MOX–Report No. 37/2013

Characterization of basin-scale systems under mechanical and geochemical compaction

Lever, V.; Porta, G.; Tamellini, L.; Riva, M.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Characterization of basin-scale systems under mechanical and geochemical compaction

Characterization of mechanical and geochemical compaction

By V. Lever[1], G. Porta[1], L. Tamellini[2,3], M. Riva[1,4]

[1]Dipartimento di Ingegneria Civile e Ambientale Politecnico di Milano, Piazza L. Da Vinci 32,
20133 Milano, Italy
[2] MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza L. Da Vinci 32, 20133
Milano, Italy
[3] CSQI-MATHICSE, Ecole Polytechnique Fédérale de Lausanne, Station 8, CH 1015, Lausanne,
Switzerland
[4]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona 85721,
USA

ABSTRACT

We present an inverse modeling procedure for the calibration of uncertain model parameters characterizing basin scale sandstone compaction due to mechanical and geochemical processes. Unknown model parameters include geophysical and geochemical system attributes as well as pressure and temperature boundary conditions. We derive a reduced model of the system based on the generalized polynomial chaos expansion (gPCE) approximation method and compute the variance-based Sobol indices for the selected uncertain parameters. The gPCE is used to approximate the model response at a low computational cost and the Sobol indices quantify the effect of each uncertain parameter on the state variables. Parameter estimation is performed within a Maximum Likelihood framework. Results are illustrated on a one-dimensional test case involving quartz cementation and mechanical compaction in sandstones. The reliability of the gPCE approximation in the context of an inverse modeling framework is assessed. The effects of (a) the strategy adopted in building the gPCE and (b) the type and spatial location of calibration data (such as temperature and porosity) on the goodness of the parameter estimates are explored by means of classical estimation error analysis and model selection criteria.

1. INTRODUCTION

Diagenesis of sedimentary basins involves the coupled action of mechanical and geochemical processes [*Wangen,* 2010]. Mechanical compaction is due to the effective stresses caused by the load of the overlying sediments after deposition. These effective stresses induce grain rearrangement and therefore porosity reduction with increasing depth. Geochemical compaction has also a large influence on the evolution of the porous matrix structure. Typical examples include quartz cementation in sandstones and smectite to illite transformation in shales [see, e.g., *Osborne and Swarbrick*, 1999; *Milliken,* 2004; *Taylor et al.*, 2010 and references therein]. In this work we focus on quartz cementation phenomena, which are particularly relevant in sandstones. These processes take place at the pore scale and are typically temperature-activated.

Basin evolution models require the solution of temperature and pressure fields. Knowledge of these quantities is crucial in several applications, e.g., quantitative assessment of saline groundwater flow and residence times in coastal reservoirs [*Kreitler,* 1989], prediction of liquid overpressure [e.g., *Jiao and Zheng*, 1998], evaluation of hydrocarbon generation and migration [e.g., *Taylor et al.,* 2010; *Zhao and Lerche*, 1993], analysis of risk assessment in drilling practice [*Nadeau,* 2011; *O'Connor et al.*, 2011]. The characteristic spatial and temporal evolutionary scales of sedimentary basin compaction processes are, respectively, of the order of kilometers and millions of years. On the other hand, the critical physical and chemical processes take place at the pore scale and are typically analyzed through laboratory experiments. A complete and rigorous model formulation which embeds the multiscale nature of the diagenetic processes is still not available. Therefore, simplified effective models are usually adopted. Empirical relationships between porosity and stresses [e.g., *Schneider et al.*, 1994] are commonly employed. Nonetheless, issues related to quartz cementation, including the role played by pressure and hydrocarbons in the precipitation/dissolution process as well as the proper identification of the source of silica, have been largely debated in the literature [*Taylor et al.*, 2010]. Although inhibition of quartz cementation due to fluid overpressure has been observed [e.g., *Osborne and Swarbrick*, 1999], widely used quartz cementation models rely on the assumption that (*i*) quartz precipitation is a temperature-driven reaction-limiting factor [e.g., *Oelkers et al.*, 1996] and (*ii*) dissolution of grains and quartz precipitation happen at the same location, meaning that the source of quartz is local [e.g., *Walderhaug* 1994, 1996; *Lander and Walderhaug*, 1999].

Outputs of basin compaction models are affected by uncertainty, mainly due to the lack of knowledge of the appropriate interpretive conceptual and mathematical model and the associated parameters. Since direct measurements of model parameters are typically scarce, parameter estimation can be performed by conditioning a given compaction model on measured state variables, such as temperature, heat flux, porosity and pressure [*Lerche*, 1991; *Zhao and Lerche*, 1993; *Tuncay and Ortoleva*, 2004; *Beha et al.*, 2008; *Huvaz et al.*, 2005].

Recently, *Formaggia et al.* [2013] presented a comprehensive simulation tool for sandstone compaction in the presence of quartz cementation. This model allows to (a) perform a global sensitivity analysis of the system states under uncertain mechanical and geochemical parameters and (b) obtain an efficient surrogate model of the compaction system. The surrogate model is based on a sparse grid sampling technique in the context of a generalized polynomial chaos expansion (gPCE) approximation of the system states [*Ghanem and Spanos*, 1991; *Xiu and Karniadakis*, 2002; *Le Maitre and Knio*, 2010]. Being a polynomial expression, the gPCE approximation of the model outputs can be evaluated at any location in space and time and for any combination of values of the uncertain parameters at a reduced computational cost. This allows obtaining a fast evaluation of the mean and the variance of the system states associated with the randomness of the model parameters, as well as of the Sobol sensitivity indices [*Sobol*, 1991; *Sudret,* 2007; *Crestaux et al*, 2009] which provide a direct quantitative measure of the influence of each uncertain parameter on the total output variance. The information embedded in the Sobol indices can be used in the context of an inverse modeling procedure to derive optimal calibration data locations [see, e.g., *Fajraoui et al.*, 2011, 2012; *Ciriello et al.*, 2013]. Probability density functions of output variables can also be computed to evaluate uncertainty propagation features through the model.

The idea of accelerating the solution of inverse problems through the use of polynomial approximations has been already discussed in literature [e.g. *Balakrishnan et al,* 2003; *Marzouk et al.*, 2007, 2009; *Fajraoui et al.*, 2011, 2012; *Ciriello et al.*, 2013; *Oladyshkin et al.*, 2013]. In this work, we analyze the feasibility of estimating the key parameters of a basin compaction model within an inverse maximum likelihood (ML) framework [e.g., *Carrera and Neuman*, 1986] where the full model is replaced by its gPCE approximation. A preliminary attempt to accelerate ML estimates with a gPCE methodology was presented by *Pence et al.* [2011] in the context of dynamical systems.

Here we employ a synthetic example to explore the influence of the joint information given by heterogeneous and uncertain state variable measurements (e.g., porosity and temperature), on our ability to properly estimate the key parameters of a basin compaction model. Recent studies [e.g. *Zhang et al.*, 2010; *Lin and Tartakovsky*, 2009] show that reduced models based on gPCE may result in inaccurate results in the presence of high nonlinearity. The distinctive feature of this work is the use of gPCE within inverse modeling for *i*) highly nonlinear coupled equations system and (*ii*) large space-time evolutionary scales typical of basin compaction models. The relevance of the spatial location of data on the quality of parameter estimation is also assessed. We highlight the parameters playing a critical role in the model through the use of the Sobol indices. An additional novel element of our study is the analysis of the way the ML framework can benefit from the adoption of anisotropic polynomial approximations, in which the surrogate model is refined only with respect to the key parameters. Here we use an a-priori anisotropic approximation strategy, where the importance of each parameter is established in advance, through human expertise or *ad-hoc* preliminary computations. The sparse grid sampling points/gPCE polynomials are then choosen accordingly, following the approach presented in [*Nobile et al.*, 2008; *Bäck et al*, 2011]. We mention that on the other hand an a-posteriori anisotropic approximation strategy could also be possible, i.e. a strategy in which the importance of each parameter is discovered during the computation, as points / polynomials get added to the approximation [e.g., *Gerstner and Greipel*, 2003; *Chkifa et al.*, 2013]. Such anisotropic strategies have been extensively discussed and applied e.g. to diffusion and groundwater flow problems [see e.g., *Beck et al.*, 2012; *Foo et al.*, 2008; *Ganapathysubramanian and Zabaras*, 2007]. To the best of authors' knowledge, the present paper is the first one using anisotropic approximation strategies in a model inversion approach.

The paper is organized as follows. In Section 2 we recall the main features of the basin compaction model and of its gPCE approximation. Section 3 is devoted to the description of the ML inverse framework and of the numerical methodology adopted. Numerical results concerning a synthetic test case are discussed in Section 4. Concluding remarks are then presented.

2. BASIN COMPACTION MODELING

In this section we briefly summarize the theoretical and numerical tools developed by *Formaggia et al.* [2013] for the analysis of mechanical and geochemical compaction in a basin-scale model. We introduce the mathematical formulation of the sandstone compaction model and then recall the numerical methodologies employed to derive the gPCE-based reduced model.

