

# Polynomial Response Surfaces for Probabilistic Risk Assessment and Risk Control via Robust Design

Sergey Oladysshkin and Wolfgang Nowak  
*SRC Simulation Technology, Institute of Hydraulic Engineering, University of Stuttgart  
Germany*

## 1. Introduction

Many engineering systems represent challenging classes of complex dynamic systems. Lacking information about their system properties leads to model uncertainties up to a level where quantification of uncertainties may become the dominant question in modeling, simulation and application tasks. Uncertainty quantification is the prerequisite for probabilistic risk assessment and related tasks. Current numerical simulation models are often too expensive for advanced application tasks that involve accurate uncertainty quantification, risk assessment and robust design. This Chapter will present recent approaches for these challenges based on polynomial response surface techniques, which reduce massively the initial complex model at surprising accuracy. The reduction is achieved via projections on orthonormal polynomial bases, which form a so-called response surface. This way, the model response to changes in uncertain parameters and design or control variables is represented by multi-variate polynomials for each output quantity of interest. This technique is known as polynomial chaos expansion (PCE) in the field of stochastic PDE solutions. The reduced model represented by the response surface is vastly faster than the original complex one, and thus provides a promising starting point for follow-up tasks: global sensitivity analysis, uncertainty quantification and probabilistic risk assessment and as well as robust design and control under uncertainty. Often, the fact that the response surface has known polynomial properties can further simplify these tasks. We will emphasize a more engineering-like language as compared to otherwise intense mathematical derivations found in the literature on PCE. Also we will make use of most recent developments in the theory of stochastic PDE solutions for engineering applications. The current Chapter provides tools based on PCE for global sensitivity analysis, uncertainty quantification and risk analysis as well as design under uncertainty (robust design).

## 2. Response surface via polynomial chaos expansion

In the present Chapter, we consider the response surface in a closed polynomial form. Obviously, a response surface can be constructed in different ways, e.g. it can be constructed directly on a dense Cartesian grid of input parameters at extremely high computational efforts. Likewise, conceptually straightforward numerical Monte Carlo (MC) simulation techniques are computationally demanding since the statistical accuracy of their predictions

depend on the number of realizations used. In this Chapter we explore an alternative methodology which demands only minimum number of model evaluations to construct a response surface (see details in Section 2.3).

Our alternative approach is through the polynomial chaos expansion (PCE) introduced by Wiener (1938). Generally, all PCE techniques can be viewed as an efficient approximation to full-blown stochastic modeling (e.g., exhaustive MC). The basic idea is to represent the response of a model to changes in 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. 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, see e.g. Fajraoui et al. (2011); Foo & Karniadakis (2010); Zhang & Lu (2004). One of the attractive features of PCE is the high-order approximation of error propagation (Ghanem & Spanos, 1991; 1990) as well as its computational speed when compared to MC (Oladyshkin, Class, Helmig & Nowak, 2011b).

The original PCE concept (Wiener, 1938) can be used only for Gaussian distributed input parameters. In recent years, the classical PCE technique was extended to the generalized polynomial chaos (gPC) (Wan & Karniadakis, 2006) which accommodates for the use of an increased, yet limited number of statistical distributions (Askey & Wilson, 1985). The PCE methods discussed above assume an exact knowledge of the probability density functions of all input parameters and they are optimal only when applied to a finite number of certain parametric probability distributions. Unfortunately, information about the distribution of data or input parameters is very limited in many realistic applications, especially in environmental engineering and sciences. Data that characterize model parameters often indicate a variety of statistical distribution shapes (e.g., bounded, skewed, multi-modal, discontinuous, etc). Also, empirical parameter distributions derived from raw data sets do in general not follow analytically known distribution shapes. For such reasons, application tasks demand further adaptation of the chaos expansion technique to a larger spectrum of distributions.

To accommodate for a wide range of data distributions, a recent generalization of PCE is the arbitrary polynomial chaos (aPC). There are only very few studies that have used aPC before, and they can only be found in mathematical stochastics (Ghanem & Doostan, 2006; Soize & Ghanem, 2004) and aerospace engineering (Witteveen & Bijl, 2006; Witteveen et al., 2007). These studies focused on proofs of existence, constructing the basis by Gram-Schmidt orthogonalization and on related, quite theoretical issues. A notable exception is the very recent study of Li et al. (2011) in the field of petroleum engineering. That study did not discuss, however, the aPC in the light of data availability which plays a critical role in applications such as those presented by Oladyshkin, Class, Helmig & Nowak (2011a); Oladyshkin & Nowak (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 a raw data sets.

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 (Oladyshkin & Nowak, 2011). The necessity to adapt to arbitrary distribution is discussed in more details in Section 4 of the Chapter.

PCE techniques can mainly be sub-divided into intrusive (Ghanem & Spanos, 1993; Matthies & Keese., 2005; Xiu & Karniadakis, 2003) and non-intrusive (Isukapalli et al., 1998; Keese & Matthies, 2003; Li & Zhang, 2007; Oladyshkin, Class, Helmig & Nowak, 2011b) approaches, i.e., methods that require or do not require modifications in the system of governing equations and corresponding changes in simulation codes. Intrusive approaches require symbolic manipulations of the governing equations and can sometimes provide semi-analytical solutions for stochastic analyses of simple problems. The most well-known method from this group is the stochastic Galerkin technique. However, the necessary symbolic manipulations may become very complex and analytically cumbersome, and cannot easily be implemented in commercial codes. For this reason, non-intrusive approaches like sparse quadrature and the probabilistic collocation method (PCM: see Li & Zhang (2007); Oladyshkin, Class, Helmig & Nowak (2011b)) have lately been receiving a quickly increasing attention. In a simple sense, PCM can be interpreted as a smart (mathematically optimal) interpolation rule of model output between different parameter sets. The polynomial interpolation may be interpreted as a response surface of the model. It is based on a minimal and optimally chosen set of model evaluations, each with a defined set of model parameters (called collocation points). The challenge here is to find a compromise between computational effort and a reasonable approximation of the physical processes.

### 2.1 Definitions and polynomial chaos expansion

In this Section, we will introduce the aPC-based framework. aPC is for arbitrary Polynomial Chaos which is a most recent generalization of PCE methods (see Section 2.2). Let  $\omega = \{\omega_1, \dots, \omega_N\}$  represent the vector of  $N$  input parameters for some model  $\Omega = f(\omega)$ . The model  $\Omega(\omega)$  may be an explicit or implicit expression (e.g., a partial or ordinary differential equation or a coupled system). To perform sensitivity analysis, uncertainty quantification and risk assessment, we wish to investigate the influence of all parameters  $\omega$  on the model output  $\Omega$ . The model output can be space and time dependent, i.e.  $\mathbf{x} = (x_1, x_2, x_3)$ .

In the following, we will approximate the model response by a truncated polynomial expansion for each point in space  $\mathbf{x}$  and time  $t$ . According to polynomial chaos theory (Wiener, 1938), the model output  $\Omega$  can be approximated by polynomials  $\Psi_j(\omega)$  as follows:

$$\Omega(\mathbf{x}, t; \omega) \approx \sum_{j=0}^M c_j(\mathbf{x}, t) \Psi_j(\omega), \tag{1}$$

where the number  $M$  of polynomials depends on the total number of analyzed input parameters ( $N$ ) and the order  $d$  of the polynomial representation, according to the combinatoric formula  $M = (N + d)! / (N!d!) - 1$ . The coefficients  $c_j$  in Eq. (1) quantify the dependence of the model output  $\Omega$  on the input parameters  $\omega$  for each desired point in space  $\mathbf{x}$  and time  $t$ . The symbol  $\Psi_j$  is a simplified notation of the multi-variate orthonormal polynomial basis for  $\omega$  including all cross-terms between different parameters, as explained below. Let us mention that, in the current state of science for polynomial chaos expansion, the random variables have to be statistically independent or may be linearly correlated, which

can be removed by adequate linear transformation. Construction of a polynomial basis for statistically dependent random variables beyond linear dependence is a very important issue for future research.

