in the expansion coefficients with higher order polynomials [55]. This issue is not encountered in the present application, however alternative accurate approaches can be also employed if needed, such as those advanced in References [55,56].

A quantitative evaluation of the fitting quality of the gPCE surrogate is provided by employing the leave-one-out (LOO) cross-validation [57,58], where an estimate of the mean-square error of the residuals between the full model and surrogate solution is given by the SSE_{LOO} defined as:

$$SSE_{LOO} = \frac{1}{n_q} \sum_{i=1}^{n_q} (u_{z,i} - \hat{u}_{z,i}^i)^2, \tag{29}$$

where $u_{z,i}$ is the full model result at a certain time t and parameter combination i and $\hat{u}_{z,i}^i$ is the surrogate solution at the same t and i with the gPCE built without the point denoted by i . In practice,

the n_q residuals can be obtained without building n_q different gPCE but using the original gPCE constructed with the whole set of collocation points [57]. The coefficient Q^2 , similar to the coefficient of determination R^2 , can be defined as:

$$Q^2 = 1 - \frac{SSE_{LOO}}{SST}, \tag{30}$$

where SST is the sum of the squared deviations of $u_{z,i}$ from their mean values μ_u at the instant t . Values of Q^2 close to 1.0 indicates a good match between the model outcome and the gPCE surrogate approximation. Table 1 reports the mean and variance of vertical displacements at the times $t = 1, 5,$ and 10 years, along with the Q^2 value. Q^2 progressively increases to the upper limit of 1 for increasing N , proving the convergence of both the polynomial expansion and the high-dimensional quadrature formula. This holds true for $t > 5$ years with a gPCE-surrogate solution less accurate at the onset of the simulation.

Table 1. Mean μ_u , variance σ_u^2 and coefficient Q^2 with the generalized Polynomial Chaos Expansion (gPCE) approximation up to degree $N = 1, N = 2$ and $N = 3$ (sparse grid).

t [yr]	$N = 1$ ($n_q = 16$)			$N = 2$ ($n_q = 81$)			$N = 3$ ($n_q = 137$)		
	μ_u [m]	σ_u^2 [m ²]	Q^2	μ_u [m]	σ_u^2 [m ²]	Q^2	μ_u [m]	σ_u^2 [m ²]	Q^2
1	-2.1×10^{-3}	2.0×10^{-6}	0.820	-2.3×10^{-3}	3.1×10^{-6}	0.931	-2.3×10^{-3}	3.0×10^{-6}	0.823
5	-1.7×10^{-1}	1.4×10^{-2}	0.860	-1.7×10^{-1}	1.5×10^{-2}	0.995	-1.7×10^{-1}	1.5×10^{-2}	0.999
10	-3.6×10^{-1}	2.6×10^{-2}	0.904	-3.7×10^{-1}	2.8×10^{-2}	0.999	-3.7×10^{-1}	2.8×10^{-2}	1.000

3.3. Sensitivity Analysis

Before proceeding with the assimilation, it is important to evaluate the actual influence of the uncertain parameters on the model output. We compute the Sobol’ indices using the gPCE-surrogate solution of the model output. The results are presented in Figure 3. First-order indices clearly show the higher impact on the total variance with a negligible influence of the factors interactions. In particular, parameters λ^* and R appear to have the major impact on the solution at several times, except for the first two years where ρ_k and ρ_μ have a non-negligible effect. However, the significance of the quantity of interest, that is, vertical displacements, is quite limited at the beginning of the production period, because both the pressure variation and the geomechanical answer of the system are still very small and possibly influenced by the measurement errors. Thus, we limited our sensitivity analysis to the time interval between 3 and 10 years. In this temporal window, based on the total Sobol’ indices $S_{T,i}$ (Figure 3c,d) and setting $\epsilon_S = 0.1$, we can remove from the parameter space the ratios ρ_k and ρ_μ . However, it is important to point out that further analysis may be needed in order to investigate the model sensitivity based on different metrics, for example, on the expected value, skewness and kurtosis [59,60]. This may be helpful to better understand the role of each parameter, thus ensuring its exclusion from the parameter space is negligible.

The new parameter space is defined by $\mathbf{p} = (\lambda^*, R)$ with $n = 2$. This allows to reduce significantly the computational effort and, for the further analysis, retain the gPCE order to $N = 3$ with a total number of high-dimensional quadrature nodes $n_q = 16$. The Sobol’ indices for this new parameter space have the same behaviour as the one shown in Figure 3.

3.4. Bayesian Update

In this section, we compare the results of state/parameter update by the assimilation of surface displacements using the classical ES approach and its variation in the form of gPCE-ES. As already mentioned before, the synthetic “true” observations \mathbf{d}^t are collected from a geomechanical model run with the “true” parameter set and perturbed by a zero-mean Gaussian noise ϵ (Figure 2b). Different “true” parameter sets are investigated to test the robustness of the proposed approach.

The gPCE-ES technology is then applied within the iterative ES-MDA framework using either constant and decreasing factors α_{MDA} to inflate the covariance matrix of measurement error.

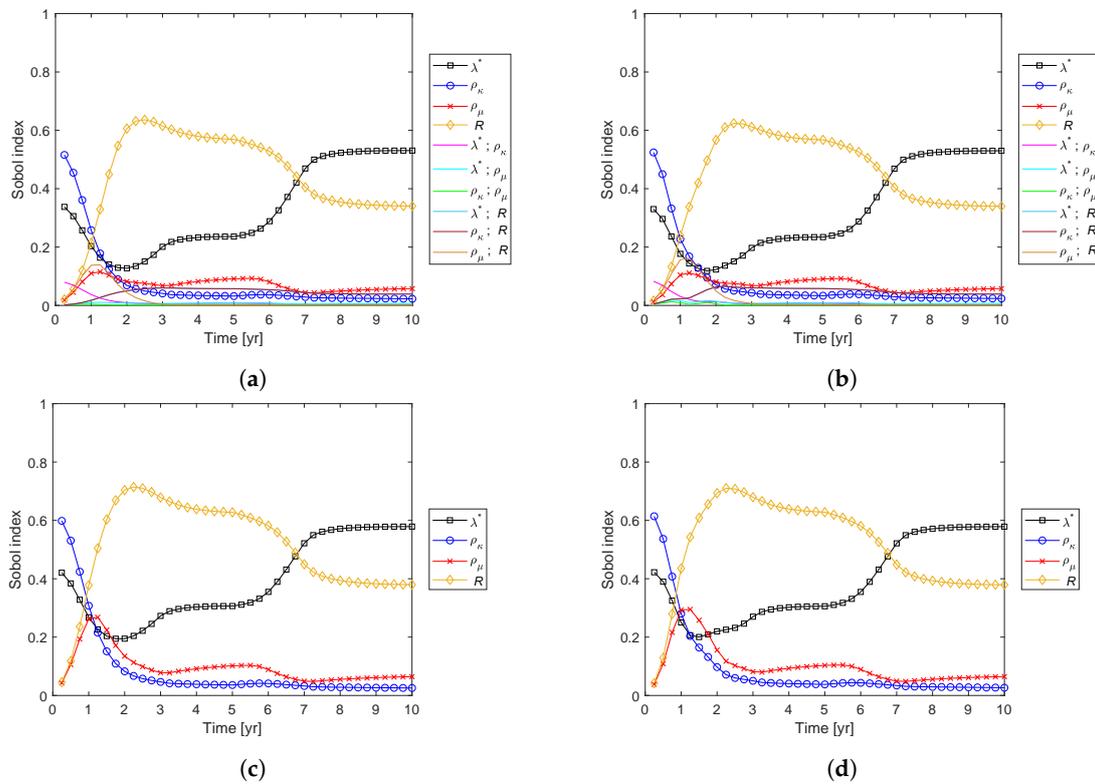


Figure 3. Sobol’ indices versus simulated time. Profiles with markers are used for the single parameters, while continuous profiles are used for the combinations of parameters. (a) Case with $n = 4$ parameters and gPCE degree $N = 2$. (b) Case with $n = 4$ parameters and gPCE degree $N = 3$ with a sparse grid. (c) The same as (a) considering the total effect. (d) The same as (b) considering the total effect.