2.1 Forward basin compaction model

Consider a one-dimensional sedimentary basin $\Omega(t) = [z_{bot}(t), z_{top}(t)]$ evolving with time t, $z_{bot}(t)$ and $z_{top}(t)$ being the bottom and the top of the domain, respectively. Mass conservation of fluid and solid phases in $\Omega(t)$ are governed respectively by

$$\frac{\partial \phi \rho^l}{\partial t} + \frac{\partial \phi \rho^l u^l}{\partial z} = q^l \tag{1}$$

$$\frac{\partial \left[\left(1-\phi\right)\rho^s\right]}{\partial t} + \frac{\partial \left[\left(1-\phi\right)\rho^s u^s\right]}{\partial z} = q^s \tag{2}$$

where ϕ is the porosity of the sediments, u^i and ρ^i indicate the velocity and the density of i-phase (with $i = s$ for the solid phase and $i = l$ for the fluid phase) respectively. The source terms q^i account for processes associated with fluid ($i = l$, e.g., water released during transformation of clay mineral) and solid ($i = s$, e.g., quartz precipitation) generation.

The Darcy flux (u^D) is given by

$$u^D = \phi\left(u^l - u^s\right) = \frac{K}{\mu^l}\left(\frac{\partial p}{\partial z} - \rho^l g\right) \tag{3}$$

where p is the pore pressure, μ^l is the fluid dynamic viscosity, g is the gravity acceleration and K is the permeability. The latter is modeled as $K(\phi) = 10^{k_1\phi - k_2}$ [*Wangen*, 2010] where k_1 and k_2 are fitting parameters which are usually determined through laboratory experiments.

The rate of porosity change due to mechanical compaction is given by

$$\frac{d\phi_M}{dt} = -\beta\left(\phi_0 - \phi_f\right)\exp\left(-\beta\sigma_C\right)\frac{d\sigma_C}{dt} \tag{4}$$

where

$$\frac{d\,\cdot}{dt} = \frac{\partial\,\cdot}{\partial t} + u^s\frac{\partial\,\cdot}{\partial z}, \tag{5}$$

Here, ϕ_0 is the initial porosity of the basin, ϕ_f is the minimum porosity value that can be attained by pure mechanical compaction, β is the soil compressibility coefficient and σ_C is the effective stress, given by subtracting the liquid pressure from the total load.

Quartz precipitation is modeled as proposed by *Walderhaug* [1996]

$$\frac{d\phi_Q}{dt} = A\frac{M_Q}{\rho_Q}a_q 10^{b_q T}\,; \qquad A = A_0\left(\frac{\phi}{\phi_{act}}\right); \;\; T > T_C \tag{6}$$

where ϕ_Q is the volumetric fraction of quartz cement, M_Q and ρ_Q are respectively the molar mass and the density of quartz, A_0 and ϕ_{act} represent the specific surface and the actual porosity at the onset of quartz precipitation, and a_q and b_q are characteristic parameters of the system. The reaction takes place only if the temperature, T, is larger than a critical value T_C (usually assumed equal to 80°C).

Finally, the temperature evolution is modeled by

$$C_T\frac{\partial T}{\partial t} + C_T\frac{\partial T}{\partial z} - \frac{\partial}{\partial z}\left(K_T\frac{\partial T}{\partial z}\right) = 0; \qquad K_T = \lambda_l^{\phi}\left[\lambda_s\right]^{1-\phi} \tag{7}$$

where $C_T(\phi) = \phi\rho^l c^l u^l + \left(1-\phi\right)\rho^s c^s u^s$ is the effective thermal capacity of the medium, K_T is the thermal conductivity, λ_s and λ_l are fluid and solid specific conductivities, c^l and c^s respectively are the liquid and solid specific thermal capacities. The nonlinear partial differential system (1)-(7) is complemented by appropriate initial and boundary conditions as detailed in Section 4.

2.2 Global sensitivity analysis and model reduction technique

Inverse modeling (or history matching) typically requires solving the forward system model for several values of the unknown parameters. This procedure depends on the methodology employed and is usually highly time consuming. In the following we alleviate the computational burden by introducing a polynomial surrogate of the full compaction model described in Section 2.1

We collect the N_P uncertain parameters, p_i, in vector $\mathbf{p} \in \Re^{N_P}$. Since, in general, no detailed information on geochemical compaction model parameters are available, each p_i is assumed to be described by a uniform distribution within the interval $\Gamma_i = [a_i, b_i]$, so that $\mathbf{p} \in \Gamma = \Gamma_1 \times \Gamma_2 \times ... \times \Gamma_{N_P}$. Any output of the full compaction model can thus be described as a function $f(\mathbf{p}) : \Gamma \rightarrow \Re$. The generalized Polynomial Chaos expansion (gPCE) allows approximating $f(\mathbf{p})$ by a sum of Q multivariate orthogonal polynomials $\psi_i(\mathbf{p})$

$$f(\mathbf{p}) \approx \sum_{i=1}^{Q} \alpha_i \psi_i(\mathbf{p}) \tag{8}$$

where α_i are real numbers called gPCE coefficients. The specific family of polynomials to be used in (8) depends on the probability distribution of the parameters. Since p_i are considered as uniformly distributed, in the following we adopt the family of multivariate Legendre polynomials [*Ghanem and Spanos*, 1991; *Xiu and Karniadakis*, 2001; *Le Maitre and Knio*, 2010]. The cornerstone of the adopted algorithm is the so-called sparse-grid sampling of Γ [*Smolyak*, 1963; *Xiu and Hestaven*, 2005; *Babuska et al*, 2007; *Formaggia et al.*, 2013], which is a generalization of the simpler Cartesian grid sampling. In the latter, one first chooses a set of points within each interval Γ_i and then builds a grid over Γ by taking the Cartesian product of such sets. The procedure is exemplified in Figure 1a. Clearly, the number of points of the resulting grid increases exponentially with the dimension of Γ, i.e., with N_P. On the other hand, the sparse grid procedure allows improving the effectiveness of sampling upon creating a grid over Γ by superposing many small Cartesian grids Γ (see Figure 1b and 2). In other words, this method is able to capture the features of the sampled function $f(\mathbf{p})$ by using a relatively small number of points, as can be seen by comparing Figures 1a and 1b. The results obtained with the full model at each point of the sparse grid are then used to build an intermediate surrogate model of $f(\mathbf{p})$ which is termed sparse grid approximation of $f(\mathbf{p})$ and which is then converted in the gPCE expansion with simple algebraic manipulations [see *Formaggia et al.*, 2013 for details].

For a given Γ, the number of collocation points of the sparse grid, N_C, and their distribution over the parameter space is determined according to the following steps [see, e.g., *Bäck et al.*, 2011]:

(a) selection of the set of polynomials that will enter the gPCE expansion (8); common examples of polynomial sets are Legendre polynomials whose maximum degree in each direction $p_1, p_2, \ldots, p_{N_p}$ does not exceed a given level $w \in N$ (which constitutes the “maximum degree” gPCE) or the set of Legendre polynomials for which the sum of degrees in each direction does not exceed a given level (which is termed “total degree” gPCE);

(b) introduction of anisotropic refinements of the gPCE approximation of $f(\mathbf{p})$; in this step it is possible to refine the gPCE model only with respect to the most relevant parameters by adopting anisotropic grids as shown, e.g., in Figure 1c.

We will follow these steps in the numerical examples described in Section 4, where we demonstrate that the gPCE can efficiently be employed as a surrogate model for the inversion procedure of a basin compaction model. The gPCE expansion also allows, after simple algebraic manipulations of the coefficients α_i, to compute the mean and the variance of $f(\mathbf{p})$ together with the Sobol indices. The latter provide a measure of the relative contribution of each parameter to the total variance of the state variables and can be used to perform a global sensitivity analysis of the system output with respect to the input parameters [*Archer et al.*, 1997; *Sudret* 2007; *Crestaux et al*, 2009]. In particular, total Sobol indices include the sum of all Sobol indices related to a single parameter and can be employed to assess the global influence of any given parameter on the uncertainty related to the model output.