## 2.2 Data-driven orthonormal basis

The support interval, weighting function or probability distribution for  $\omega$  is determined from available information (modeler's experience, expert opinion, general prior information or field data) and reflects the uncertainty or expected range of variation of input parameters. This implies that the polynomial basis and related projections should adapt to the acquired information, such that they approximate the model best where the probability density of the parameters is highest. In order to construct such a data-driven polynomial basis that considers all the available information about the input parameters  $\omega$ , let us define the set of polynomials  $\{P_j^{(0)}, \dots, P_j^{(d)}\}$  of degree  $d$  for the parameters  $\omega_j$  as an orthonormal basis in the parameter space. The polynomial  $P_j^{(k)}(\omega_j)$  of degree  $k$  in an individual parameter  $\omega_j$  can be written as a simple linear combination of the different powers  $i$  of  $\omega_j$ :

$$P_j^{(k)}(\omega_j) = \sum_{i=0}^k p_{i,j}^{(k)} \omega_j^i, \quad k = 0 \dots d, \quad j = 0 \dots N. \quad (2)$$

Here  $p_{i,j}^{(k)}$  are the coefficients within the polynomial  $P_j^{(k)}(\omega_j)$ . Assuming that the input parameters within  $\omega$  are independent (Ghanem & Spanos, 1991), the multi-dimensional basis can be constructed as a simple product of the corresponding univariate polynomials:

$$\Psi_k(\omega) = \prod_{j=1}^N P_j^{(\alpha_j^k)}(\omega_j), \quad \sum_{j=1}^N \alpha_j^k \leq M, \quad k = 1 \dots N, \quad (3)$$

where  $\alpha_j^k$  is a multivariate index that contains the combinatoric information how to enumerate all possible products of individual univariate basis functions. In other words, the index  $\alpha$  can be seen as  $M \times N$  matrix, which contains the corresponding degree (e.g. 0, 1, 2, etc.) for parameter number  $j$  in expansion term  $k$ . The multivariate basis allows to represent the reaction of a model  $\Omega$  to several ( $N$ ) parameters  $\omega_j$  ( $j = 0 \dots N$ ), as an  $N$ -dimensional polynomial response surface, defined by the expansion in Eq. (1).

We will show now, how to construct the data-driven orthogonal polynomial basis for each individual component  $\omega_j$  from the vector  $\omega$ . The main idea of the data-driven approach, see Oladyshkin, Class, Helmig & Nowak (2011a); Oladyshkin & Nowak (2011), consists in constructing the coefficients  $p_{i,j}^{(k)}$  for Eq. 2 in such a way that the polynomials in Eq. (2) form a basis that is orthonormal in precisely the given input distributions of model parameters. It does so without posing any restrictions to the statistical distribution shapes or weighting functions that available data, expert opinion or modeler experience may assume.

According to Oladyshkin & Nowak (2011), an orthogonal polynomial basis up to order  $d$  can be constructively defined for any arbitrary probability measure, given that  $\omega_j$  has finite statistical moments (e.g., mean, variance, skewness, etc) up to order  $2d - 1$ . The unknown polynomial coefficients  $p_{i,j}^{(k)}$  can be defined (among other available options for construction:

Abramowitz & Stegun (1965); Stieltjes (1884); Witteveen & Bijl (2006)) from the following matrix equation:

$$\begin{bmatrix} \mu_{0,j} & \mu_{1,j} & \dots & \mu_{k,j} \\ \mu_{1,j} & \mu_{2,j} & \dots & \mu_{k+1,j} \\ \dots & \dots & \dots & \dots \\ \mu_{k-1,j} & \mu_{k,j} & \dots & \mu_{2k-1,j} \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} p_{0,j}^{(k)} \\ p_{1,j}^{(k)} \\ \dots \\ p_{k-1,j}^{(k)} \\ p_{k,j}^{(k)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \\ 1 \end{bmatrix}. \tag{4}$$

Here,  $\mu_{i,j}$  are the  $i^{th}$  non-central (raw) statistical moments for random variable  $\omega_j$ . It becomes evident from Eq. (4) that statistical moments are the only required form of information on the input distributions. This property carries over to all other types of Taylor or polynomial chaos expansions.

The above orthogonal polynomial basis can be used directly for analysis. However, a normalized basis has further useful properties. For example, the mean and variance of  $\Omega(\omega)$  according to the expansion (1) is given by simple analytical relations (see Eq. (15) in Section 4 of this Chapter), by virtue of the orthonormality property. This follows from the general properties of Fourier expansions, which encompass all expansions in orthonormal bases. The squared coefficients of such expansions are called the spectrum (compare with the Fourier transformation). The sum of squared coefficients is the total energy or variance. Due to orthonormality, most terms cancel out in subsequent steps. For example, if we consider a stochastic process in the probability space  $(\Lambda, A, \Gamma)$  with space of events  $\Lambda$ ,  $\sigma$ -algebra  $A$  and probability measure  $\Gamma$  (see e.g. Grigoriu (2002)) and if  $\hat{P}_j^{(k)}(\omega_j)$  is an orthonormal basis, then by definition of orthogonality:

$$\int_{\omega_j \in \Lambda} \hat{P}_j^{(k)}(\omega_j) \hat{P}_j^{(l)}(\omega_j) d\Gamma(\xi) = \delta_{kl}. \tag{5}$$

The orthonormal polynomial basis can be obtained as:

$$\hat{P}_j^{(k)}(\omega_j) = \frac{P_j^{(k)}}{\|P_j^{(k)}\|}, \quad \|P_j^{(k)}\|^2 = \int_{\omega_j \in \Lambda} [P_j^{(k)}(\omega)]^2 d\Gamma(\omega_j), \tag{6}$$

where  $\|P_j^{(k)}\|$  is the normalizing constant of the polynomial  $P_j^{(k)}$  for space of events  $\Lambda$  (where  $\omega_j \in \Lambda$ ) with probability measure  $\Gamma$ .

Because the square of a polynomial of order  $k$  yields a polynomial of order  $2k$ , normalization according to Eq. (6) requires the statistical moments of  $\omega$  up to order  $2d$ .

When the input data set is small, the sample moments are only uncertain estimates of real moments. Hence, a direct application of the method presented becomes less robust. In that case, it would be useful to apply some standard methods to achieve robustness in the estimation of moments, such as bootstrapping (e.g. Efron (1987)). Moreover, especially in such cases, expert opinion can be very useful. In the presented approach, an expert will have total freedom of data interpretation (not restricted to the selection among standard PDFs) and can provide much more sophisticated information (e.g. lower and higher moments,

etc.). Such an expert opinion (in a most general sense) will be incorporated directly without any additional transformation or additional subjectivity when translating it to the stochastic numerical framework.