3.4.1. ES and gPCE-ES

We assimilate a time series of vertical displacements collected at a point on the top surface located approximately at the center of the reservoir (Figure 2a). The synthetic observations are given by the geomechanical model run with what we assume to be the “true” parameter set ($\lambda^* = 0.011, R = 1.25$). The purpose is to constrain the forecast response of both state variables, for example, surface displacements and material parameters.

Figure 4 shows the prior and updated ensembles of the uncertain parameters, λ^* (left sub-panels) and R (right sub-panels), obtained by progressively increasing the number of MC realizations. The assimilation is carried out by both the standard ES with the full forward model (red dots) and the proposed gPCE-ES approach (blue dots). As expected, the mean of the updated ensembles tends to approach the “true” value for both parameters, though with possible oscillations far from convergence, for example at $n_{MC} = 50$. We provide the results for $n_{MC} = (16, 25, 41, 50, 75, 100)$ in order to analyze the cases when the number of realizations is either smaller than, equal to, or larger than, the number of observations $m = 41$. Since the total cost of an ES application can be regarded as roughly equal to the number of forward model runs, we can measure the computational cost, η , just by counting the number of MC simulations. Therefore, the computational cost η_g associated to gPCE-ES is constant and equal to 16, that is, the number n_q of forward model runs required to compute the surrogate gPCE model. By distinction, with the standard ES the computational cost, η_f , increases with n_{MC} . The uncertainty reduction is significantly improved if $n_{MC} \geq m$. In case $n_{MC} < m$ we retain 99% of the non-zero singular values computed to invert the matrix in Equation (26). Generally speaking, the parameter R better approaches the “true” value as compared to λ^* .

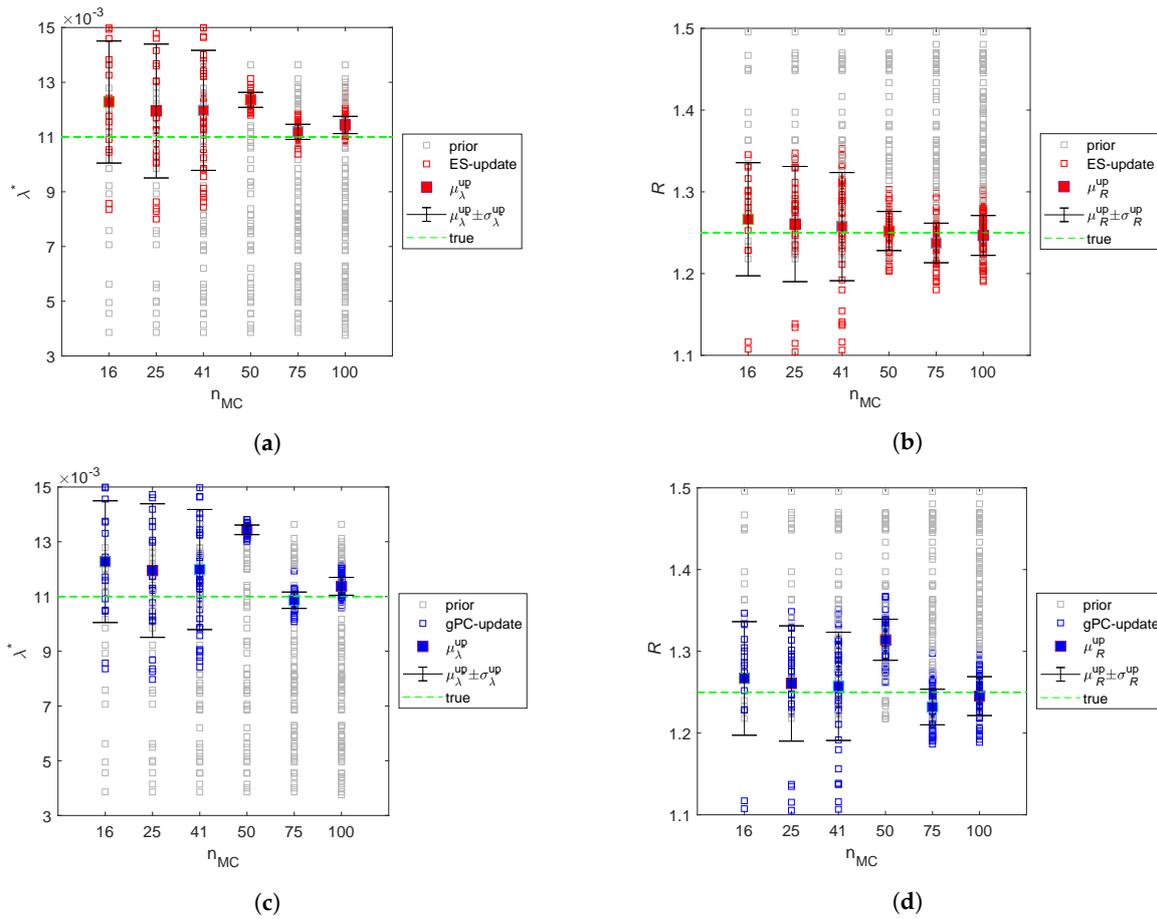


Figure 4. Parameter ensembles for increasing number of Monte Carlo realizations n_{MC} . Panels (a) and (b) use the standard Ensemble Smoother (ES) with the full forward model for λ^* and R , respectively. Panels (c,d) provide the same outcome as (a,b) by using the gPCE-ES approach. μ and σ are the mean and standard deviation of the updated (up) ensembles.

Figures 5 and 6 show the update of the state variable u_z , that is, the vertical displacement in time at the monitoring point, by increasing n_{MC} with ES and gPCE-ES, respectively. The observations are generally well-captured by both approaches, with no significant differences between the methods. Hence, the proposed surrogate model appears to capture very well the overall behaviour of the forward model. To measure the quality of the approximation provided by gPCE-ES, we define the metrics AE (average error) and AES (average ensemble spread) defined as:

$$AE = \frac{1}{n_{MC}} \sum_{k=1}^{n_{MC}} |\psi_k - \psi^{true}| \quad (31)$$

$$AES = \frac{1}{n_{MC}} \sum_{k=1}^{n_{MC}} |\psi_k - \mu_\psi|, \quad (32)$$

where ψ generally stands for state variables or parameters, and μ_ψ its mean. The metrics are computed both for prior and updated ensembles of states and parameters, providing a quantitative measure of the assimilation quality. For the state u_z , time-averaged values are used to compute the metrics. AE and AES reductions (Table 2) larger than 90% are obtained for both ES and gPCE-ES, thus confirming that the surrogate model is able to reproduce almost exactly the full forward model behaviour.