3. INVERSE MODELING

In this Section, we describe the Maximum Likelihood (ML) approach that we adopt to derive ML estimates $\hat{\mathbf{p}}$ of $\mathbf{p}$ on the basis of porosity and/or temperature measurements. We set

$$\phi_i^* = \phi_i + \varepsilon_{\phi_i}^* \quad i = 1, ..., N_\phi \tag{9}$$

$$T_j^* = T_j + \varepsilon_{T_j}^* \quad j = 1, ..., N_T \tag{10}$$

where ϕ_i and T_j are, respectively, the (unknown) true values of ϕ and T at measurement points z_i and z_j at time t, ϕ_i^* and T_j^* are their (known) measured values affected by zero-mean (unknown) measurement errors, $\varepsilon_{\phi_i}^*$ and $\varepsilon_{T_j}^*$. In practical applications of basin-scale problems the time t at which measurements are taken usually coincides with the current time [e.g., *Zhao and Lerche*, 1993; *Taylor et al.*, 2010]. Following the work of *Carrera and Neuman* [1986] we assume (*a*) $\varepsilon_{\phi_i}^*$ and $\varepsilon_{T_j}^*$ to be multivariate Gaussian, (*b*) absence of spatial correlation and crosscorrelation between measurement errors of ϕ and T, and (*c*) covariance matrix of measurements errors of ϕ, $\mathbf{C}_\phi$, and *T*, $\mathbf{C}_T$, to be known up to positive statistical parameters, σ_ϕ^2, and σ_T^2, i.e.,

$$\mathbf{C}_\phi = \sigma_\phi^2 \mathbf{V}_\phi \qquad\qquad \mathbf{C}_T = \sigma_T^2 \mathbf{V}_T \tag{11}$$

where σ_ϕ^2, and σ_T^2 are typically unknown and estimated during inversion. According to assumption (*b*) $\mathbf{V}_\phi$ and $\mathbf{V}_T$ become diagonal matrices. Furthermore, in the following we assume $\mathbf{V}_\phi = \mathbf{V}_T = \mathbf{I}$, i.e., the prior estimation errors of ϕ and T are constant in space.

The ML estimate $\hat{\mathbf{p}}$ of $\mathbf{p}$ is obtained by minimizing the negative log likelihood (*NLL*) criterion [*Carrera and Neuman*, 1986; *Medina and Carrera*, 2003] that, in the absence of direct measurements of **p**, becomes

$$NLL = \frac{J_\phi}{\sigma_\phi^2} + \frac{J_T}{\sigma_T^2} + N_\phi \ln \sigma_\phi^2 + N_T \ln \sigma_T^2 + N_D \ln(2\pi) \tag{12}$$

where $N_D = N_\phi + N_T$. The quantities J_ϕ and J_T are, respectively, the porosity and the temperature residual criteria and are defined as

$$J_\phi = \left(\mathbf{\Phi} - \mathbf{\Phi}^*\right)^{\mathrm{T}} \mathbf{V}_\phi^{-1} \left(\mathbf{\Phi} - \mathbf{\Phi}^*\right); \tag{13}$$

$$J_T = \left(\mathbf{T} - \mathbf{T}^*\right)^{\mathrm{T}} \mathbf{V}_\phi^{-1} \left(\mathbf{T} - \mathbf{T}^*\right) \tag{14}$$

where superscript $^{\mathrm{T}}$ denotes transpose, $\mathbf{\Phi}^*$ is the vector of porosity measurements, $\mathbf{T}^*$ is the vector of temperature measurements, $\mathbf{\Phi}$ and $\mathbf{T}$ are the vectors of conditional porosity and temperature values evaluated according to the forward model (1)-(7) at measurement locations. Note that $\mathbf{\Phi}$ and $\mathbf{T}$ depend on the parameter vector $\mathbf{p}$. It is thus clear that the minimization of *NLL* requires the solution of the system (1)-(7) for a (typically large) number of $\mathbf{p}$ values. This task can be extremely CPU time consuming, especially in the presence of strong model nonlinearities. Therefore, in this work we explore the feasibility of replacing (1)-(7) by the gPCE approximation of $\mathbf{\Phi}$ and $\mathbf{T}$, which can be efficiently evaluated for any particular value of $\mathbf{p}$. For notational convenience, in the following we use $\mathbf{\Phi}$ and $\mathbf{T}$ to refer to the gPCE solution. Therefore, σ_ϕ^2 and σ_T^2 include both measurement and model errors, the latter being due to the use of the gPCE approximation. Note that if σ_ϕ^2 and σ_T^2 are known, minimization of (12) is equivalent to minimizing the general least squares criterion

$$J = J_\phi + \lambda J_T \qquad (15)$$

where $\lambda = \sigma_\phi^2 / \sigma_T^2$. Small values of λ imply that porosity data are assumed to be more reliable than temperature data, and hence the minimization of J will be essentially equivalent to the minimization of J_ϕ; the opposite holds for large values of λ.

In general, σ_ϕ^2 and σ_T^2 (and therefore λ) are unknown *a priori*. In principle, these statistical quantities could be estimated jointly with $\mathbf{p}$ by minimizing (12). However, such estimation is likely to be unstable [*Carrera and Neuman*, 1986]. *Carrera and Neuman* [1986] suggested to avoid this problem by (*i*) minimizing (15) with respect to $\mathbf{p}$ for a set of λ values, (*ii*) obtaining for each λ the corresponding ML estimates of σ_ϕ^2 and σ_T^2 as

$$\hat{\sigma}_\phi^2 = \frac{J_{\min}}{N_D} \qquad \hat{\sigma}_T^2 = \frac{\hat{\sigma}_\phi^2}{\lambda} \qquad (16)$$

where $J_{\min}$ is the minimum value of J evaluated at step (*i*), (*iii*) evaluating *NLL* by (12) for each set of ($\mathbf{p}$, $\hat{\sigma}_\phi^2$, $\hat{\sigma}_T^2$), and (*iv*) choosing the set ($\mathbf{p}$, $\hat{\sigma}_\phi^2$, $\hat{\sigma}_T^2$) for which *NLL* is minimum. *Riva et al.* [2009, 2011] have demonstrated that an improved estimate of λ can be obtained on the basis of the Bayesian criterion [*Kashyap*, 1982]

$$KIC = NLL - N_P \ln 2\pi - \ln\left|\mathbf{Q}\right| \qquad (17)$$

where $|\mathbf{Q}|$ is the Cramer-Rao lower bound approximation of the determinant of the covariance matrix of the estimation error, i.e.,

$$\mathbf{Q} = \sigma_\phi^2 \left(\mathbf{J}_\phi{}^T \mathbf{V}_\phi^{-1} \mathbf{J}_\phi + \lambda \mathbf{J}_T{}^T \mathbf{V}_T^{-1} \mathbf{J}_T \right)^{-1} \tag{18}$$

where $\mathbf{J}_k$ ($k = \phi$, T) is the Jacobian matrix including the derivatives of the output state variables (porosity or temperature) with respect to the model parameters evaluated at measurement locations at the values of $\mathbf{p}$ at the current iteration of the inversion procedure. Note that evaluation of $\mathbf{J}_k$ usually requires to solve several times the forward model to approximate the derivatives of the state variables with respect to the model parameters. A key point of the gPCE framework is that $\mathbf{J}_k$ can be obtained analytically, as ϕ and T are approximated by polynomial functions. For an extensive discussion of the reliability of *KIC* and *NLL* in driving the choice of λ see, e.g., *Ye et al* [2008], *Tsai and Li* [2008], *Riva et al* [2011].

In summary, we propose here to obtain ML estimates of the parameters characterizing a basin-scale system subject to mechanical and geochemical compaction according to the following steps:

1. Derivation of the gPCE surrogate model. N_P uncertain model parameters are required to be selected. This step is developed upon sampling the parameters space Γ with a sparse grid and solving the compaction problem (1)-(7) at each point of the sparse grid. Numerical evaluation of (1)-(7) is performed according to the lagrangian approach proposed by *Formaggia et al.* [2013].
2. Minimization of J for selected λ values. The minimization of J is performed through the Nelder-Mead simplex search method [*Lagarias et al.*, 1998]. During this step, the gPCE is evaluated for each space-time location where measurements are available and for each

tentative value $\hat{\mathbf{p}}$ computed by the minimization algorithm. We repeat the minimization procedure with different initial parameters guesses, to avoid detecting local minima.

3. ML estimation of σ_{ϕ}^{2} and σ_{T}^{2} by (16) for each λ.
4. Selection of λ by minimizing (with respect to λ) *NLL* and/or *KIC*.

4. ILLUSTRATIVE EXAMPLE

4.1 Problem Definition and Global Sensitivity Analysis

We illustrate the proposed methodology on a synthetic basin compaction test case similar to the one analyzed by *Formaggia et al.* [2013]. The total sedimentation time we consider is 200 Ma (millions of years) and the sedimentation rate is fixed to 30 m/Ma. Temperature and pressure values at the top of the basin ($z = z_{top}$) are assigned, and are respectively equal to 295 K and $\gamma_{\text{sea}} h_{\text{sea}}$ (i.e., the hydrostatic pressure of the overlying sea depth, h_{sea}, γ_{sea} being the specific weight of seawater). For our purposes we assume h_{sea} to be constant in time, thus disregarding possible erosion phenomena. The bottom of the basin is assumed to be impermeable ($u^D = 0$) and subject to a given a geothermal gradient, G_T .

Amongst the several parameters characterizing the system (1)-(7), *Formaggia et al.* [2013] showed that uncertainty typically associated with the three parameters β, a_q and h_{sea} bears the largest influence on porosity profiles, while temperature is mostly affected by a_q and h_{sea}. In this study we also consider uncertainty in the geothermal gradient G_T, which is expected to influence both temperature and porosity distributions.