### 2.3 Non-intrusive determination of the coefficients

The remaining task is to evaluate the coefficients  $c_j$  in Eq. (1). Therefore, we will use the non-intrusive probabilistic collocation method (PCM) (Li & Zhang, 2007; Oladyskhin, Class, Helmig & Nowak, 2011b). The collocation formulation does not require any knowledge of the initial model structure, i.e., of  $\Omega$ . It only requires knowledge on how to obtain the model output for a given set of input parameters, which allows to treat the model  $\Omega$  like a "black-box". The distinctive feature of non-intrusive approaches is that any simulation model can be considered a "black-box", i.e. commercial software can be used without any modifications required. The idea of PCM is to evaluate the model exactly  $M$  times, which allows to directly fit the polynomial representation of  $\Omega$ , see Eq. (1), with its  $M$  unknown coefficients  $c_j$  to the obtained  $M$  model results. The  $M$  model evaluations are performed with  $M$  different parameter sets  $\{\omega_1^{(i)}, \dots, \omega_N^{(i)}\}, i = 1, \dots, M$ , called collocation points.

This leads to the following system (Villadsen & Michelsen, 1978) of linear equations:

$$\mathbf{M}_\Psi(\omega)\mathbf{V}_c(\mathbf{x}, t) = \mathbf{V}_\Omega(\mathbf{x}, t; \omega) \quad (7)$$

where  $\mathbf{V}_c$  is the  $M \times 1$  vector of unknown coefficients  $c_j$  in expansion (1), the  $M \times 1$  vector  $\mathbf{V}_\Omega$  contains the model output for each collocation point, and the  $M \times M$  matrix  $\mathbf{M}_\Psi$  contains the polynomials evaluated at the collocation points:

$$\begin{aligned} \mathbf{M}_\Psi &= \left\{ \Psi_i(\omega_1^{(i)}, \dots, \omega_N^{(i)}) \right\}, \quad i = 1 \dots M, j = 1 \dots M; \\ \mathbf{V}_\Omega &= \left\{ \Omega_i(x, y, z, t, \omega_1^{(i)}, \dots, \omega_N^{(i)}) \right\}, \quad i = 1 \dots M; \\ \mathbf{V}_c &= \{c_i(x, y, z, t)\}, \quad i = 1 \dots M. \end{aligned} \quad (8)$$

The vectors  $\mathbf{V}_c$  and  $\mathbf{V}_\Omega$  are space- and time-dependent, whereas the matrix  $\mathbf{M}_\Psi$  does not depend on space and time and can be generated once for the given expansion degree and parameter number.

The solution  $\mathbf{V}_c$  of the system (7) depends on the selection of collocation points. According to Villadsen & Michelsen (1978), the optimal choice of collocation points corresponds to the roots of the polynomial of one degree higher ( $d + 1$ ) than the order used in the chaos expansion ( $d$ ). Once the orthonormal polynomial basis is constructed using data (or assumptions on data), the collocation points become as well data-driven and optimally distributed in the space of input parameters. This strategy is based on the theory of Gaussian integration (e.g., Abramowitz & Stegun (1965)), and allows exact numerical integrations of order  $d$  given  $d + 1$  values of the function to be integrated.

The data-driven polynomial basis (see Section 2.2) defines the positions of the collocation points specific to the distribution of input parameters at hand and, thus, indicates what are the optimal parameter sets for model evaluation using all available information about the input parameters.

For multi-parameter analysis, the number of available points from the original optimal integration rule is  $(d + 1)^N$ , which is larger than the necessary number  $M$  of collocation

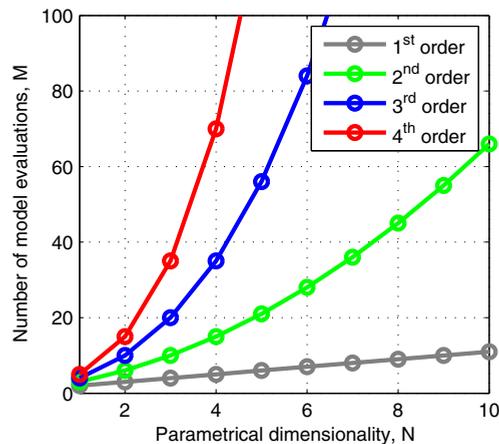


Fig. 1. Computational costs: number of model evaluations for different orders

points. The full tensor grid can be used for low-order (1<sup>st</sup>, 2<sup>nd</sup>) analysis of few parameters, but for higher-order analysis of many parameters the tensor grid suffers from the curse of dimensionality ((d + 1)<sup>N</sup> points). In that case, a smart choice of a sparse subset of the tensor grid becomes necessary. For this reason, the collocation approach became more popular in the last years. Probabilistic collocation (Li & Zhang, 2007; Oladyshkin, Class, Helmig & Nowak, 2011b) chooses the collocation points from the full tensor grid according to their probability weight, i.e. their importance as specified by the available probability distribution of ω. This simply means to select the collocation points from the most probable regions of the input parameters’ distribution (see Oladyshkin, Class, Helmig & Nowak (2011b)) and the modeler can extract a lot of information in the main range of the parameter distribution.

From the practical point of view, the computational costs of the data-driven chaos expansion combined with non-intrusive collocation approach are proportional to the number of terms in the chaos expansion multiplied by the time of a single model evaluation. The number of model evaluations (number of expansion terms M) for the construction of response surface via collocation approach depends on crossing order of expansion and presented in Figure 1.

**2.4 Efficient convergence of data-driven expansion**

The data-driven polynomial chaos expansion presented in Section 2.2 provides a simple and efficient tool for analysis of stochastic systems. Let us illustrate the efficiency of analysis within the data-driven versus the conventional expansion. For that, we will consider the simple exponential decay differential equation which was already used in Xiu & Karniadakis (2003) for illustration of the Askey scheme:

$$\frac{dY(t)}{dt} = -\zeta_{Ph}Y, \quad Y(0) = 1 \tag{9}$$

Let Y<sub>PC</sub> be the solution obtained using the polynomial chaos expansion (1) for the problem defined in equation (9). We use a Monte Carlo simulation as reference solution and define the relative error between the polynomial chaos expansion solution Y<sub>PC</sub> and the Monte Carlo solution Y<sub>MC</sub> as ε = |Y<sub>PC</sub> - Y<sub>MC</sub>| / |Y<sub>MC</sub>|.

We will apply the chaos expansion to equation (9) with both intrusive, like the Galerkin method (Ghanem & Spanos, 1993; Matthies & Keese., 2005; Xiu & Karniadakis, 2003), and non-intrusive approaches, like the probabilistic collocation method (Isukapalli et al., 1998; Li & Zhang, 2007; Oladyskhin, Class, Helmig & Nowak, 2011b). Both approaches depend on the distribution of the random variable  $\zeta$ . If the random variable  $\zeta$  is not distributed in same space as the polynomial basis, an additional conversion is required, such as e.g. Gaussian anamorphosis (Wackernagel, 1998). For each method, we will construct two expansions: conventional PCE mapped to Gaussian and data-driven aPC. Figure 2 demonstrates the convergence of the mean value (at time  $t = 1$ ) for a small selection of exemplary distributions of the model input  $\zeta$  (e.g. Rayleigh, Weibull, Log-normal). A similar situation has been observed for the convergence of the variance. More details are presented in the study by Oladyskhin & Nowak (2011). Chaos expansion in the data-driven (optimal) polynomial basis without transformation shows at least an exponential convergence. Let us mention, that the data-driven basis provides identical results for both intrusive and non-intrusive methods, because numerical integration is exact to the necessary degree when using the roots of the  $d + 1$  order polynomial from the optimal basis, and because no transformation from physical to normal space is necessary. Convergence with a non-optimal basis (here: Hermite) after transformation strongly depends on the nonlinearity of the required transformation.