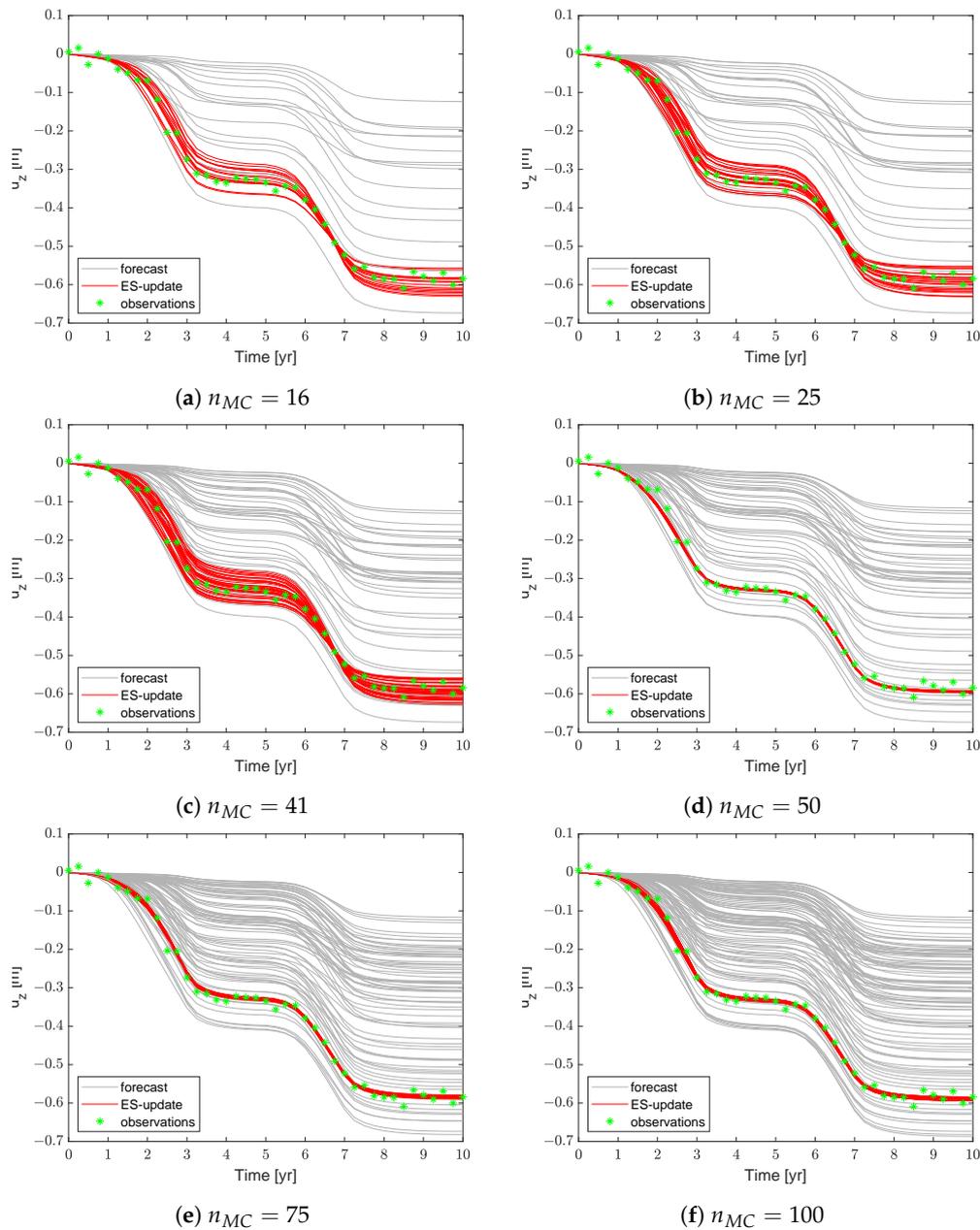


Figure 5. State ensembles for increasing number of Monte Carlo realizations n_{MC} . In this case, the forecast ensemble (the grey one) has been obtained through the geomechanical model. The update ensemble (in red) is directly the output of the ES technique, that is, it does not derive from the geomechanical model run with the update ensembles of the parameters.

Based on these outcomes, we consider now only the gPCE-ES framework. Figure 7a shows the gPCE-ES convergence with n_{MC} . We recall that the computational cost is roughly constant, $\eta_g = 16$. The performance index AE^u decreases with n_{MC} for both parameters and stabilizes after $\sim 100 n_{MC}$. If a different “true” parameter set is used, for example, $(\lambda^* = 0.005, R = 1.5)$, we obtain the behaviour shown in Figure 7b. In this case, the mean converges to the “true” value at a much faster rate for λ^* , while R remains poorly constrained. The “true” parameter sets are chosen such that they provide very different “true” reference simulations, as shown in Figure 2b.