All these uncertain parameters are assumed to be uniformly distributed within the intervals (*Min*, *Max*) reported in Table 1. Selected bounds are consistent with the results of previous sensitivity analysis [*Walderhaug*, 1994; *Lander and Walderhaug*, 1999; *Wangen*, 2010;

Formaggia et al., 2013]. Relative ranges of parameters are computed as $\left|\min\left(p_i\right)-\max\left(p_i\right)\right|/\overline{p}_i$, i.e., as the relative variation of the interval with respect to its mean value. A large relative range is considered for a_q and it is linked to the high level of uncertainty associated with the estimation of reaction kinetics parameters. The remaining parameters are assumed constant. In particular, we set: $b_q = 0.022$ °C^{-1} [*Walderhaug*, 1994], and $k_2 = 16.94$ [*Wangen*, 2010; *Formaggia et al.*, 2013]. At the initial simulation time, we assume a layer of material of 500 m thickness. Initial porosity distributions is assigned through standard Athy's law [e.g., *Schneider*, 1994] for mechanical compaction.

Figure 3 depicts the vertical profiles of the average temperature and porosity obtained by the gPCE approximation at the final deposition time. Here, we have used a "total degree" gPCE (see Section 2) at level $w = 3$, which is adequate to resolve the complexity of the input/output mapping (see also Section 4.2 for a further discussion). Figure 3 also reports the uncertainty envelopes obtained by summing and subtracting the associated standard deviation to the mean profiles. Figure 4 shows the Sobol indices associated with the results plotted in Figure 3. The mean porosity (Figure 3a) initially (z > - 2000 m) decreases with decreasing z following an exponential trend, as described by (4). This behavior is due to mechanical compaction and is strongly influenced by β and h_{sea}, as shown in Figure 4a. Quartz cementation starts at about z ≈ - 2000 m where the Sobol indices related to a_q and G_T increase. For z < - 2000 m the porosity rapidly decreases to zero and its variance tends to increase. Mean temperature (Figure 3b) increases almost linearly with depth until z ≈ - 2000 m. It can be noted that the temperature gradient decreases when quartz cementation starts to become relevant. This behavior is associated with the decrease of accessible pore space, which influences the thermal conductivity of the medium at large depth values.

Figure 4b reveals that the temperature distribution is highly influenced by G_T and h_{sea}, as these parameters are strictly related to the boundary conditions of the thermal problem. Parameter a_q plays also a role at high depths, highlighting the strong correlation between the vertical distributions of temperature and porosity when the quartz precipitation is active.

4.2 Inversion modeling and results

The reference porosity $\mathbf{\Phi}_{true}$ and temperature $\mathbf{T}_{true}$ fields have been generated by solving (1)-(7) with $\mathbf{p} = \mathbf{p}_{true}$ (see Table 1). The profiles $\mathbf{\Phi}_{true}$ and $\mathbf{T}_{true}$ obtained at the final simulation time (t = 200 Ma) are shown in Figure 3. We sample $\mathbf{\Phi}_{true}$ and $\mathbf{T}_{true}$ at 300 equally spaced locations along the z-axis to obtain the information employed in the inversion procedure. In order to simulate measurements errors, the calibration data $\mathbf{\Phi}^*$ and $\mathbf{T}^*$ are obtained by superimposing a white Gaussian noise having a variance of σ_ϕ^2 and σ_T^2 to $\mathbf{\Phi}_{true}$ and $\mathbf{T}_{true}$, respectively. In the following, we investigate the impact of (*i*) the order w of the gPCE approximation, (*ii*) the type of calibration data available, and (*iii*) the spatial distribution of the data on the quality of **p** estimates.

4.2.1 Analysis of the gPCE approximation in the inversion procedure

We start by assuming that only porosity data are available (i.e., λ = 0 in (15)) and compare the outputs of the inversion procedure obtained with various orders w of the gPCE polynomial approximation (8) of the porosity. We set $\mathbf{\Phi}^* = \mathbf{\Phi}_{true}$. Therefore, the only source of error in the calibration data is due to the gPCE approximation of porosity. Here, we investigate two different gPCE strategies, namely (a) an isotropic sampling strategy, according to which the same accuracy is adopted to approximate the dependence of the porosity on each parameter; and (b) an anisotropic strategy, in which we consider different accuracies of the approximation with respect to each parameter. With reference to the isotropic setting, we use the total degree gPCE at two levels $w = 3,4$. In the context of the anisotropic setting, we set a polynomial order 3 for each parameter with the exception of a_q, for which we use polynomials up to order 6. This choice leads to a sparse grid sampling that concentrates sampling points along the a_q direction in the parameters space Γ (see Figure 1c for an example in a two-dimensional parameter space). This choice is motivated by the observation that the dependence of ϕ on a_q has a complex behavior, due to the fact that a_q appears in the exponential quartz cementation rate (6), which results in a highly nonlinear input-output relationship, and therefore requires a special refinement [*Formaggia et al.*, 2013]. Moreover, as noted above (see also Table 1), the relative range of a_q is much larger than that associated with the other uncertain parameters. In general, in the presence of real data, sparse grids refinement may be selected on the basis of information provided by a global sensitivity analysis (e.g., Sobol indices), estimates of parameter uncertainty and/or large relative ranges.

The key results obtained are listed in Table 2. The lowest values of *NLL*, *KIC* and σ_ϕ are obtained with the anisotropic gPCE approximation, thus identifying the latter as the best forward (surrogate) model. The relative errors $\eta\left(p_i\right) = \left|\hat{p}_i - p_i^*\right| / p_i^*$ (with $i = \beta, a_q, G_T, h_{sea}$) associated with each estimated parameter are also reported in Table 2. As expected, η_i decreases with w by adopting an isotropic grid in the parameter space. In particular, the parameter a_q is poorly estimated when $w = 3$ and an isotropic sampling is performed. The anisotropic refinement of the sparse grid provides relative errors which are always smaller that 1%. Remarkably, adoption of the anisotropic refinement leads to improved results with respect to those obtained through an isotropic gPCE with w = 4. This degree of accuracy is also associated with a considerably reduction of the CPU time, i.e., of about 60% when compared against the isotropic gPCE with w = 4 (see also the number N_C of sparse grid points required in the two cases and reported in Table 2). For validation purposes, the model inversion has also been performed through the use of a standard genetic algorithm [*Storn and Price*, 1997] by relying on the true forward model (1)-(7) (details not shown). The outcome of this analysis are practically coincident with the anisotropic gPCE-based solutions while the CPU time increases by one order of magnitude, thus corroborating the usefulness of the approach we propose for inversion purposes.

Similar results have been obtained by considering that only temperature data are available, as shown in Table 3. Comparing the results listed in Tables 2 and 3 we note that the use of temperature data leads to slightly improved accuracy in the estimates of G_T, h_{sea} and a_q with respect to what can be obtained using only porosity data in the setting we analyzed (compare the relative errors η in Table 3 and Table 2). On the other hand, relying on porosity data allows obtaining an improved estimate of β. This result is consistent with the Sobol indices analysis reported in Figure 4, where it is clear that β does not influence significantly the temperature distribution.

4.2.2 Choice of calibration dataset

In this paragraph we consider the joint use of porosity and temperature in the inversion procedure. We set the measurement error standard deviations to $\sigma_\phi = 3.00 \times 10^{-2}$ and $\sigma_T = 10.00$ K, corresponding to coefficients of variations of the order of 10% of the interval comprised between minimum and maximum values displayed by the two variables (see Figure

3). Based on the analysis reported in Section 4.2.1, we resort to the anisotropic refinement of the sparse grid for the gPCE. As described in Section 2, we perform different inversions for selected λ values. Figure 5 depicts the way *NLL* and *KIC* vary with λ. We note that the two curves are approximately flat within the interval $6\times10^{-6} < \lambda < 1\times10^{-5}$, with a minimum located at $\lambda = 7\times10^{-6}$. This result is consistent with the true reference value of λ, which is given by $\sigma_\phi^2 / \sigma_T^2 = 9\times10^{-6}$. The ML estimates of the standard deviation of the porosity and temperature measurement error evaluated by (16) are $\hat{\sigma}_\phi = 2.65\times10^{-2}$ and $\hat{\sigma}_T = 10.01$ K and are indeed very close to the true values.

Figure 6 reports the ratio between the ML estimate, $\hat{p}_i$, of each parameter and the true values obtained using (*i*) only $\mathbf{\Phi}^*$ data (Figure 6a), (*ii*) only $\mathbf{T}^*$ data (Figure 6b), and (*iii*) $\mathbf{\Phi}^*$ and $\mathbf{T}^*$ data jointly upon setting $\lambda = 7\times10^{-6}$ (Figure 6c). Figure 6 also reports the uncertainty bands of width $\pm\hat{\sigma}_{p_i} / p_{true}$, where $\hat{\sigma}_{p_i}$ is evaluated by (18).