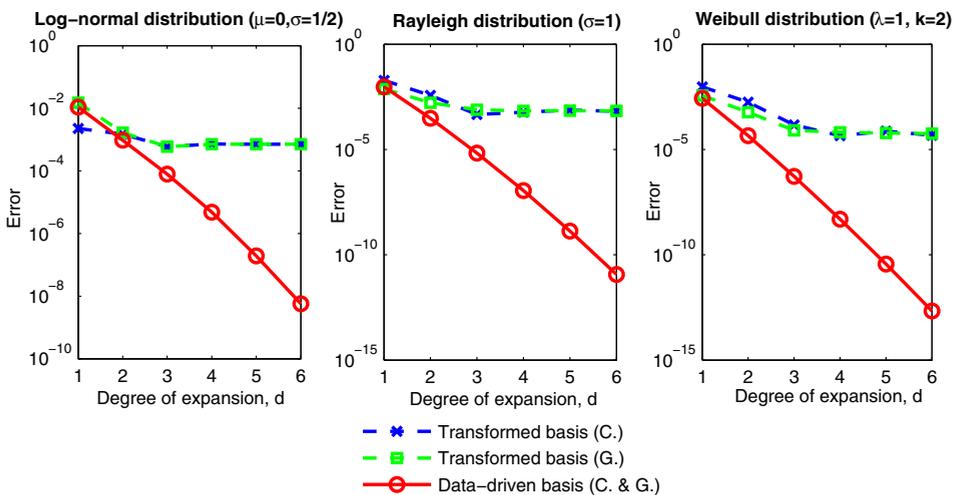


Fig. 2. Convergence estimating the mean value based on optimal data-driven basis and non-optimal transformed basis using Galerkin projection (G.) and Collocation approach (C.)

**2.5 Conclusions to section 2**

We presented the data-driven approach (aPC) for construction of a response surface based on a global orthonormal polynomial basis for arbitrary distributions. The arbitrary distributions can be either discrete, continuous, or discretized continuous and can be specified either through a few statistical moments, analytically as PDF/CDF, numerically as a histogram, or theoretically through the even more general format of a probability measure. 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 (Oladyshkin, Class, Helmig & Nowak, 2011a; Oladyshkin & Nowak, 2011).

The method only demands the existence of a finite number of moments, and does not require the exact knowledge or even existence of probability density functions. An interesting aspect is that only moments up to twice the order of expansion matter. Therefore, any PDFs fitted to input data will lead to the same statistical moments for model output, if the PDFs coincide in the required moments up to twice the order of expansion. For the same reason, if one still desires to fit a single PDF to the input data set, we recommend maximum entropy or minimum relative entropy methods applied such that the moments relevant for PCE are matched. This differs drastically from fitting low-parametric distributions to only lower moments of the available data, because this would modify the remaining relevant moments up to twice the order of expansion. More details on this issue will be provided in Section 4.

### 3. Global sensitivity analysis

Understanding the general role of parameters in models and the impact of varying model parameters on the response of prediction models is a relevant subject in various fields of science and engineering. Characterizing the impact of parameter variations is known as sensitivity analysis and can be subdivided into local and global analysis (Saltelli et al., 2008; Sobol, 1990; Sudret, 2008). In many cases of practical interest, we wish to perform a Global Sensitivity Analysis (GSA) in order to analyze a model as such or to investigate, quantify and rank the effects of parameter variation or parameter uncertainty on the overall model uncertainty. GSA can also be used to: (1) quantify the relative importance of each individual input parameter in the final prediction (Anderson & Burt, 1985; Sobol, 1990; Winter et al., 2006); (2) aid engineers to produce more robust designs; and finally (3) help decision makers to allocate financial resources towards better uncertainty reduction.

For example, the field of subsurface contaminant hydrology requires uncertainty estimates due to the ubiquitous lack of parameter knowledge caused by spatial heterogeneity of hydraulic properties in combination with incomplete characterization (Dagan, 1989; Rubin, 2003). For such reasons, we need to rely on probabilistic tools to predict contaminant levels and their overall health effects, and to quantify the corresponding uncertainties. Having efficient computational approaches to estimate uncertainty and to perform GSA in hydrogeological applications (and many other fields of science and engineering that feature uncertain dynamic or distributed systems) is desirable. It can inform modelers about the relevance of processes or parameters in the models they compile, and can inform engineers and decision makers about which parameters require most attention and where characterization efforts should be allocated such that prediction uncertainty can be minimized. Hence, there is an ever-increasing demand for having a GSA method that efficiently quantifies uncertainty and parameter relevance in complex and non-linear systems. An important recommendation to keep in mind is that GSA should be global not only in the sense of looking at the entire range of possible parameter variations. It should also be used to assess the importance of parameters on a global, final model output or post-processing result that is relevant to generate new insight, or relevant for final decisions. GSA should not merely be applied to model-internal quantities that are of secondary importance for the scientific or application task at hand (Saltelli et al., 2008).

In this Section, we tackle GSA based on the aPC technique, following the line of work on aPC by Oladyshkin, Class, Helmig & Nowak (2011a); Oladyshkin & Nowak (2011) and synthesizing the GSA approach by Oladyshkin, de Barros & Nowak (2011). Because the presented framework accounts for arbitrary bounds or weighting functions for input parameters, it provides a weighted global sensitivity. In some sense, the novel sensitivity indexes introduced by Oladyshkin, de Barros & Nowak (2011) can be perceived as a generalization of the Morris method (Morris, 1991) to weighted analysis. Up to presence the Morris method considers a uniform importance of input parameters within pre-defined intervals. We also generalize the Sobol indices (Sobol, 1990) for GSA to the aPC context, and provide a novel GSA measure which resembles a weighted square norm of sensitivities. Compared to Sobol indices, the new weighted sensitivity measure is absolute rather than relative. The advantage of an absolute index over a relative one is that it is a quantitative expression for the (averaged) derivative (slope), and hence keeps the original meaning of a sensitivity as known from linear, local analysis.

Performing GSA requires to evaluate the model at many points in the space of the input parameters. The correct choice of such points within the parameter space is important for adequate and efficient assessment of sensitivity. The aPC approach explicitly offers a method for optimal choice of these points, based on the generalized mathematical theory of Gaussian integration (e.g., Abramowitz & Stegun (1965)), see Section 2.

Foglia et al. (2007) pointed out that, according to their experience, one can get 70% of the insight from 2% of the model runs when using local sensitivity analysis methods versus global methods. The big advantage of aPC-based GSA (or more generally: GSA based on any PCE technique), is that one can obtain global sensitivity information at computational costs that are hardly larger than those for local analysis. The reason is the following: Local methods use infinitesimally small spacing between parameter sets for model evaluation to get numerical derivatives evaluated at a single point. The aPC based-method places the parameter sets for model evaluation at an optimized spacing in parameter space. This can be interpreted as fitting secants (or polynomials for non-linear analysis) to the model response. These secants (polynomials) approximate the model over the entire parameter space in a weighted least-square sense (compare with the best unbiased ensemble linearization approach described by Nowak (2009)). This is more beneficial to computing a tangent or local second derivatives (compare FORM, SORM methods, e.g., Jang et al. (1994)) that approximate the model well just around one point in the parameter space.

This Section provides an alternative procedure to perform GSA that is computationally efficient and highly flexible. In particular, due to aPC, the GSA introduced here and in Oladyshkin, de Barros & Nowak (2011) can be interpreted as exploiting a smart (mathematically optimal) interpolation rule of model output between optimally chosen sets of input parameters, where the number of model evaluations is minimal. Compared to earlier works that related GSA to classical PCE the presented approach: (1) emphasizes a more engineering-like language as compared to otherwise intense mathematical derivations and is based on a clear 3-step procedure to perform sensitivity analysis, (2) provides easy-to-use semi-analytical expressions for frequent use in applications, (3) generalizes PCE-based GSA to arbitrary probability distributions of the investigated parameters, and moreover, (4) allows to align the complexity and order of analysis with the reliability and detail level of statistical information on the input parameters.