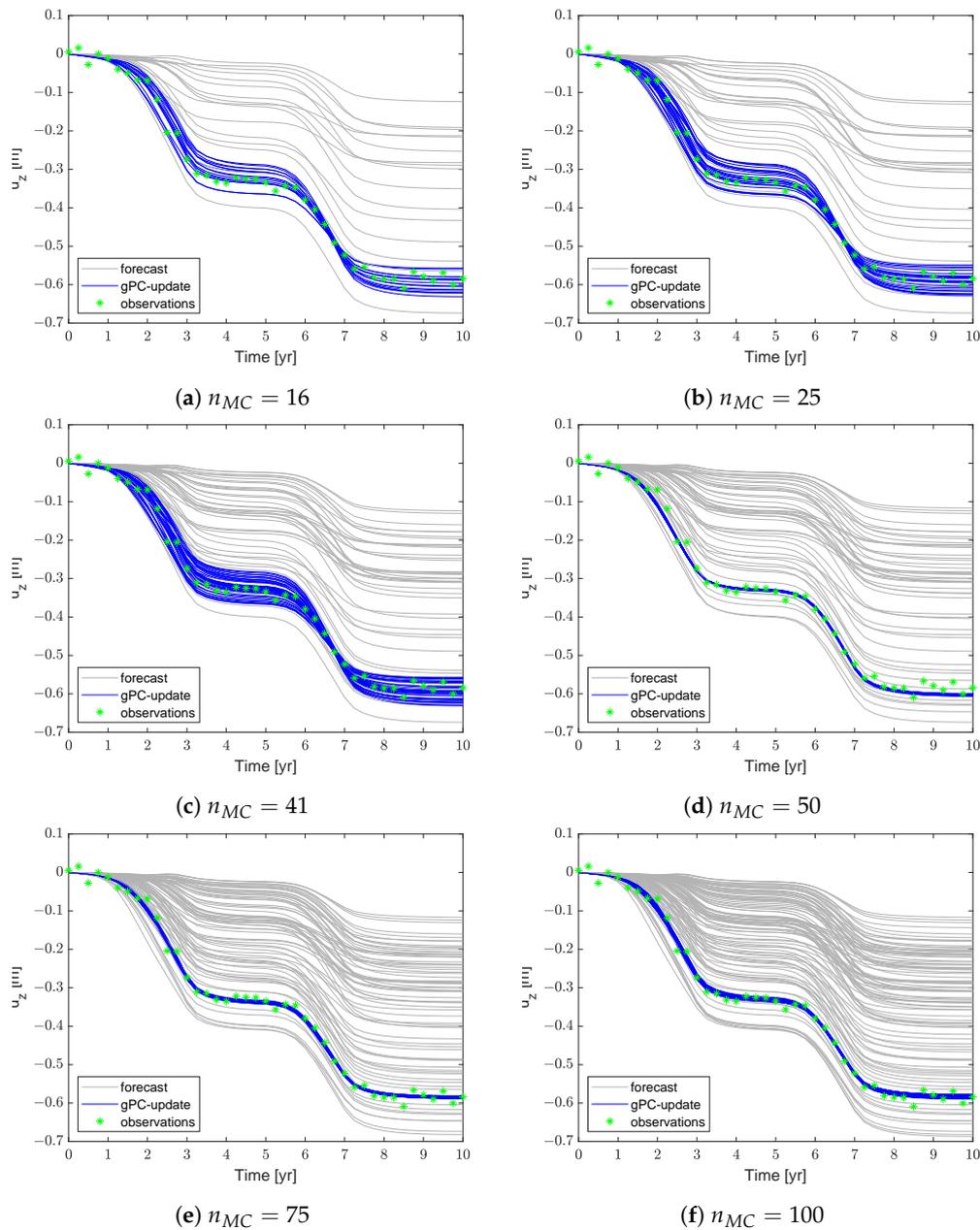


Figure 6. The same as Figure 5 with the forecast ensembles, in grey, obtained through the gPC expansion.

3.4.2. gPCE-MDA

The ES-MDA approach is based on a multiple assimilation process to provide a better model constraint especially on strongly non-linear models. Despite the potential improvement with respect to the standard ES method, the MDA approach is rarely used in practice mainly because of the large computational cost associated to the generation of several ensembles. The introduction of an inexpensive surrogate model in place of the full forward model can make also the MDA technique more affordable in real-world problems.

We consider two test cases differing in the selection of the inflation factors α_{MDA} : (i) constant $\alpha_{MDA} = n_{MDA}$, and (ii) decreasing $\alpha_{MDA} = (18.66, 14, 8, 4, 2)$. The number of successive assimilations is fixed a-priori and equal to $n_{MDA} = 5$. Each assimilation step is performed by using ensembles with $n_{MC} = 3000$ realizations. This ensemble size allows to achieve the convergence also with the ES technique, as shown in Figure 7. Figure 8 shows the mean and the standard deviation of the updated parameter ensembles in the five successive assimilation steps considering the “true” parameter set

$\mathbf{p}^t = (0.005, 1.5)$. In this case, the differences with respect to a single ES assimilation (blue bar in Figure 8) are evident, that is, with multiple assimilation steps the solution appears to better approach the truth and with a reduced variability spread. This behaviour is clearly noticeable for the parameter R (Figure 8b). Similar to ES, the MDA effectiveness seems to vary when the set of observations changes. For example, with the choice of $\mathbf{p}^t = (\lambda^* = 0.011, R = 1.25)$ the improvement of multiple assimilation steps with respect to a single ES update is not significant.

Table 2. Values of AE and AES (prior and update) for the state and parameter ensembles computed by ES and gPCE-ES.

	ES			gPCE-ES		
	$n_{MC} = 25$	$n_{MC} = 50$	$n_{MC} = 100$	$n_{MC} = 25$	$n_{MC} = 50$	$n_{MC} = 100$
$AE^p(\mathbf{u}_z)$ [m]	0.1667	0.1642	0.1656	0.1667	0.1642	0.1656
$AE^u(\mathbf{u}_z)$ [m]	0.0190	0.0104	0.0099	0.0189	0.0117	0.0101
$AES^p(\mathbf{u}_z)$ [m]	0.0834	0.0836	0.0803	0.0833	0.0836	0.0803
$AES^u(\mathbf{u}_z)$ [m]	0.0149	0.0010	0.0016	0.0149	0.0010	0.0017
$AE^p(\lambda^*)$	0.0028	0.0029	0.0029	0.0028	0.0029	0.0029
$AE^u(\lambda^*)$	0.0021	0.0014	0.0005	0.0021	0.0024	0.0004
$AES^p(\lambda^*)$	0.0027	0.0025	0.0024	0.0027	0.0025	0.0024
$AES^u(\lambda^*)$	0.0020	0.0002	0.0002	0.0020	0.0001	0.0003
$AE^p(R)$	0.1109	0.1065	0.1069	0.1109	0.1065	0.1069
$AE^u(R)$	0.0567	0.0186	0.0192	0.0568	0.0641	0.0184
$AES^p(R)$	0.0787	0.0724	0.0729	0.0787	0.0724	0.0729
$AES^u(R)$	0.0538	0.0186	0.0188	0.0538	0.0196	0.0182

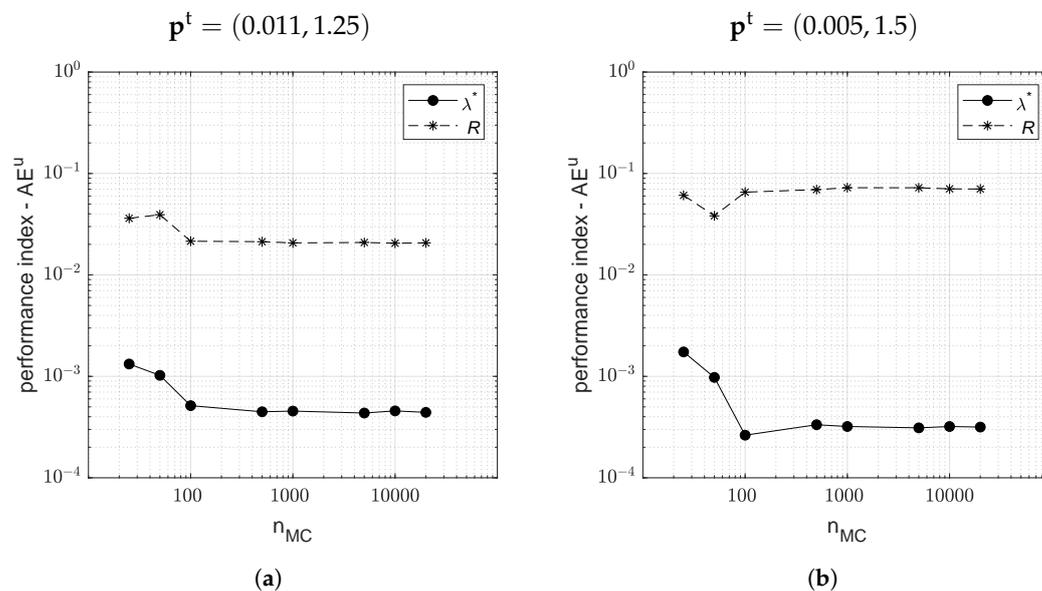


Figure 7. Convergence of gPCE-ES with increasing ensemble size for (a) $\mathbf{p}^t = (0.011, 1.25)$ and (b) $\mathbf{p}^t = (0.005, 1.5)$.

