variables through a response surface that is defined with the help of an orthonormal polynomial basis in the parameter space. In simple words, the dependence of model output on all relevant input parameters is approximated by a high-dimensional polynomial. The resulting polynomials are functions of the model parameters. This projection can be interpreted as an advanced approach to statistical regression.

The PCE offers an efficient and accurate high-order way of including non-linear effects in stochastic analysis (e.g., Zhang and Lu, 2004; Foo and Karniadakis, 2010; Fajraoui et al., 2011). One of the attractive features of PCE is the higher-order in uncertainty quantification (e.g., Ghanem and Spanos, 1990, 1991; Le Maî tre and Knio, 2010), as well as its computational speed when compared to other methods for uncertainty quantification performed on the full model, such as MC (Oladyshkin et al., 2011). Due to its elegant reduction of models to polynomials, it allows performing many tasks analytically on the expansion coefficients. Alternatively, it allows performing excessive MC on the polynomials since they are vastly faster to evaluate than the original model.

Unfortunately, the original PCE concept (Wiener, 1938) is optimal only for Gaussian distributed input parameters. To accommodate for a wide range of data distributions, a recent generalization of PCE is the arbitrary polynomial chaos (aPC Oladyshkin et al., 2011). Compared to earlier PCE techniques, the aPC adapts to arbitrary probability distribution shapes of input parameters and, in addition, can even work with unknown distribution shapes when only a few statistical moments can be inferred from limited data or from expert elicitation. The arbitrary distributions for the framework can be either discrete, continuous, or discretized continuous. They can be specified either analytically (as probability density/cumulative distribution functions), numerically as histogram or as raw data sets. This goes beyond the generalization of PCE in methods such as the generalized polynomial chaos (gPC) or the multi-element gPC (ME-gPC) (Wan and Karniadakis, 2007; Xiu and Karniadakis, 2002). The aPC approach provides improved convergence in comparison to classical PCE techniques, when applied to input distributions that fall outside the range of classical PCE. A more specific discussion and review of involved techniques will follow in Sections 2.1–2.3.

With an introduction to response methods via the aPC, we describe here the theoretical background that we use in our modeling procedure. The related techniques for sensitivity and risk analysis used in this work are explained in Sections 4 and 5.

2.1. Definitions and polynomial chaos expansion

Suppose that we approximate a problem by a functional Υ, which represents the model responses Γ for the input variables Θ:

$$\Gamma \approx \Upsilon(\Theta). \tag{1}$$

Like all PCE methods, the aPC is a stochastic approach to approximate the response surface. Considering the uncertainty in the input variables, the aPC constructs a set of polynomial basis function and expands the solution in this basis. Thus, the response vector Γ in Eq. (1) can be approximated by Oladyshkin and Nowak (2012):

$$\Gamma \approx \sum_{i=1}^{n_c} c_i \Pi_i(\Theta). \tag{2}$$

Here, n_c is the number of expansion terms, c_i are the expansion coefficients, and Π_i are the multi-dimensional polynomials for the variables $\Theta = [\theta_1, \ldots, \theta_n]$, and n is the considered number of modeling parameters. If the model response $\Gamma(\Theta)$ depends on space and time, then so do the expansion coefficients c_i.

The number n_c of unknown coefficients c_i results from the number of possible polynomials with total degree equal to or less than d. This number depends on the degree d of the approximating polynomial, and the number of considered parameters n:

$$n_c = \frac{(d+n)!}{d!n!}. \tag{3}$$

2.2. Data-driven orthonormal basis

All polynomials Π_i in expansion (2) are orthogonal, i.e., they fulfill the following condition:

$$\int_{I\in\Omega} \Pi_l \Pi_m p(\Theta) d(\Theta) = \delta_{lm}, \tag{4}$$

where I is the support of Ω, δ is the Kronecker symbol, and $p(\Theta)$ is the probability density function for the input parameters. We obtain the orthonormal basis with the moments-based method proposed in Oladyshkin and Nowak (2012) and Oladyshkin et al. (2011). Orthonormality has the advantage that many subsequent analysis steps are accessible to relatively simple analytical solutions.

Knowledge on variability never is so perfect such that we could express the probability of model parameter values in a unique distribution function. Available data are mostly scarce, and fitting a density function to observed frequencies is often biased by subjective choices of the modeler. Oladyshkin et al. (2011) argued that, with aPC, it is possible to use available probabilistic information with no additional formal knowledge requirements for their probability distributions, only based on the statistical moments of the available data. They showed that, it is possible to calculate estimates for the mean, variance, and higher order moments of the model response $\Gamma(\Theta)$ even with incomplete information on the uncertainty of input data, provided in the form of only a few statistical moments up to some finite order.

2.3. Non-intrusive determination of the coefficients

The next task is to compute the coefficients c_i in Eq. (2). Generally, all PCE techniques can be sub-divided into intrusive (Ghanem and Spanos, 1993; Matthies and Spanos, 2005; Xiu and Karniadakis, 2003) and non-intrusive (Keese and Matthies, 2003; Isukapalli et al., 1998; Li and Zhang, 2007; Oladyshkin et al., 2011) approaches, i.e., methods that require or do not require modifications in the system of governing equations and corresponding changes in simulation codes. The challenge in choosing between the methods is to find a compromise between computational effort for model evaluations and a reasonable approximation of the physical processes by the interpolation.

For our study, we prefer the probabilistic collocation method (PCM: see Oladyshkin et al., 2011, 2011; Li and Zhang, 2007) from the group of non-intrusive approaches like sparse quadrature (Babuška et al., 2007; Xiu and Hesthaven, 2005; Gerstner and Griebel, 2003; Barthelmann et al., 2000). In a simple sense, PCM can be interpreted as a smart (mathematically optimal) interpolation and extrapolation rule of model output between and beyond different input parameter sets. It is based on a minimal and optimally chosen set of model evaluations, each with a defined set of model parameters (called collocation points). For this reason, the collocation approach became more popular in the last years. Also, the collocation formulation does not require any knowledge of the initial model structure. It only requires knowledge on how to obtain the model output for a given set of input parameters, which allows treating the model like a “black-box”. The distinctive feature of non-intrusive approaches is that any simulation model can

Please cite this article in press as: Ashraf, M., et al., Geological storage of CO_2: Application, feasibility and efficiency of global sensitivity analysis and risk assessment using the arbitrary polynomial chaos. Int. J. Greenhouse Gas Control (2013), http://dx.doi.org/10.1016/j.ijggc.2013.03.023