When only porosity data are used, values of $\hat{p}_i / p_{true,i}$ are comprised in the interval of width $\pm\hat{\sigma}_{p_i} / p_{true}$ around the corresponding estimated value (Figure 6a). Moreover we observe that $0.8 < \hat{p}_i / p_{true,i} < 1.2$, i.e., the relative errors $\eta(p_i)$ are always smaller than 20%. True values of the parameters lie within the range of width $\pm\hat{\sigma}_{p_i}$ around the mean value for G_T, h_{sea} and a_q while the mechanical compaction parameter β is underestimated. Uncertainty related to the estimate of a_q is considerably larger than that associated with the other three parameters.

Using only temperature data $\mathbf{T}^*$ leads to overestimating all parameters, as shown in Figure 6b. In this case, significant prediction errors are observed for β and a_q. This is consistent with the vertical distribution of the Sobol indices (Figure 4b), which shows that β and a_q influence only marginally temperature, as compared to G_T and h_{sea}.

When porosity and temperature measurements are jointly considered (Figure 6c) the parameter estimates are close to their true counterparts and their estimation uncertainty is considerably reduced. These results suggest that the characterization of a basin subject to mechanical and geochemical compaction greatly benefits by the joint availability of porosity and temperature data.

Results of Figure 6 are complemented by Table 4, where we list estimates $\hat{\sigma}_{\phi}$ and $\hat{\sigma}_T$ obtained through the different calibration procedures, together with the associated CPU time. Standard deviations of measurement errors are accurately estimated (within 10% of error with respect to their true values).

4.2.3 Analysis of the influence of the spatial distribution of available data

Finally, we assess the influence of the spatial location of available calibration data on the accuracy and efficiency of the inversion procedure. As previously discussed, the influence of the selected uncertain parameter on the output variables can be quantified through the Sobol indices. Here, we show how the knowledge of the Sobol indices enables one to identify which parameters can be accurately estimated when data are available in specific zones of the domain. In particular, we consider the following two zones of width equal to 1000 m that, according to Figure 4, allow separating the effects of different groups of uncertain parameters: (*i*) an upper zone, for -500 m $\leq z \leq -1500$ m where no quartz cementation is observed, and (*ii*) a deep zone, for -2500 m $\leq z \leq -3500$ m. In the upper interval, porosity and temperature respectively depend on (β, h_{sea}) and (G_T, h_{sea}). On the other hand when -3500 m $< z < -2500$ m both porosity and temperature are chiefly influenced by a_q and G_T (see Figure 4).

We consider the following six scenarios, depending on the location and type of available data: (1) only porosity data, $\mathbf{\Phi}_{up}^{*}$, are available within the interval –500 m < z < –1500 m; (2) only temperature data $\mathbf{T}_{up}^{*}$ are available within the interval –500 m < z < –1500 m; (3) only porosity data, $\mathbf{\Phi}_{lo}^{*}$, are available within the interval –2500 m < z < –3500 m; (4) only temperature data, $\mathbf{T}_{lo}^{*}$, are available within the interval –500 m < z < -1500 m; (5) porosity and temperature data, $\left(\mathbf{\Phi}_{up}^{*}, \mathbf{T}_{up}^{*}\right)$, are jointly available within the interval –500 m < z < –1500 m; (6) porosity and temperature data, $\left(\mathbf{\Phi}_{lo}^{*}, \mathbf{T}_{lo}^{*}\right)$, are jointly available within the interval –2500 m < z < –3500 m. In test cases 5 and 6 we set $\lambda = 7 \times 10^{-6}$, according to the results obtained in Section 4.2.2.

Figure 7a reports the ratio between ML estimated parameters and their true values for test cases 1 and 3, where only porosity data are available. Uncertainty bands of width $\pm\hat{\sigma}_{p_i} / p_{true}$ are also reported for each parameter, with the exception of a_q and G_T in test case 1, where $\hat{\sigma}_{a_q} / a_{q,true}$= 279.74 and $\hat{\sigma}_{G_T} / G_{T,true}$ = 309.12. Calibration of β and h_{sea} through porosity data leads to acceptable results, especially when the dataset $\mathbf{\Phi}_{up}^{*}$ is considered. On the other hand, it is clear that porosity data $\mathbf{\Phi}_{up}^{*}$ are not suited to estimate a_q and G_T. This can be explained by observing that quartz precipitation is not active in the upper region of the domain and consequently porosity and temperature are not linked, i.e., G_T cannot influence porosity distribution. As expected, uncertainty associated with the estimates $\hat{a}_q$ and $\hat{G}_T$ is largely reduced when the dataset $\mathbf{\Phi}_{lo}^{*}$ is employed.

Results obtained through temperature data are shown in Figure 7b. The estimates of β are not accurate in this case. This is consistent with results of Section 4.2.2 and with the information embedded in the Sobol indices (Figure 4b), which show that β has a negligible influence on temperature. Note that calibration through $\mathbf{T}_{lo}^{*}$ yields an unphysical negative value for the estimated $\hat{\beta}$. The parameter a_q is significantly overestimated when $\mathbf{T}_{lo}^{*}$ is used. On the other hand, dataset $\mathbf{T}_{up}^{*}$ leads to a negative value of $\hat{a}_q$, which is not compatible with the physical meaning of a_q. The geothermal gradient G_T and the sea level h_{sea} are well calibrated through both $\mathbf{T}_{up}^{*}$ and $\mathbf{T}_{lo}^{*}$. This is consistent with the observation that G_T and h_{sea} highly influence temperature at any location. As expected, the accuracy in G_T is improved and uncertainty is reduced when $\mathbf{T}_{lo}^{*}$ is considered, while the opposite holds for h_{sea}. Figure 7c shows that the quality of the estimate of each parameter significantly increases when porosity and temperature data are jointly available (notice the different vertical scale axis). The only inaccurate result has been obtained for the calibration of a_q, when only data in the upper part of the basin are available. As noted above, this result is due to the fact that quartz cementation is not active at shallow locations where data $\left(\mathbf{\Phi}_{up}^{*}, \mathbf{T}_{up}^{*}\right)$ are observed.

5. CONCLUSIONS

We develop and present a methodology for model inversion of nonlinear basin-scale mechanical and geochemical compaction processes based on a reduced model of the system (gPCE) and a Maximum Likelihood (ML) approach. The gPCE of porosity and temperature distributions is derived upon relying on a (generally) anisotropic sparse grid approximation of the problem outputs in the parameter space. We illustrate the proposed technique in the context of a one-dimensional synthetic test case when compaction occurs for mechanical stress and precipitation of quartz. Our work leads to the following major conclusions.

1. Anisotropic grids can be efficiently employed to refine the sparse grid approximation and increase the accuracy of the parameter estimates.
2. Inversion performed with only porosity data renders acceptable estimates of the considered uncertain parameters. However, large uncertainty is associated with the estimate $\hat{a}_q$, which determines quartz cementation kinetics. This result is associated with (*i*) relatively large uncertainty bounds assigned to the parameter, (*ii*) the nonlinear relationship between a_q and porosity. Relying only on temperature data lead to significant overestimation of both a_q and β. This result is consistent with Sobol indices showing that a_q and β do not have a strong influence on the thermal problem.
3. When porosity and temperature measurements are jointly considered all parameter estimates are close to their true counterparts and their estimation uncertainty is considerably reduced.
4. The Sobol indices can be used to identify the parameters which can be accurately estimated when data are available in specific zones of the domain. This implies that Sobol indices can drive optimal selection of measurement locations also in the context of the type of complex nonlinear processes we consider, as previously suggested by *Fajraoui et al.* [2011, 2012] and *Ciriello et al*. [2013] for relatively simple laboratory scale transport scenarios.
5. In the upper part of the basin, porosity depends mainly on β and h_{sea}, while temperature is greatly affected by G_T and h_{sea} variations. On the other hand, both porosity and temperature are mainly influenced by a_q and G_T at the largest depths investigated. This

has a direct influence on reliability and accuracy of parameter estimates. In particular, our results suggest that proper characterization of quartz cementation kinetics requires availability of porosity and temperature data at deep locations in the basin.

Future developments of the present work involve parameter estimation in the presence of heterogeneous basins, involving low permeability inclusions giving rise to fluid overpressure. Application of the proposed methodology to field measurements is also envisioned.

ACKNOWLEDGMENT

The authors are grateful for the partial financial support from Eni SpA.

REFERENCES

Archer, G. E. B., A. Saltelli, and I.M. Sobol (1997), Sensitivity measures, ANOVA-like techniques and the use of bootstrap. J. Stat. Comput. Simul. 58(2), 99–120.

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

Bäck, J., F. Nobile, L. Tamellini, and R. Tempone (2011), Stochastic Spectral Galerkin and Collocation methods for PDEs with random coefficients: a numerical comparison, in Spectral and High Order Methods for Partial Differential Equations, eds. J. Hestaven and E. Ronquist, Lecture Notes in Computational Science and Engineering, 76, 43-62, Springer.

Balakrishnan, S., A. Roy, M. G. Ierapetritou, G. P. Flach, and P. G. Georgopoulos (2003), Uncertainty reduction and characterization for complex environmental fate and transport models: an empirical Bayesian framework incorporating the stochastic response surface method. Wat Resour Res 39(12) SBH81-SBH813.