The choice of the sequence of inflation factors α_{MDA} does not modify significantly the final assimilation outcome (Figure 8).

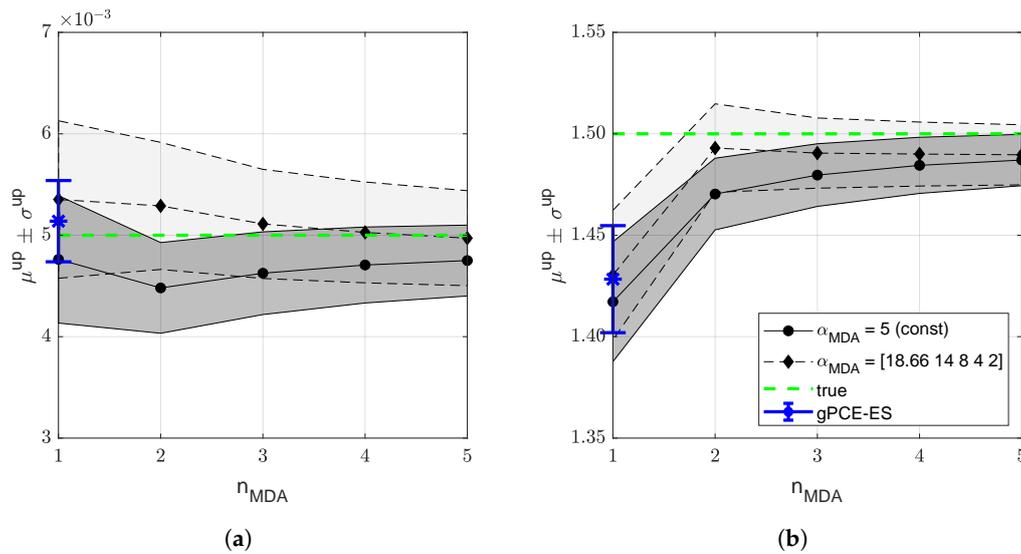


Figure 8. Mean and standard deviation of the updated ensembles obtained with gPCE-MDA in case of constant and decreasing α_{MDA} for parameter (a) λ^* and (b) R considering the “true” parameter set $\mathbf{p}^t = (0.005, 1.5)$. The blue bar refers to the results obtained with the gPCE-ES algorithm.

4. Discussion and Conclusions

In this work, we develop a gPCE-based framework to quantify and reduce the prediction uncertainty in geomechanical problems. The idea is to exploit the low computational cost of a gPCE-based surrogate model to—(i) perform a global sensitivity analysis and capture the material parameters mostly affecting the expected prediction, and (ii) update the variation ranges of both the state variables of interest and the governing parameters via data assimilation. We use an ensemble-based approach, such as the ES and its variation in the MDA form. Due to the high computational cost of the forward model runs required by the ensemble approaches, the proposed gPCE-ES and gPCE-MDA algorithms, if able to effectively approximate the full model behaviour, can allow for the use of large ensembles and successive assimilations at an affordable cost. The proposed approach is tested in a large-scale geomechanical model dealing with the prediction of land subsidence due to a hydrocarbon production program from a deep reservoir. To describe the constitutive behaviour of the deep rocks, we use a strongly non-linear visco-elasto-plastic model based on the VN law [3,43]. We show that high-accuracy surrogate models can be achieved via gPCE. Through the computation of the Sobol’ indices, in the specific application considered herein we demonstrate that the material parameters mainly impacting on the outcome of interest, that is, land subsidence in time, are those controlling the rock deformability in loading conditions and the size of the yield surface, λ^* and R . The initial variability ranges of such parameters are progressively reduced through data assimilation considering: (i) standard ES, (ii) gPCE-ES, and (iii) gPCE-MDA. The gPCE-ES provides highly accurate results comparable to those obtained with the native ES approach but at a significantly lower computational cost. It is worth recalling that the high efficiency of gPCE is mostly due to the low dimension of the parameter space. In fact, the gPCE approach suffers from the so-called “curse of dimensionality” due to the exponential increase of the cost with the number of independent stochastic variables.

The gPCE-ES is used to test the ES convergence rate at an increasing size of the ensemble. By fast polynomial evaluation we can analyse ensembles up to 20,000 realizations, which are not affordable without a surrogate model. In general, the parameter mean approaches the “true” solution with increasing MC realizations. Figure 9 shows that ensemble sizes greater than ~ 2000 realizations do not provide any significant improvement on the update of the state variable, that is, the mean-ensemble value of the surface displacements coincides with the observation vector. This also implies a limitation to the best match of the parameter.

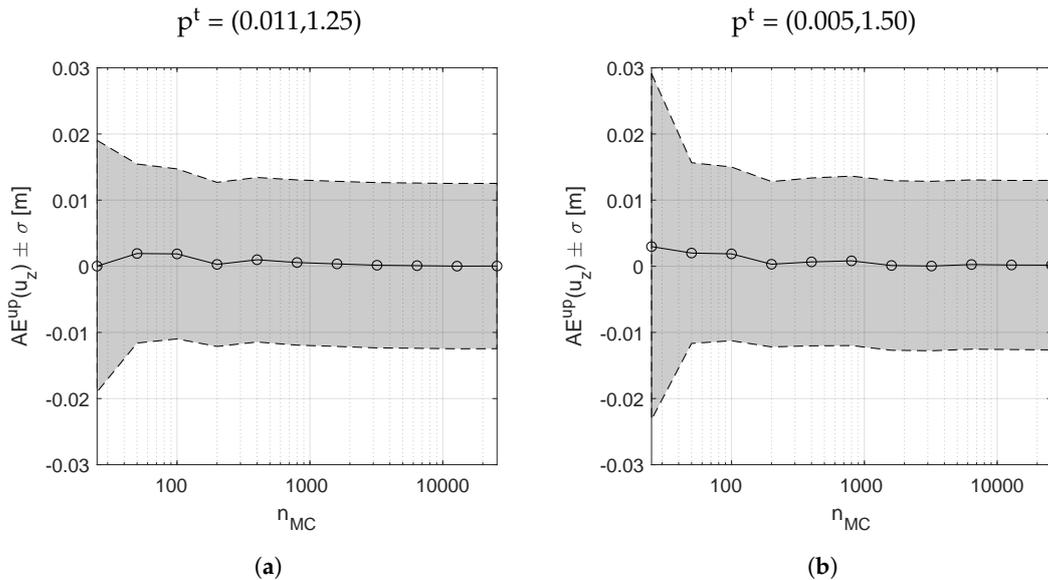


Figure 9. Mean and standard deviation for the updated displacements on the observation point varying the size of the ensembles.