### 3.1 Sobol sensitivity indices

Classical PCE has already been used for sensitivity analysis in different fields of applications (Crestaux et al., 2009; Sudret, 2008). Within the data-driven PCE framework (de Barros, Oladyshkin & Nowak, 2011; Oladyshkin et al., 2011) used in this work, the so-called Sobol indices for sensitivity estimation (Sobol, 1990) can be computed analytically based on the PCE (see Crestaux et al. (2009); Plischke (2010); Sudret (2008)) using Eq. (10):

$$S_{i_1, \dots, i_s} = \frac{\sum_{j=1}^M \chi_j c_j^2}{\sum_{j=1}^M c_j^2}, \quad \chi_j = \left\{ \begin{array}{ll} 1, & \text{if } \alpha_j^k > 0, \quad \forall j \in (i_1, \dots, i_s) \\ 0, & \text{if } \alpha_j^k = 0, \quad \exists j \in (i_1, \dots, i_s) \end{array} \right\}'$$

where  $S_{i_1, \dots, i_s}$  is the Sobol index that indicates what fraction of the of total variance of  $\Omega$  can be traced back to the joint contributions of the parameters  $\omega_{i_1}, \dots, \omega_{i_s}$ . The index selection operator  $\chi_j$  indicates where the chosen parameters  $\omega$  numbered as  $i_1, \dots, i_s$  (i.e.,  $\omega_{i_1}, \dots, \omega_{i_s}$ ) have simultaneous contributions within the overall expansion. In plain words, it enumerates all polynomial terms that contain the specified combination  $i_1, \dots, i_s$  of model parameters.

A complementing metric for sensitivity analysis is the *Total Index* introduced by Homma & Saltelli (1996). It expresses the total contribution to the variance of the model output  $\Omega$  due to the uncertainty of an individual parameter  $\omega_j$  in all cross-combinations with other parameters:

$$S_j^T = \sum_{(i_1, \dots, i_s): j \in (i_1, \dots, i_s)} S_{i_1, \dots, i_s} \tag{10}$$

where  $S_j^T$  simply sums up all Sobol indices in which the variable  $\omega_j$  appears, both as univariate and joint influences.

### 3.2 Weighted sensitivity indices

The weighted sensitivity index introduced by Oladyshkin, de Barros & Nowak (2011) reflects the square slope  $\partial\Omega/\partial\omega_j$ , but averaged over the statistical distributions or weighing functions of  $\omega_1, \dots, \omega_N$ :

$$S_{\omega_j}^2 = \int_{\omega_1 \in \Lambda} \dots \int_{\omega_N \in \Lambda} \left[ \frac{\partial\Omega(\boldsymbol{\omega})}{\partial\omega_j} \right]^2 d\Gamma(\omega_1) \dots d\Gamma(\omega_N) \tag{11}$$

So, the weighted global index for GSA using an arbitrary probability measure  $\Gamma$  can be used for reflection of the expected magnitude of parameter variation. Compared to Sobol indices (see Section 3.1) which analyze the spectral energy (variance) contribution of individual parameters to the model  $\Omega$ , the new weighted measure looks at the spectral energy within the derivatives  $\partial\Omega/\partial\omega_j$ . The averaged sensitivity index  $S_{\omega_j}$  can be explicitly expressed as:

$$S_{\omega_j}^2 = \sum_{k=0}^M c_k^2 \sum_{i=0}^{\alpha_j^k - 1} \left[ b_i^{(\alpha_j^k - 1)} \right]^2 P_j^{(i)}(\omega_j). \tag{12}$$

where the re-collection coefficients  $b_i^{(\alpha_j^k-1)}$  are defined as solution of the corresponding linear system:

$$\begin{bmatrix} b_0^{(\alpha_j^k-1)} \\ b_1^{(\alpha_j^k-1)} \\ \dots \\ b_{\alpha_j^k-1}^{(\alpha_j^k-1)} \end{bmatrix} \begin{bmatrix} p_{0,j}^{(0)} & 0 & 0 & 0 \\ p_{0,j}^{(1)} & p_{1,j}^{(1)} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ p_{0,j}^{(\alpha_j^k-1)} & p_{1,j}^{(\alpha_j^k-1)} & \dots & p_{\alpha_j^k-1,j}^{(\alpha_j^k-1)} \end{bmatrix}^T = \begin{bmatrix} p_{1,j}^{(\alpha_j^k)} \\ 2p_{2,j}^{(\alpha_j^k)} \\ \dots \\ \alpha_j^k p_{\alpha_j^k,j}^{(\alpha_j^k)} \end{bmatrix} \tag{13}$$

This index reflects the influence of a parameter  $\omega_j$  onto the model output  $\Omega$  in a similar fashion to the total index defined in Eq. (10). However, the new multivariate index in Eq. (11) does not rely on comparison among different parameters, i.e., it is an absolute measure. This is an advantage over the existing Sobol-based total index which is only a comparative and relative measure.

**3.3 Three-step algorithm for data-adaptive global sensitivity analysis**

This Section summarizes the computational algorithm for GSA based on aPC. The important feature in this computational algorithm is that it can be performed for arbitrary distributions of the input data  $\omega$ . The entire algorithm for the desired degree of precision  $d$  and number of parameters  $N$  is based on the following 3 steps:

*Step 1. Characterize the model parameters to be investigated:* Compute the raw moments  $\mu_{k,j}$  ( $k = 1 \dots 2d$ ) of the input data for each input parameter  $\omega_j$  ( $j = 1 \dots N$ ). If a probability density function (PDF) is provided, then we can evaluate the actual theoretical moments *in lieu* of the raw data moments of a given data set. Alternatively, expert elicitation may serve to provide opinions on these moments, e.g., by guessing a distribution shape and using its moments.

*Step 2. Approximate the model response surface by polynomials:* For the specific moments  $\mu_{k,j}$ , compute the coefficients of the optimal polynomial basis in Eq. (3) using the system of linear equations Eq. (4) and the normalization in Eq. (6). In addition, compute the coefficients of the expansion using the system of linear equations, Eq. (7), to represent the model response.

*Step 3. Compute the desired sensitivity information from the polynomials:* With the original model reduced to a multi-variable polynomial, the sensitivity indices can be obtained analytically without any heavy additional computational efforts. This is achieved by using the relations provided in Eqs.(10) or (12) provided in the two upcoming Sections.

In summary, the algorithm described above has the following advantages:

1. It constructs an optimal orthonormal polynomial basis for any desired distribution of data, whereas previous PCE techniques were limited to a small number of statistical distributions. Since only statistical moments are relevant, the input distributions may be either discrete, continuous, or discretized continuous and can be specified either through some statistical moments, analytically as PDF, numerically as histograms, or theoretically through the even more general format of probability measures.
2. The algorithm performs an optimal projection of the physical model onto a polynomial basis with minimum computational effort. This reduces the original model to a set of polynomials with many useful properties that allow immensely fast evaluation and offer a list of analytical relations.

3. The presented algorithm can easily be implemented without any deep knowledge related to the theory of chaos expansion and projection techniques.
4. The model  $\Omega$  does not have to be modified and no specific properties are required. Hence, it may be given in any arbitrary form for simple analytical solution up to a multi-scale multi-physical simulation software framework.