Beck, J., F. Nobile, L. Tamellini, and R. Tempone (2012), A quasi-optimal sparse grids procedure for groundwater flows, in Spectral and High Order Methods for Partial Differential Equations, eds. M. Azaiez, H. El Fekih and J. S. Hesthaven, Lecture Notes in Computational Science and Engineering, in press, Springer. Also available as EPFL-MATHICSE technical report 46/2012.

Beha, A., R. O .Thomsen, and R. Littke (2008), A rapid method of quantifying the resolution limits of heat-flow estimates in basin models. J Petrol. Geol., 31, 167-178.

Carrera, J., and S. P. Neuman (1986), Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, Wat. Resour. Res., 22 (2),199-210.

Chkifa, A., A. Cohen, R. Devore, and C. Schwab (2013), Sparse adaptive taylor approximation algorithms for parametric and stochastic elliptic PDEs, ESAIM: Math. Model. Num., 47(1), 253–280.

Ciriello, V., A. Guadagnini, V. Di Federico, Y. Edery, and B. Berkowitz (2013), Comparative analysis of formulations for conservative transport in porous media through sensitivity-based parameter calibration, Water Resour. Res., in press.

Crestaux, T., O. Le Maıtre, and J. M. Martinez (2009), Polynomial chaos expansion for sensitivity analysis, Reliab. Eng. Syst. Safety, 94, 7, 1161–1172.

Fajraoui, N., F. Ramasomanana, A. Younes, T. Mara, P. Ackerer, and A. Guadagnini (2011), Use of global sensitivity analysis and polynomial chaos expansion for interpretation of non reactive transport experiments in laboratory scale porous media. Water Resour. Res, 47(2), W02521.

Fajraoui, N., T. A. Mara, A. Younes, and R. Bouhlila (2012), Reactive Transport Parameter Estimation and Global Sensitivity Analysis Using Sparse Polynomial Chaos Expansion, Water, Air, & Soil Pollution, 223(7), 4183-4197.

Foo, J., X. Wan. and G. E. Karniadakis (2008), The multi-element probabilistic collocation method (ME-PCM): Error analysis and applications, J. Comput. Phys., 227, 9572–9595.

Formaggia L., A. Guadagnini, I. Imperiali, V. Lever, G. M. Porta, M. Riva, A. Scotti, and L. Tamellini (2013), Global sensitivity analysis through polynomial chaos expansion of a basin-scale geochemical compaction model, Computat. Geosci., 17(1), 25-42.

Ganapathysubramanian, B., and N. Zabaras (2007), Sparse grid collocation schemes for stochastic natural convection problems, J. Comput. Phys., 225(1), 652–685.

Gerstner, T., and M. Griebel (2003), Dimension-adaptive tensor-product quadrature, Computing, 71(1), 65–87.

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

Huvaz, O., R. O. Thomsen, and S. Noeth (2004), A method for analyzing geothermal gradient histories using the statistical assessment of uncertainties in maturity models. J. Petrol. Geol., 38, 107-118.

Jiao, J. J., and C. Zheng, (1998), Abnormal fluid pressures caused by deposition and erosion of sedimentary basins. J. Hydrol., 204, 124-137.

Kashyap, R. L., (1982), Optimal choice of AR and MA parts in autoregressive moving average models. IEEE T. Pattern. Anal. 4(2):99–104.

Kreitler, C. H., (1989), Hydrogeology of sedimentary basins. J. Hydrol., 106, 29-53.

Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright (1998), Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions, SIAM J. of Optimiz., 9(1), 112-147.

Lander, R. H., and O. Walderhaug (1999), Predicting porosity through simulating sandstone compaction and quartz cementation. AAPG Bull., 83(3) 433-449.

Le Maitre, O., and O. Knio (2010), Spectral methods for uncertainty quantification, Scientific Computation, Springer.

Lerche, I., (1991), Inversion of Dynamical Indicators in Quantitative Basin Analysis Models. I. Theoretical Considerations. Math. Geol., 23, 817-832.

Lin, G., and A. M. Tartakovsky (2009), An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow and solute transport in randomly heterogeneous porous media. Adv. Water Resour., 34, 1527-1538.

Marzouk, Y. M., H. N. Najm, and L. A. Rahn (2007), Stochastic spectral methods for efficient Bayesian solution of inverse problems. J. Comput. Phys., 224 (2007), 560-586.

Marzouk, Y. M., and D. Xiu (2009), A stochastic Collocation approach to Bayesian Inference in Inverse Problems, Commun. Comput. Phys., 6(4), 826-847.

Milliken, K. L., (2004), Late diagenesis and mass transfer in sanstone-shale sequences. Treatise Geochem., 7, 115-158.

Medina, A., and J. Carrera (2003), Geostatistical inversion of coupled problems: dealing with computational burden and different types of data. J. Hydrol., 281, 251-264.

Nadeau, P. H., (2011), Lessons learned from the golden zone concept for understanding overpressure development, and drilling safety in energy exploration. Working paper, Deepwater Horizon Study Group.

Nobile, F., R. Tempone, and C. G. Webster (2008), An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal., 46, 2411–2442.

O'Connor, S. , R. Swarbrick, R. Lahann (2011), Geologically-driven pore fluid pressure models and their implications for petroleum exploration. Introduction to thematic set, Geofluids, 11, 343-348.

Oelkers, E. H., P. A. Bjorkum, and W. M. Murphy (1996), A petrographic and computational investigation of quartz cementation and porosity reduction in North Sea sandstones. Am. J. Sci., 296, 420–452.

Oladyshkin, S., H. Class, and W. Nowak (2013), 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, Computat. Geosci., in press.

Osborne, M. J., and R. E. Swarbrick (1999), Diagenesis in North Sea HPHT clastic reservoirs consequences for porosity and overpressure prediction, Mar. Petrol. Geol., 16, 337-353.

Pence, B. L., H. K. Fathy, and J. L: Stein (2011), Recursive maximum likelihood parameter estimation for state space systems using polynomial chaos theory, Automatica, 47, 2420-2424.

Riva, M., A. Guadagnini, S. P. Neuman, E. Bianchi Janetti, and B. Malama (2009), Inverse analysis of stochastic moment equations for transient flow in randomly heterogeneous media. Adv. Water Resour., 10, 271–280

Riva, M., M. Panzeri, A. Guadagnini, and S. P. Neuman (2011), Role of model selection criteria in geostatistical inverse estimation of statistical data and model parameters. Water Resour. Res., 47,W07502.

Schneider, F., M. Bouteca, and G. Vasseur (1994), Validity of the porosity/effective stress concept in sedimentary basin modeling, First Break, 12, 321-326.

Smolyak, S. (1963), Quadrature and interpolation formulas for tensor products of certain classes of functions, Dokl. Akad. Nauk. SSSR, 4 240-243.

Sobol, I. M., (1991), Global sensitivity indices for nonlinear mathematical modelsand their Monte Carlo estimates, Math. Comput. Simulat., 55(1-3), 271–280.

Storn, R., and K. Price (1997), Differential Evolution – A simple end efficient heuristic for global optimization over continuous spaces, J. Global Optim., 11, 341-359.

Sudret, B., (2007), Global sensitivity analysis using polynomial chaos expansion, Reliab. Eng. Syst. Safety, 93(7), 964-979.

Taylor, T. R., M. R. Giles, L. A. Hathon, T. N. Diggs, N. R. Braunsdorf, G. V. Birbiglia, M. G. Kittridge, C. I. Macaulay, and I. S. Espejo (2010), Sandstone diagenesis and reservoir quality prediction: Models, myths, and reality, AAPG Bull., 94(8) 1093-1132.

Tsai, F.T.C., and X. Li. (2008), Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window, Water Resour. Res., 44, W09434.

Tuncay, K., and P. Ortoleva (2004), Quantitative basin modeling: present state and future developments towards predictability, Geofluids, 4, 23-39.

Walderhaug, O., (1994), Precipitation rates for quartz cement in sandstones determined by fluid-inclusion microthermometry and temperature-history modeling, J. Sed. Resear, A64, 324–333.

Walderhaug, O., (1996), Kinetic modeling of quartz cementation and porosity loss in deeply buried sandstone reservoirs. AAPG Bull., 80, 731-745.

Wangen, M., (2010), Physical Principles of Sedimentary Basin Analysis. Cambridge University Press, Cambridge (UK).

Xiu, D., and J. Hestaven (2005), High-order collocation methods for differential equations with random inputs, SIAM J. Sci. Comput. 27, 1118-1139.

Xiu, D., and G. E. Karniadakis (2002), The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput., 24(2), 619-644

Ye, M., P. D. Meyer, and S. P. Neuman (2008), On model selection criteria in multimodel analysis. Water Resour. Res., 44, W03428.

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

Zhao, K., and I. Lerche (1993), Inversion of dynamical indicators in quantitative basin analysis models. II. synthetic tests and a case history using dynamical indicator tomography. Math. Geol., 25, 107-123.