We analyse the gPCE-ES convergence rate by using two parameter sets to define the “true” reference observations, with values located in the middle and on the boundary of the expected variability ranges. Different results are obtained, with a poor constraint for R when the “true” value is closer to the upper bound of the variability range. Then, we test again these sets with the gPCE-MDA approach by repeatedly assimilating the same observations a number of times with an inflated measurement error covariance matrix. We use both constant and decreasing values for the inflating factors, without any significant difference between the two approaches. The quality of the results for constraining R is significantly improved. The gPCE-MDA is obtained at the cost of computing the gPCE coefficients only, and by replacing the forward model runs required at any MDA step with fast polynomial evaluations. We demonstrate that MDA may outperform the standard ES and this option can become particularly attractive if combined with the gPCE approximation, which reduces considerably the required computational effort.

Our results show different parameter-constrain capabilities of the proposed approach depending on the choice of the reference “true” point in the parameter space. Figure 10 illustrates this behaviour in the space defined by parameters λ^* and R . We randomly sampled 500 “true” reference parameter sets and perform the data assimilation by means of the gPCE-ES with $n_{MC} = 3000$ realizations and constant noise standard error. Figure 10a,b show the error $e = |\mu^u - p^t|$ between the mean updated value and the true value for parameters λ^* and R , respectively. Figure 10c shows the maximum vertical displacements assimilated at final simulation time. Clearly the $u_{z,max}$ increases by increasing λ^* and diminishing R . Generally, λ^* (Figure 10a) presents higher errors with $1.35 < R < 1.45$ and the parameter seems to better approach the true reference value with increasing values of $u_{z,max}$. The constrain on parameter R is more complex to analyse. A poor match is obtained at the four corners of the parameter space and minimum errors are obtained for $u_{z,max} \simeq -0.6$ m and $u_{z,max} \simeq -0.3$ m. Based on these results, further analyses would be needed to investigate the role of the parameter interactions on the model outcome linked to the influence of the varying time-behaviour parameter importance, with model non-linearity and measurements noise providing for additional crucial factors affecting the inverse problem solution.

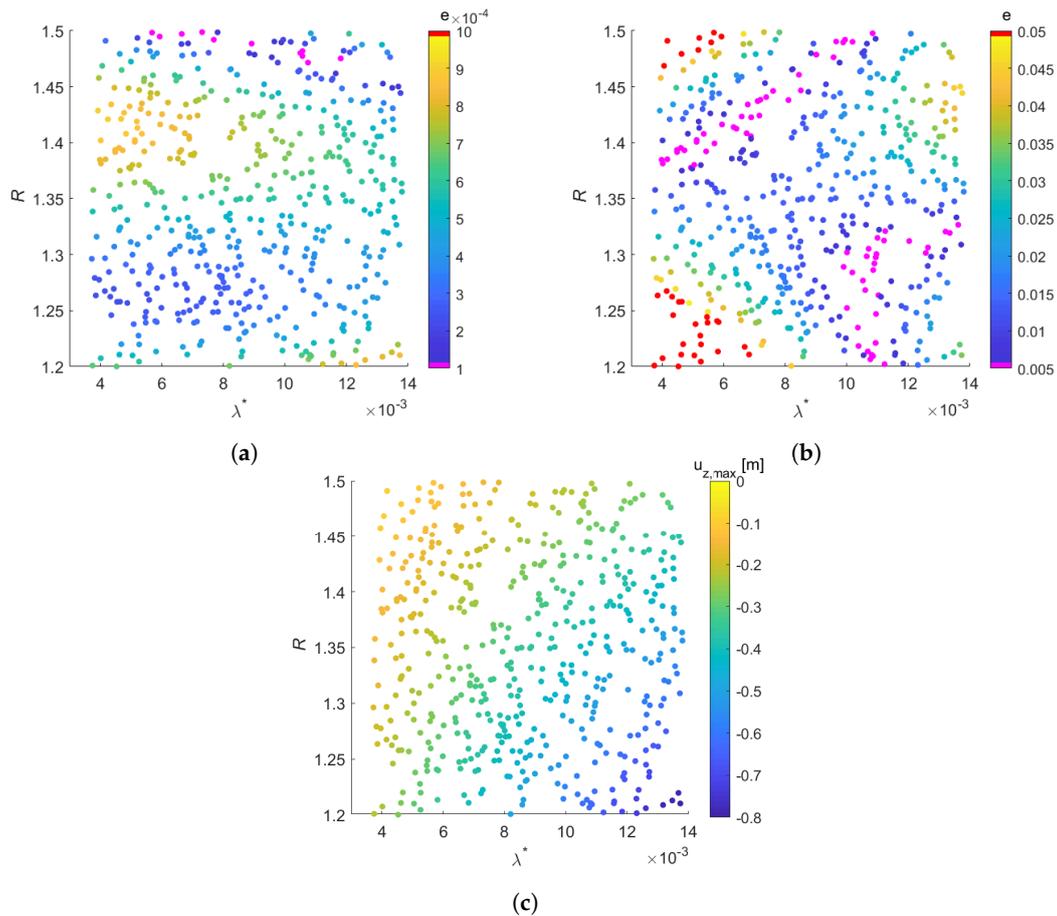


Figure 10. Error $e = |\mu^u - p^t|$ between the mean updated value and the true value for parameters (a) λ^* and (b) R at 500 randomly-sampled ‘true’ reference simulations. Sub-panel (c) shows the maximum subsidence assimilated at each assimilation.

Author Contributions: Conceptualization, C.Z., L.G., and M.F.; methodology, C.Z.; software, C.Z. and L.G.; data curation, L.G.; writing–original draft preparation, C.Z. and L.G.; writing–review and editing, M.F. and P.T.; supervision, M.F. and P.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The Authors wish to thank the University of Padova Strategic Research Infrastructure Grant 2017 ‘‘CAPRI: Calcolo ad Alte Prestazioni per la Ricerca e l’Innovazione’’ for the availability of the computational resources used in this work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UQ	Uncertainty Quantification
SA	Sensitivity Analysis
ES	Ensemble Smoother
FE	Finite Element
gPCE	generalized Polynomial Chaos Expansion
MC	Monte Carlo
MDA	Multiple Data Assimilation
PSI	Permanent Scatterer Interferometry
VN	Vermeer-Neher