### 3.4 Illustration: sensitivity analysis for human health risk

#### 3.4.1 Problem formulation

We will demonstrate the presented GSA methodology for a contaminant transport problem in a 3D heterogeneous aquifer and the resulting human health risk for an exposed population. The aquifer has a hydraulic conductivity tensor  $\mathbf{K}(\mathbf{x})$  and constant effective porosity  $n_e$ . For illustration purposes, we consider flow to be incompressible, single-phased, at steady-state, free of boundary effects. The Random Space Functions for the log-conductivity  $Y = \ln K$  is statistically characterized by its mean  $\langle Y \rangle$ , variance  $\sigma_Y^2$ , covariance model  $C_Y$  (assumed here exponential) and its integral scale  $I_{Y,i}$ , where  $i = 1, 2$  and  $3$  for  $\mathbf{x} = (x_1, x_2, x_3)$ . We will consider  $I_Y \equiv I_{Y,1} = I_{Y,2}$  and  $I_{Y,v} \equiv I_{Y,3}$  with anisotropy ratio  $f = I_3/I_Y$ . Statistical stationarity for  $Y$  is assumed. Flow is uniform-in-the-average with mean velocity  $\langle \mathbf{u}(\mathbf{x}) \rangle \equiv (U, 0, 0)$ . In this test case, we will make use of the semi-closed expressions for one-particle displacement covariance function from Dagan (1988).

In most cases, decision makers are interested in quantifying adverse effects in human health due to contaminated groundwater exposure. For this case, the environmental performance metric of interest is human health risk, which depends not only on the hydrogeological parameters but also on the physiological and behavioural parameters of the exposed individual. For illustration purposes, we will consider the increased lifetime cancer risk model from the EPA (USEPA, December 1989;D), although many other risk models exist as discussed in the literature (de Barros, Ezzedine & Rubin, 2011; Maxwell & Kastenber, 1999; Siirila et al., 2011). The increased lifetime cancer risk formulation for the groundwater ingestion pathway is given by:

$$r = a\bar{C}, \quad \bar{C} = \max \left[ \frac{1}{ED} \sum_{t=0}^{t+ED} C(t) \right]_{t=0}^{\infty}, \tag{14}$$

where  $\bar{C}$  is the maximum running average over the exposure duration  $ED$  (years) defined in Maxwell & Kastenber (1999) and  $a$  is an uncertain health parameter (see Oladyshkin, de Barros & Nowak (2011) for details).

#### 3.4.2 Importance ranking of modeling parameters

We will apply the presented framework introduced in Section 3.3 for the human health risk analysis. These sensitivity indices are useful to quantify the simultaneous influence of model parameters, especially when the number of parameters becomes large and visualization of a multivariate model response becomes unfeasible and confusing. We will analyze an uncertain contaminant spill location and uncertain parameters in the model for health risk. Let us consider the integral scale  $I_Y$ , variance  $\sigma_Y^2$ , longitudinal velocity  $U$  and position  $x_s, y_s, z_s$  of the source with the source volume  $V_0$  and the slope of the risk model  $a$  as analyzed parameters  $\omega_j$  in our analysis.

According to Step 1 of the algorithm presented in Section 3.3, we first characterize all input parameters. For illustrative purpose we will consider the following distributions of modelling parameters:

- For  $\omega_1$ :  $\omega_1 = 1 + 2\omega'_1$ , with  $\omega'_1$  following a beta distribution with  $\alpha = 2$  and  $\beta = 2$ ;
- For  $\omega_2$ : Uniformly distributed within the interval  $[0.1, 0.7]$ ;
- For  $\omega_3$ : Log-normally distributed with  $\mu = 3.6$  and  $\sigma = 0.3$ .
- For  $\omega_4$ :  $\omega_4 = 9 + 2\omega'_4$ , where  $\omega'_4$  is beta distributed with  $\alpha = 2$  and  $\beta = 2$ ;
- For  $\omega_5$  and  $\omega_6$ :  $\omega_{5,6} = -1 + 2\omega'_{5,6}$ , where  $\omega'_{5,6}$  is beta distributed with  $\alpha = 2$  and  $\beta = 2$ .
- For  $\omega_7$ :  $\omega_7$  is lognormal distributed with  $\mu = 1.7$  and  $\sigma = 0.14$ .

These PDFs are used in Step 1 to calculate the corresponding raw moments of input data. According to Step 2 of the algorithm (Section 3.3), we compute the coefficients of the orthonormal polynomial basis using Eq. (4) and Eq. (6) for given raw moments and for the desired degree of expansion  $d$ . Also, we consider the response surface of the model according to Eq.(1) and Eq.(7). At the last phase of analysis, Step 3 of the algorithm (Section 3.3), the desired quantitative sensitivity information is extracted from the polynomial response surface.

All sensitivity indices can be constructed as it was shown above. For brevity, we will focus on the total and weighted indices only. Figure 3 illustrates the total (left plot) and weighted (right plot) sensitivity indices at different degrees of expansion. This figure shows the overall influence of all model parameters on the total human health effect (final prediction), which implicitly includes a time integral over the concentration history (see Eq. 14). For this particular case, we observe that the risk is more sensitive towards hydrogeological parameters than the health parameters. In the case study related to health risk (Figure 3), the new weighted measure converges faster in comparison to Total indices with increasing order of the expansion. Total indices show the main reaction of the model to the analyzed parameters, but as relative quantities do not provide faster stabilization in the ranking of analyses parameters in comparison to absolute weighted indices. The presented approach for GSA provides a relatively accurate approximation even for moderate orders (see also the convergence study for aPC in general in Oladyshkin & Nowak (2011)). Thus, global and weighted sensitivity analysis can already be performed at computational costs that are only slightly larger than those of local analysis.

### 3.5 Conclusions to section 3

We have presented an alternative method for global sensitivity analysis (GSA), which is based on the arbitrary polynomial chaos expansion (aPC). Compared to existing polynomial-based GSA methods, it can accommodate for all types of statistical distributions or weighting functions of the input parameters. This approach is denoted as *data-adaptive* because the aPC method can be applied even in situations where precise statistical information (e.g., known parametric distributions) for the input parameters is not available. If desired, the method can work directly with raw sampled data sets to represent the uncertainty and possible variational ranges of input data. The presented methods allow experts to choose freely of technical constraints the shapes of their statistical assumptions, and they allow to align the complexity and order of analysis with the reliability and detail level of statistical information on the input parameters.

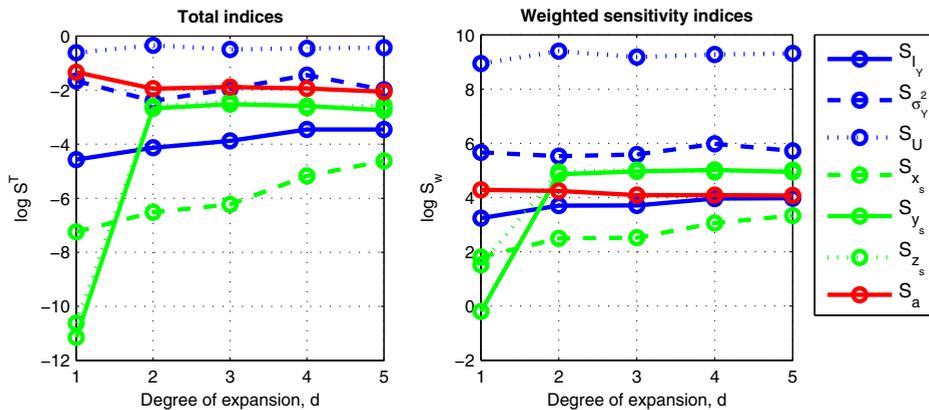


Fig. 3. Convergence of total (left plot) and weighted (right plot) sensitivity indices for the health risk prediction.