TABLES

Parameter	*Min*	*Max*	Relative range	$\mathbf{p}_{true}$
β [Pa^{-1}]	5×10^{-8}	7×10^{-8}	0.33	5.8×10^{-8}
a_q [mol m^{-2} s^{-1}]	0.40×10^{-18}	3.56×10^{-18}	1.60	1.8×10^{-18}
G_T [°C m^{-1}]	2.70×10^{-2}	3.30×10^{-2}	0.20	3.10×10^{-2}
h_{sea} [m]	450.0	550.0	0.20	520.0

Table 1. Selected uncertain parameters, associated range of variability and relative range of variation; $\mathbf{p}_{\mathrm{true}}$ indicates the parameter values used to generate the reference porosity and temperature fields.

	Isotropic gPCE w = 3	***Isotropic gPCE w = 4***	***Anisotropic gPCE***
J	6.02×10^{-4}	2.58×10^{-4}	0.29×10^{-4}
$\hat{\sigma}_\phi$	1.34×10^{-3}	8.78×10^{-4}	2.97×10^{-4}
NLL	- 3458.60	- 3740.30	- 4465.42
KIC	-3631.98	-3876.85	-4649.95
$\eta(\beta)$	0.83%	0.22%	0.02%
$\eta(a_q)$	20.58%	3.07%	0.45%
$\eta(G_T)$	3.92%	0.74%	0.09%
$\eta(h_{sea})$	1.47%	1.04%	0.07%
N_C	137	385	153
CPU time [s]	1663	4017	2266

Table 2. Main statistics of calibration performed using only porosity data, isotropic gPCE with w = 3, 4 and anisotropic gPCE. Number of collocation points and CPU times are also listed.

	Isotropic gPCE w = 3	***Isotropic gPCE w = 4***	***Anisotropic gPCE***
J	37.20×10^{-3}	6.03×10^{-3}	0.03×10^{-3}
$\hat{\sigma}_T$ [K]	1.02×10^{-2}	4.25×10^{-3}	1.79×10^{-3}
NLL	-2084.99	-2690.99	-3271.02
KIC	-2285.38	-2855.63	-3478.80
$\eta(\beta)$	1.70%	0.10%	0.16%
$\eta(a_q)$	0.68%	0.46%	0.02%
$\eta(G_T)$	0.18%	0.02%	0.02%
$\eta(h_{sea})$	0.11%	0.03%	0.01%
N_C	137	385	153
CPU time [s]	1559	6249	1820

Table 3. Main statistics of calibration performed using only temperature data with isotropic gPCE with *w* = 3 and 4 and anisotropic gPCE. Number of collocation points and CPU times are also listed.

	$\mathbf{\Phi}^*$	$\mathbf{T}^*$	$\mathbf{\Phi}^*, \mathbf{T}^*$
$\hat{\sigma}_\phi$	0.028	-	0.026
$\hat{\sigma}_T$ [K]	-	9.820	10.01
CPU time [s]	1487	1911	3211

Table 4. Estimates of standard deviations of measurement error of porosity and temperature obtained using only porosity data , $\mathbf{\Phi}^*$, only temperature data, $\mathbf{T}^*$, and both types of data, ($\mathbf{\Phi}^*, \mathbf{T}^*$). CPU times are also listed. Corresponding parameter estimates are depicted in Figure 6.

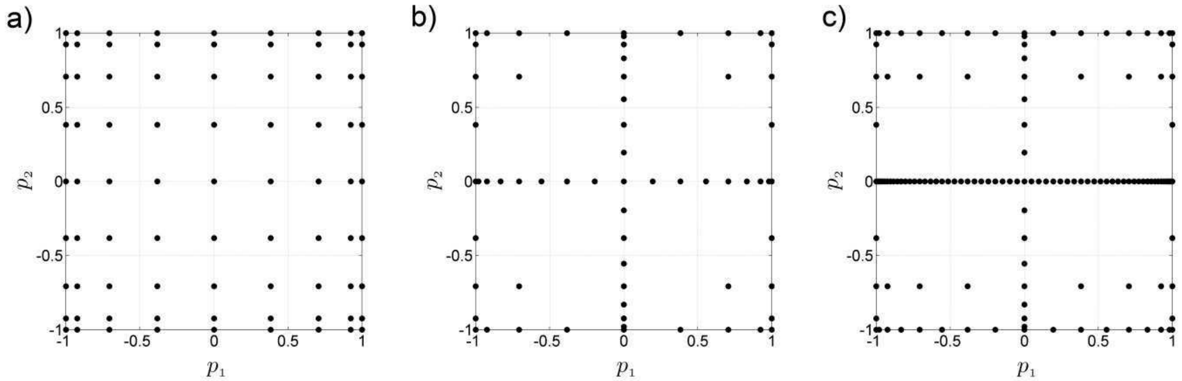

Figure 1. Three different sampling strategies of a two-dimensional parameter space $\Gamma = [-1,1] \times [-1,1]$: Cartesian grid (a); isotropic sparse grid (b); and anisotropic sparse grid with refinement along the direction of parameter p_1 (c).

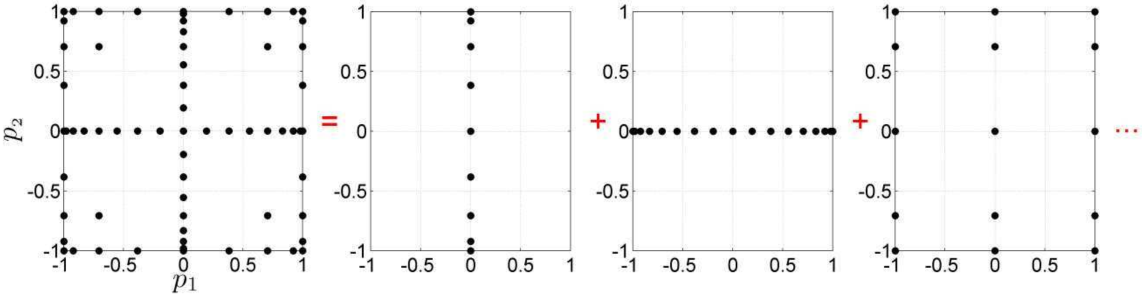

Figure 2. Graphical example of a sparse grid construction as a superimposition of tensor grids.

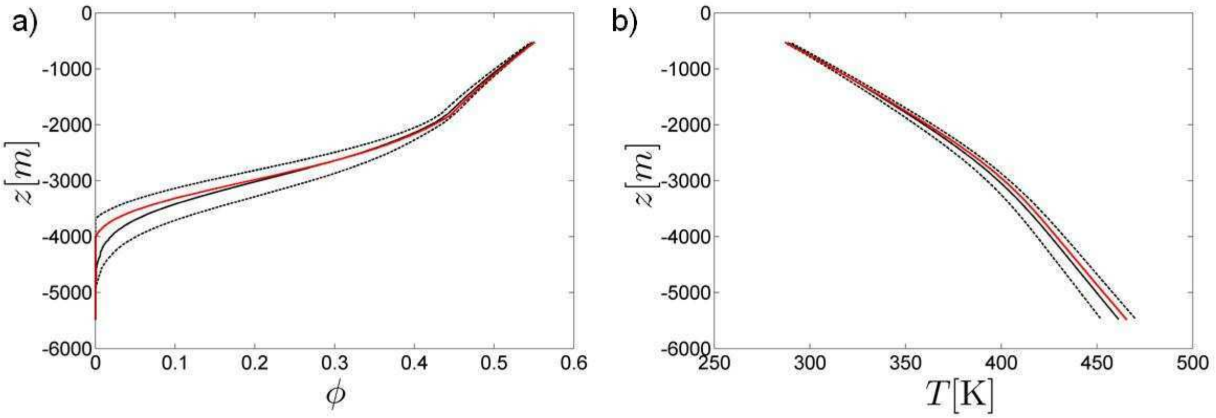

Figure 3. Vertical distribution of mean porosity (a) and temperature (b) (black solid lines) at final simulation time (t = 200 Ma). Intervals of width corresponding to one standard deviation about the mean are also shown as dashed black curves. Red curves represent the reference porosity and temperature fields.

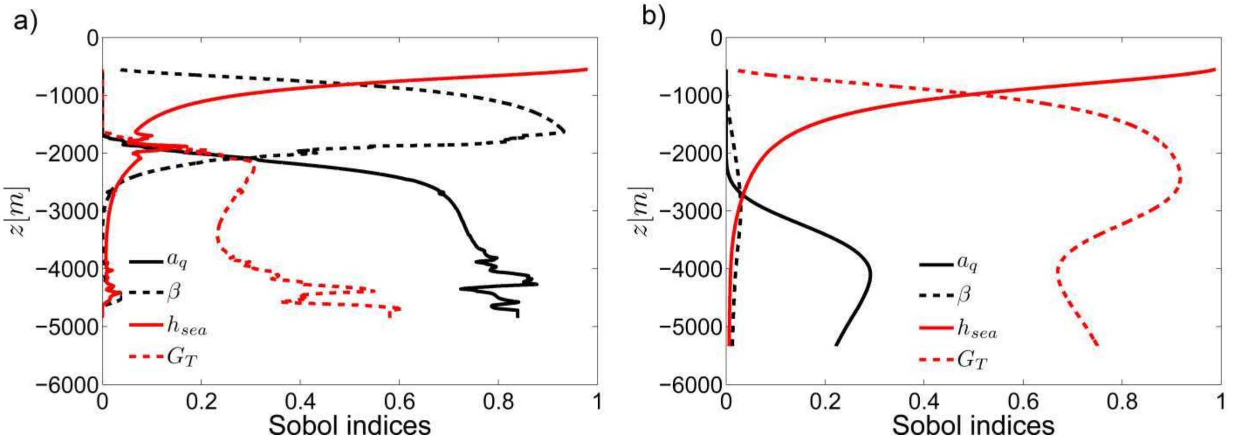

Figure 4. Total Sobol indices associated with porosity (a) and temperature (b) at the final simulation time (t = 200 Ma).