References

1. Gambolati, G.; Ferronato, M.; Teatini, P.; Deidda, R.; Lecca, G. Finite element analysis of land subsidence above depleted reservoirs with pore pressure gradient and total stress formulations. *Int. J. Numer. Anal. Methods Geomech.* **2001**, *25*, 307–327. [[CrossRef](#)]
2. Spiezia, N.; Ferronato, M.; Janna, C.; Teatini, P. A two-invariant pseudoelastic model for reservoir compaction. *Int. J. Numer. Anal. Methods Geomech.* **2017**, *41*, 1870–1893, doi:10.1002/nag.2704. [[CrossRef](#)]
3. Isotton, G.; Teatini, P.; Ferronato, M.; Janna, C.; Spiezia, N.; Mantica, S.; Volontè, G. Robust numerical implementation of a 3D rate-dependent model for reservoir geomechanical simulations. *Int. J. Numer. Anal. Methods Geomech.* **2019**, *43*, 2752–2771, doi:10.1002/nag.3000. [[CrossRef](#)]
4. Lanier, J.; Caillerie, D.; Chambon, R.; Viggiani, G.; Bésuelle, P.; Desrues, J. A general formulation of hypoplasticity. *Int. J. Numer. Anal. Methods Geomech.* **2004**, *28*, 1461–1478. [[CrossRef](#)]
5. Chandong, C.; Zoback, M. Viscous creep in room-dried unconsolidated Gulf of Mexico shale (II): development of a viscoplasticity model. *J. Pet. Sci. Eng.* **2010**, *72*, 50–55. [[CrossRef](#)]
6. Oka, F.; Kimoto, S.; Higo, Y.; Ohta, H.; Sanagawa, T.; Kodaka, T. An elasto-viscoplastic model for diatomaceous mudstone and numerical simulation of compaction bands. *Int. J. Numer. Anal. Methods Geomech.* **2011**, *35*, 244–263. [[CrossRef](#)]
7. Cassiani, G.; Brovelli, A.; Hueckel, T. A strain-rate-dependent modified Cam-Clay model for the simulation of soil/rock compaction. *Geomech. Energy Environ.* **2017**, *11*, 42–51. [[CrossRef](#)]
8. Nguyen, S.K.; Volontè, G.; Musso, G.; Brignoli, M.; Gemelli, F.; Mantica, S. Implementation of an elasto-viscoplastic constitutive law in Abaqus/Standard for an improved characterization of rock materials. In Proceedings of the 51st U.S. Rock Mechanics/Geomechanics Symposium, San Francisco, CA, USA, 25–28 June 2017.
9. Ferronato, M.; Castelletto, N.; Gambolati, G.; Janna, C.; Teatini, P. II cycle compressibility from satellite measurements. *Geotechnique* **2013**, *63*, 479. [[CrossRef](#)]
10. Zoccarato, C.; Baù, D.; Ferronato, M.; Gambolati, G.; Alzraiee, A.; Teatini, P. Data assimilation of surface displacements to improve geomechanical parameters of gas storage reservoirs. *J. Geophys. Res. Solid Earth* **2016**, *121*, doi:10.1002/2015JB012090. [[CrossRef](#)]
11. Zoccarato, C.; Baù, D.; Bottazzi, F.; Ferronato, M.; Gambolati, G.; Mantica, S.; Teatini, P. On the importance of the heterogeneity assumption in the characterization of reservoir geomechanical properties. *Geophys. J. Int.* **2016**, *207*, 47–58, doi:10.1093/gji/ggw259. [[CrossRef](#)]
12. Fokker, P.A.; Wassing, B.B.; van Leijen, F.J.; Hanssen, R.F.; Nieuwland, D.A. Application of an ensemble smoother with multiple data assimilation to the Bergermeer gas field, using PS-InSAR. *Geomech. Energy Environ.* **2016**, *5*, 16–28, doi:10.1016/j.gete.2015.11.003. [[CrossRef](#)]
13. Gazzola, L.; Ferronato, M.; Frigo, M.; Janna, C.; Teatini, P.; Zoccarato, C.; Antonelli, M.; Corradi, A.; Dacome, M.; Mantica, S. Uncertainty quantification and reduction through Data Assimilation approaches for the geomechanical modeling of hydrocarbon reservoirs. In Proceedings of the 53rd US Rock Mechanics Geomechanics Symposium ARMA, New York, NY, USA, 23–26 June 2019.
14. Emerick, A.; Reynolds, A. Ensemble smoother with multiple data assimilation. *Comput. Geosci.* **2013**, *55*, 3–15, doi:10.1016/j.cageo.2012.03.011. [[CrossRef](#)]
15. Bottazzi, F.; Della Rossa, E. A Functional Data Analysis Approach to Surrogate Modeling in Reservoir and Geomechanics Uncertainty Quantification. *Math. Geosci.* **2017**, *49*, 517–540, doi:10.1007/s11004-017-9685-y. [[CrossRef](#)]
16. Botti, M.; Pietro, D.A.D.; Maître, O.L.; Sochala, P. Numerical approximation of poroelasticity with random coefficients using Polynomial Chaos and Hybrid High-Order methods. *Comput. Methods Appl. Mech. Eng.* **2019**, *361*, 112736, doi:10.1016/j.cma.2019.112736. [[CrossRef](#)]
17. Ganesh, A.; Zhang, B.; Chalaturnyk, R.J.; Prasad, V. Uncertainty quantification of the factor of safety in a steam-assisted gravity drainage process through polynomial chaos expansion. *Comput. Chem. Eng.* **2020**, *133*, 106663, doi:10.1016/j.compchemeng.2019.106663. [[CrossRef](#)]
18. Castiñeira, D.; Jha, B.; Juanes, R. Uncertainty Quantification and Inverse Modeling of Fault Poromechanics and Induced Seismicity: Application to a Synthetic Carbon Capture and Storage (CCS) Problem. In Proceedings of the ARMA-2016-151, American Rock Mechanics Association, 50th U.S. Rock Mechanics/Geomechanics Symposium, Houston, TX, USA, 26–29 June 2016.