The proposed method incorporates the full range of possible simulation outcomes for the investigated model parameters as it approximates the model’s full response surface by multivariate polynomials. The existing polynomial-based method to compute Sobol indices was extended to the much more general aPC framework. While Sobol’s indices look at the contributions from individual parameters to the energy (variance) of the model, our weighted sensitivity indices look at the energy norm of derivatives of the model with respect to individual parameters. The resulting sensitivity measure shows better convergence properties than Sobol analysis. It can convey the information of a global sensitivity analysis at computational costs that are almost low as those of linear analysis.

#### 4. Uncertainty quantification and probabilistic risk assessment

This Section presents a reasonably accurate method for uncertainty quantification and probabilistic risk-assessment at acceptable computational costs. The lack of information about distributed properties leads to model uncertainties up to a level where the quantification of uncertainties becomes the dominant question in application tasks and may override the influence of secondary physical processes. Often numerical simulation models are inadequate for stochastic simulation techniques based on brute-force Monte Carlo simulation and related approaches, because even single deterministic simulations may require parallel high-performance computing. In the current Section, we suggest and apply a massive stochastic model reduction technique based on non-intrusive polynomial expansion as defined in Section 2. As a second focus for this Section, we continue discussion of data availability and subjectivity raised in the Section 2.

Unfortunately, precise information on distribution shapes for all uncertain input parameters is very rare in realistic applications, such as underground reservoir simulations, groundwater modeling, etc. Applied research on real-world systems, like modeling CO<sub>2</sub> storage, groundwater flow, etc., often faces the problem of immensely limited information about the model parameters involved, e.g. reservoir permeability, porosity, etc. With only limited data available or even in total absence of data, not even probability density functions representing the lack of knowledge can be easily inferred in a justified manner. Moreover, even if some

amount of data is available, the statistical distribution of the corresponding model parameters can be nontrivial, e.g. bounded, skewed, multi-modal or discontinuous. In any case, the attempt to construct probability density functions from samples of limited size or from sparse information introduces additional room for subjectivity into the analysis. Thus, applied tasks demand the direct handling of arbitrary distribution shapes and sparse data sets without additional assumptions.

The purpose of the current Section is to work with a highly parsimonious and yet purely data-driven description of uncertainty, applying the arbitrary polynomial chaos (aPC). It was shown in Section 2 that statistical moments are the only source of information that is propagated in all types of polynomial expansion-based stochastic approaches. Thus, exact probability density functions for uncertain input parameters do not have to be known and do not even have to exist. This avoids the necessity of assuming subjectively or speculating on the exact shapes of probability distributions. The new freedom in distribution shapes gained with aPC opens the path to accessing with PCE even those applications where data samples of limited size merely allow the inference of a few moments, and one would not be able to construct a probability density function without introducing subjective assumptions and hence dangerous sources of bias.

#### 4.1 Evaluation of statistics

In the presented approach, the statistics of the model output are based directly on the model and the specified moments of input data, see Section 2. If a model output  $\Omega(\omega)$  is expanded in the normalized polynomial basis Eq. (6), then characteristic statistical quantities of  $\Omega(\omega)$  can be evaluated directly. For example, the mean and variance of  $\Omega(\omega)$  are given by the following simple analytical relations:

$$\text{mean}(\Omega) = c_1, \quad \text{var}(\Omega) = \sum_{j=2}^N c_j^2, \quad (15)$$

where the latter is a result of Parseval's Theorem (e.g. Siebert (1986)).

Likewise, all moments of  $\Omega$  up to the order of expansion can be obtained analytically, based only on expansion coefficients and the used moments of the input parameters. Therefore, all probability distributions of input parameters that share the same  $2d$  moments will, in any polynomial chaos expansion of order  $d$ , lead to the same moments of model output of order  $d$ . For the same reason, PCE expansions of order  $d$  are unaffected by subjective choices concerning input parameter distributions that only affect moments beyond the order  $2d$ . This is the case, for example, for maximum entropy PDFs.

Cumulative distribution functions (CDF), probability density functions (PDF) and other arbitrary statistics of model output can be evaluated via Monte Carlo analysis of the polynomial response surface that results from the expansion Eq. (1). This is very fast, because the polynomial surface is much faster to evaluate than the original model equations. However, the latter approach will ask for full knowledge of the input probability density function involved, which we try to avoid. If Monte Carlo cannot be avoided, e.g. for computing exceedance probability in risk analysis, we suggest using maximum entropy PDFs, or alternatively, a sufficiently large data set (if available) that can be used directly as Monte-Carlo realizations of input parameters.

### 4.2 Problem of data availability

Unfortunately, information is not available for all sources of uncertainty. In such cases, additional assumptions become necessary. The classical way would be to introduce a theoretical (parametric) probability distribution, e.g. with an assumed mean and variance. For example, some expert could introduce a lognormal distribution for the leaky-well permeability (see details in Section 4.3.1) with parameters defined from the benchmark values (Class et al., 2009), see Figure 4.

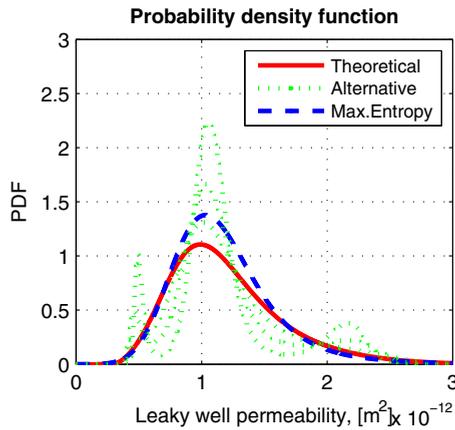


Fig. 4. Assumed stochastic distribution: theoretical PDF (red line), alternative PDFs (green lines) with the same first four moments and Maximum Entropy PDF with the same first four moments (blue line).

Establishing a full theoretical probability density function involves a strong assumption on all higher moments up to infinite order, and assumes implicit knowledge of the exact shape, e.g. also of the extreme-value tails. The data-driven approach strongly alleviates this situation, because it can handle a set of moments directly (e.g. the mean, variance, skewness, peakedness), without any further assumptions on higher-order moments and without having to introduce a PDF. In the current Section, we will use this freedom, and obtain only a small number of required moments directly from a large database or via experts, without asking for a full PDF.

For evaluating arbitrary complex output statistics without having a sufficiently large raw data set, we would recommend the Maximum Entropy PDF also shown in Figure 4 to draw a sufficiently large sample for Monte Carlo analysis of the polynomial. For example, a second-order expansion requires knowing the first four moments. Figure 4 illustrates possible distributions of stochastic variables, where the red line represents a theoretical PDF and the green lines represent a small collection of alternative PDFs with the same first four moments. The data-driven method does not require a choice between these alternatives, but only uses common information in the form of the required moments (here: up to order four). Existence of finite moments (up to a certain required order) is a necessary and sufficient condition for the proposed framework. Usually, this condition is easily fulfilled for a large spectrum of practical applications, especially for moderate degrees of expansion.