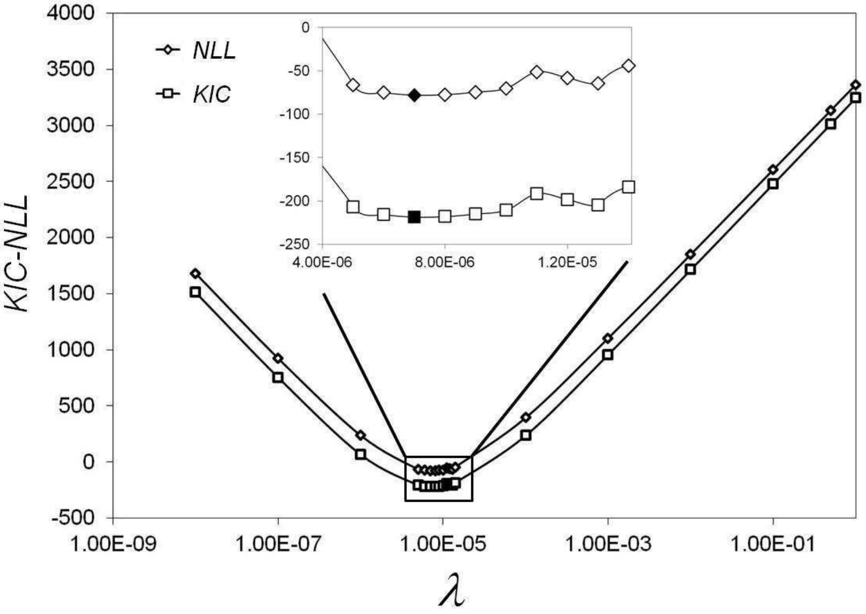

Figure 5. *KIC* and *NLL* versus λ. The insert shows the details of the behavior of these curves around the minimum value. Solid symbols correspond to minima of *KIC* and *NLL*.

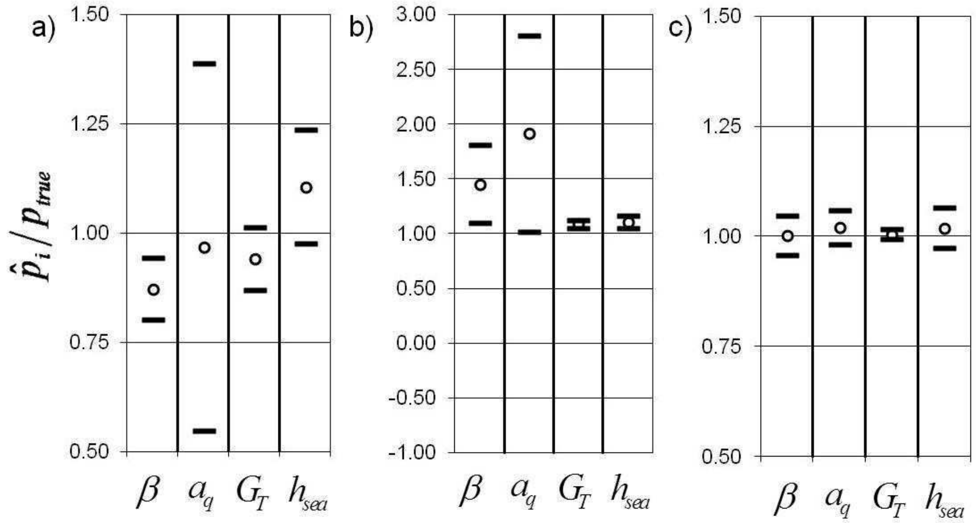

Figure 6. Normalized ML estimates of model parameters obtained through porosity (a), temperature (b), porosity and temperature data (c). Results in (c) are obtained by setting $\lambda = 7 \times 10^{-6}$. Symbols (–) indicate uncertainty bands of width $\pm \hat{\sigma}_{p_i} / p_{true}$.

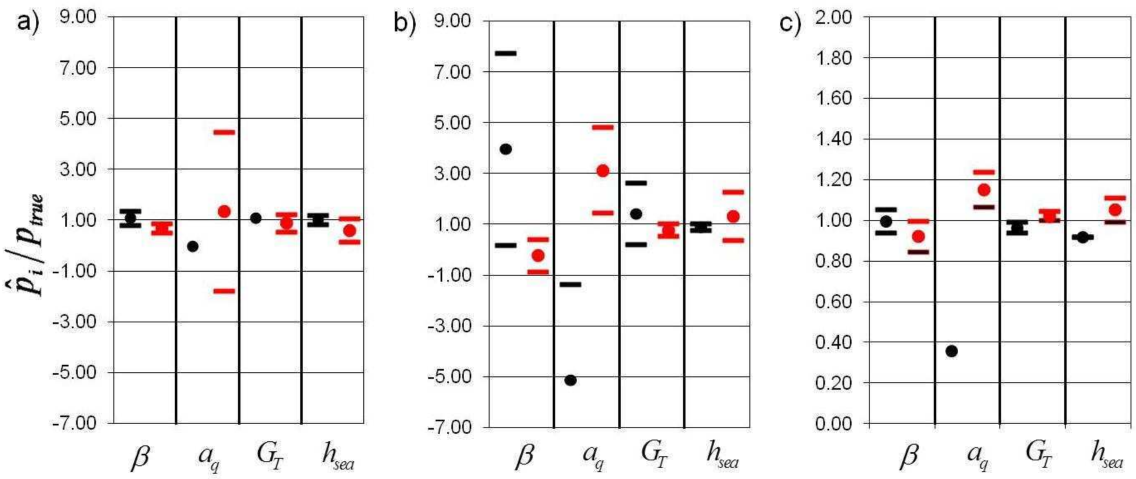

Figure 7. Normalized ML estimates of model parameters through porosity (a), temperature (b), porosity and temperature data (c). Black symbols refer to results obtained through calibration datasets (a) $\mathbf{\Phi}^*_{up}$, (b) $\mathbf{T}^*_{up}$, (c) $\left(\mathbf{\Phi}^*_{up},\mathbf{T}^*_{up}\right)$; red symbols to (a) $\mathbf{\Phi}^*_{lo}$, (b) $\mathbf{T}^*_{lo}$, (c) $\left(\mathbf{\Phi}^*_{lo},\mathbf{T}^*_{lo}\right)$. Results in (c) are obtained by setting $\lambda = 7\times10^{-6}$. Symbols (–) indicate uncertainty bands of width ± $\hat{\sigma}_{p_i}$ / p_{true}.

MOX Technical Reports, last issues

Dipartimento di Matematica “F. Brioschi”,
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

32/2013 Taddei, T.; Perotto, S.; Quarteroni, A.
Reduced basis techniques for nonlinear conservation laws

31/2013 Dassi, F.; Ettinger, B.; Perotto, S.; Sangalli, L.M.
A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain

30/2013 Cagnoni, D.; Agostini, F.; Christen, T.; de Falco, C.; Parolini, N.; Stevanović, I.
Multiphysics simulation of corona discharge induced ionic wind

29/2013 Lassila, T.; Manzoni, A.; Quarteroni, A.; Rozza, G.
Model order reduction in fluid dynamics: challenges and perspectives

28/2013 Ekin, T.; Ieva, F.; Ruggeri, F.; Soyer, R.
Statistical Issues in Medical Fraud Assessment

27/2013 Tagliabue, A.; Dede', L.; Quarteroni, A.
Isogeometric Analysis and Error Estimates for High Order Partial Differential Equations in Fluid Dynamics

24/2013 Mazzieri, I.; Stupazzini, M.; Guidotti, R.; Smerzini, C.
SPEED-SPectral Elements in Elastodynamics with Discontinuous Galerkin: a non-conforming approach for 3D multi-scale problems

25/2013 Cattaneo, Laura; Zunino, Paolo
Computational models for coupling tissue perfusion and microcirculation

26/2013 Ieva, F.; Paganoni, A.M.
Detecting and visualizing outliers in provider profiling via funnel plots and mixed effect models

23/2013 Srensen, H.; Goldsmith, J.; Sangalli, L.M.
An introduction with medical applications to functional data analysis