19. Zoccarato, C.; Ferronato, M.; Franceschini, A.; Janna, C.; Teatini, P. Modeling fault activation due to fluid production: Bayesian update by seismic data. *Comput. Geosci.* **2019**, *23*, 705–722. [[CrossRef](#)]
20. Verde, A. Global Sensitivity Analysis of Geomechanical Fractured Reservoir Parameters. In Proceedings of the 49th U.S. Rock Mechanics/Geomechanics Symposium, San Francisco, CA, USA, 28 June–1 July 2015; American Rock Mechanics Association: Alexandria, VA, USA, 2015.
21. Rezaei, A.; Nakshatrala, K.B.; Siddiqui, F.; Dindoruk, B.; Soliman, M. A global sensitivity analysis and reduced-order models for hydraulically fractured horizontal wells. *Comput. Geosci.* **2020**, doi:10.1007/s10596-019-09896-7. [[CrossRef](#)]
22. Wiener, N. The homogeneous chaos. *Am. J. Math.* **1938**, *60*, 897–936, doi:10.2307/2371268. [[CrossRef](#)]
23. Ghanem, R.G.; Spanos, P.D. *Stochastic Finite Elements: A Spectral Approach*; Springer: New York, NY, USA, 1991; (Reedited by Dover Publications: Mineola: NY, USA, 2003).
24. Xiu, D.; Karniadakis, G. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **2002**, *24*, 619–644. [[CrossRef](#)]
25. Najm, H.N. Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annu. Rev. Fluid Mech.* **2009**, *41*, 35–52, doi:10.1146/annurev.fluid.010908.165248. [[CrossRef](#)]
26. Le Maître, O.P.; Knio, O.M. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*; Scientific Computation, Springer: Dordrecht, The Netherlands, 2010, doi:10.1007/978-90-481-3520-2. [[CrossRef](#)]
27. Xiu, D. *Numerical methods for stochastic computations A Spectral Method Approach*; Princeton University Press: Princeton, NJ, USA, 2010.
28. Ghanem, R. Stochastic finite elements with multiple random non-Gaussian properties. *J. Eng. Mech.* **1999**, *125*, 26–40. [[CrossRef](#)]
29. Sudret, B.; Berveiller, M. Stochastic finite element methods in geotechnical engineering. In *Reliability-Based Design in Geotechnical Engineering: Computations and Applications*; Taylor & Francis: New York, NY, USA, 2008; pp. 260–297.
30. Li, W.; Lu, Z.; Zhang, D. Stochastic analysis of unsaturated flow with probabilistic collocation method. *Water Resour. Res.* **2009**, *45*, 1–13, doi:10.1029/2008WR007530. [[CrossRef](#)]
31. Fajraoui, N.; Ramasomanana, F.; Younes, A.; Mara, T.A.; Ackerer, P.; Guadagnini, A. Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media. *Water Resour. Res.* **2011**, *47*, W02521, doi:10.1029/2010WR009639. [[CrossRef](#)]
32. Oladyshkin, S.; Class, H.; Helmig, R.; Nowak, W. An integrative approach to robust design and probabilistic risk assessment for CO₂ storage in geological formations. *Comput. Geosci.* **2011**, *15*, 565–577. [[CrossRef](#)]
33. Formaggia, L.; Guadagnini, A.; Imperiali, I.; Lever, V.; Porta, G.; Riva, M.; Scotti, A.; Tamellini, L. Global sensitivity analysis through polynomial chaos expansion of a basin-scale geochemical compaction model. *Comput. Geosci.* **2013**, *17*, 25–42, doi:10.1007/s10596-012-9311-5. [[CrossRef](#)]
34. Deman, G.; Konakli, K.; Sudret, B.; Kerrou, J.; Perrochet, P.; Benabderrahmane, H. Using sparse polynomial chaos expansions for the global sensitivity analysis of groundwater lifetime expectancy in a multi-layered hydrogeological model. *Reliab. Eng. Syst. Saf.* **2016**, *147*, 156–169. [[CrossRef](#)]
35. Maina, F.Z.; Guadagnini, A. Uncertainty quantification and global sensitivity analysis of subsurface flow parameters to gravimetric variations during pumping tests in unconfined aquifers. *Water Resour. Res.* **2018**, *54*, 501–518. [[CrossRef](#)]
36. Li, J.; Xiu, D. A generalized polynomial chaos based ensemble Kalman filter with high accuracy. *J. Comput. Phys.* **2009**, *228*, 5454–5469. [[CrossRef](#)]
37. Marzouk, Y.; Xiu, D. A stochastic collocation approach to Bayesian inference in inverse problems. *Commun. Comput. Phys.* **2009**, *6*, 826–847. [[CrossRef](#)]
38. Saad, G.; Ghanem, R. Characterization of reservoir simulation models using a polynomial chaos-based ensemble Kalman filter. *Water Resour. Res.* doi:10.1029/2008WR007148. Available online: <https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2008WR007148> (accessed on 15 May 2020). [[CrossRef](#)]
39. Oladyshkin, S.; Class, H.; Nowak, W. Bayesian updating via bootstrap filtering combined with data-driven polynomial chaos expansions: Methodology and application to history matching for carbon dioxide storage in geological formations. *Comput. Geosci.* **2013**, *17*, 671–687, doi:10.1007/s10596-013-9350-6. [[CrossRef](#)]

40. Elsheikh, A.H.; Hoteit, I.; Wheeler, M.F. Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates. *Comput. Methods Appl. Mech. Eng.* **2014**, *269*, 515–537. [[CrossRef](#)]
41. Biot, M.A. General theory of three-dimensional consolidation. *J. Appl. Phys.* **1941**, *12*, 155–164. [[CrossRef](#)]
42. Coussy, O. *Poromechanics*; John Wiley & Sons: Hoboken, NJ, USA, 2003.
43. Vermeer, P.; Neher, H. A soft soil model that accounts for creep. In Proceedings of the International Symposium “Beyond 2000 in Computational Geotechnics”, Amsterdam, The Netherlands, 18–20 March 1999; pp. 249–261.
44. Xiu, D. Efficient collocational approach for parametric uncertainty analysis. *Commun. Comput. Phys.* **2007**, *2*, 293–309.
45. Sobol’, I. Sensitivity estimates for nonlinear mathematical models. *Mat. Model.* **1990**, *2*, 112–118. in English in Sobol’ (1993)
46. Saltelli, A.; Marco, R.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global Sensitivity Analysis. The Primer*; John Wiley & Sons, Ltd.: Chichester, UK, 2007; doi:10.1002/9780470725184.fmatter. [[CrossRef](#)]
47. Sudret, B. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.* **2008**, *93*, 964–979, doi:10.1016/j.res.2007.04.002. [[CrossRef](#)]
48. Crestaux, T.; Le Maître, O.; Martinez, J.M. Polynomial chaos expansion for sensitivity analysis. *Reliab. Eng. Syst. Saf.* **2009**, *94*, 1161–1172, doi:10.1016/j.res.2008.10.008. [[CrossRef](#)]
49. Van Leeuwen, P.; Evensen, G. Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather. Rev.* **1996**, *124*, 2898–2913. [[CrossRef](#)]
50. Emerick, A.; Reynolds, A. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilation. *Comput. Geosci.* **2012**, *16*, 639–659, doi:10.1007/s10596-012-9275-5. [[CrossRef](#)]
51. Emerick, A. Analysis of the performance of ensemble-based assimilation of production and seismic data. *J. Pet. Sci. Eng.* **2016**, *139*, 219–239, doi:10.1016/j.petrol.2016.01.029. [[CrossRef](#)]
52. Evensen, G. Analysis of iterative ensemble smoothers for solving inverse problems. *Comput. Geosci.* **2018**, *22*, 885–908, doi:10.1007/s10596-018-9731-y. [[CrossRef](#)]
53. Baù, D.; Ferronato, M.; Gambolati, G.; Teatini, P. Basin-scale compressibility of the Northern Adriatic by the radioactive marker technique. *Geotechnique* **2002**, *52*, 605–616. [[CrossRef](#)]
54. Smolyak, S.A. Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions. *Dokl. Akad. Nauk SSSR* **1963**, *148*, 1042–1043.
55. Constantine, P.; Eldred, M.; Phipps, E. Sparse pseudospectral approximation method. *Comput. Methods Appl. Mech. Eng.* **2012**, 229–232, 1–12, doi:10.1016/j.cma.2012.03.019. [[CrossRef](#)]
56. Conrad, P.R.; Marzouk, Y.M. Adaptive Smolyak pseudospectral approximations. *SIAM J. Sci. Comput.* **2013**, *35*, A2643–A2670. [[CrossRef](#)]
57. Blatman, G. Adaptive Sparse Polynomial Chaos Expansions for Uncertainty Propagation and Sensitivity Analysis. Ph.D. Thesis, Université Blaise Pascal, Clermont-Ferrand, France, 2009.
58. Sudret, B. Polynomial chaos expansions and stochastic finite element methods. In *Risk and Reliability in Geotechnical Engineering*; CRC Press: Boca Raton, FL, USA, 2014; pp. 265–300.
59. Borgonovo, E. A new uncertainty importance measure. *Reliab. Eng. Syst. Saf.* **2007**, *92*, 771–784. [[CrossRef](#)]
60. Dell’Oca, A.; Riva, M.; Guadagnini, A. Moment-based metrics for global sensitivity analysis of hydrological systems. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 6219–6234, doi:10.5194/hess-21-6219-2017. [[CrossRef](#)]