### 4.3 Illustration: risk assessment for carbon dioxide storage

#### 4.3.1 Problem formulation

We consider the benchmark leakage problem of injected CO<sub>2</sub> into overlying formations through a leaky well defined by Class et al. (2009). In the current Section, we present an illustrative example for a homogenized system, i.e. we consider spatial heterogeneity only through zonation according to different geological media. However, the technique presented can be extended to many classes of heterogeneous systems, where spatially correlated heterogeneous parameter fields can be decomposed into their uncorrelated principal components using the KL-expansion (e.g. Li & Zhang (2007)), if heterogeneity does not span over too many scales. CO<sub>2</sub> is injected into a deep aquifer, spreads within the aquifer and, upon reaching a leaky abandoned well, rises to a shallower aquifer. The goal of the simulation is to quantify the leakage rate which depends on the pressure build-up in the aquifer due to injection and on the plume evolution. The leaky well is at the center of the domain and the injection well is 100m away. Both aquifers are 30m thick and the separating aquitard has a thickness of 100m. The leaky well is modeled as a porous medium with a higher permeability than the formation. The CO<sub>2</sub> leakage rate is defined in the benchmark study as the total CO<sub>2</sub> mass flux integrated over a control plane midway between the top and bottom aquifer, divided by the injection rate, in percent.

The benchmark problem assumes that fluid properties such as density and viscosity are constant, all processes are isothermal, CO<sub>2</sub> and brine are two separate and immiscible phases, mutual dissolution is neglected, the pressure conditions at the lateral boundaries are constant over time, the formation is isotropic rigid and chemically inert, and capillary pressure is negligible. Within the presented approach, the physical complexity of the problem can be increased to an arbitrary extent because of the non-intrusive black-box conception of the probabilistic collocation method (Isukapalli et al., 1998; Li & Zhang, 2007). All relevant parameters used for the simulation are given in the paper by Oladyshkin, Class, Helmig & Nowak (2011a). In the current study, the benchmark problem is simulated using DuMuX, a multi-scale multi-physics toolbox for the simulation of flow and transport processes in porous media (Flemisch et al., 2007).

To illustrate the proposed methodology, we will consider three parameters uncertain: reservoir absolute permeability, reservoir porosity and permeability of the leaky well. The distributions of absolute permeability and porosity were taken from the U.S. National Petroleum Council Public Database (which includes 1270 reservoirs), see also the work of Kopp et al. (2009). This choice reflects the situation of site screening, where site-specific data and data that allow heterogeneity within geological units to be resolved are not yet available. Instead, we use data sets that represent macroscopic properties of supposedly similar sites as prior knowledge.

Formally, the investigated uncertain parameters correspond to the components of the vector  $\omega = \{\omega_1, \omega_2, \omega_3\}$  and represent the input parameters of Equation (1). The model output quantities  $\Omega$  considered here are pressure and saturation values as a function of space and time, and the CO<sub>2</sub> leakage rate through the leaky well as a function only of time.

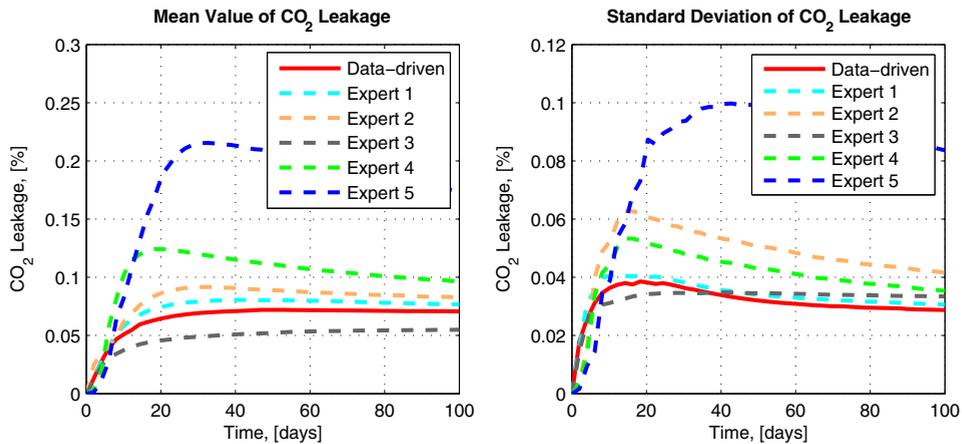


Fig. 5. Estimation of mean value (left plot) and standard deviation (right plot) of the CO<sub>2</sub> leakage rate: expert opinions (dashed lines) and data-driven approach (solid lines)

#### 4.3.2 Data-driven uncertainty quantification

Similar to GSA in Section 3 the procedure for quantifying uncertainty in CO<sub>2</sub> storage can be divided into three main steps. The first step is to construct the polynomial basis according to the data considered in Section 4.3.1. The second step is to set up the chaos expansion and obtain the required coefficients  $c_i$  for expansion (1) using the non-intrusive probabilistic collocation method. In the third step, we evaluate mean and variance of output statistics according to Section 4.1, see Figure 5.

The three steps mentioned above are straightforward and exploit the input data directly, i.e. in a data-driven manner. Involving additional assumptions on input data is possible, but can be dangerous and will import more room for subjectivity into the analysis. To illustrate the drastic impact of subjectivity that can be introduced into analysis, we performed the following simple experiment. The sample data (reservoir permeability and porosity) are distributed to five different and independent experts. The task of each expert consisted in constructing a theoretical probability density function for each parameter, which in their opinion would describe the statistics of the raw data best. As a common way of data description, all experts provided two-parametrical distributions for each input parameter (permeability and porosity). All experts proposed very different assumptions and techniques to match the permeability and porosity distributions. The responses of all experts were collected and used as input for modeling the benchmark problem defined in Sections 4.3.1. The results of this experiment are illustrated in Figure 5 using second-order polynomial chaos expansion. It shows the resulting mean value (left plot) and standard deviation (right plot) of the CO<sub>2</sub> leakage rate over time. Here, dashed lines correspond to the results based on subjective expert opinions, and solid lines correspond to the results from our purely data-driven approach. All test cases presented in Figure 5 were performed under the same conditions in all other aspects. The differences in predicting leakage rates based on the different experts are a consequence of their different interpretations and opinions on how to treat the data. This example clearly demonstrates that the room for subjectivity (when assigning theoretical

stochastic distributions to real data and thus modifying their higher-order moments) can lead to significant differences in the predicted values.

The data-driven polynomial chaos expansion proposed here for stochastic analysis is based directly on the moments of sampled data without intermediate steps of data re-interpretation. This avoids the subjectivity usually introduced when choosing among a small limited number of theoretical distributions to represent a natural phenomenon, and so avoids the problem illustrated in Figure 5. Complex models often exhibit stronger non-linearity, which can amplify the impact of subjectivity especially for assumptions on extreme value tails and higher-order moments. Subjectivity within data interpretation can cause the same order of magnitude in uncertainty or error as numerical, conceptual and stochastic approximation error, see details in the paper by Oladyskhin, Class, Helmig & Nowak (2011a).

#### 4.3.3 Efficient probabilistic risk assessment

The data-driven polynomial chaos expansion provides a simple but powerful tool for stochastic modeling and, in this case study, for the probabilistic risk assessment of CO<sub>2</sub> storage. The most integrative characteristic of the overall benchmark problem is the total leakage of CO<sub>2</sub>. To compare a quantitative characteristic that is most important in probabilistic risk assessment, we computed the cumulative probability function of the CO<sub>2</sub> leakage rate after 30 days (Figure 6). The cumulative density function represents the probability that the CO<sub>2</sub> leakage is less than or equal to a particular value. In the current work, we also apply the conventional PCE approach for direct comparison with the data-driven approach. The principal difference lies in constructing the polynomial basis. The conventional approach is based on Hermite polynomials, which are optimal for Gaussian random variables. Because the random variables  $\omega$  are not distributed Gaussian in compliance with the Hermite polynomial basis, an additional conversion is required. A large number of methods are based on transformations of the model variables  $\omega_{ph}$  from physical space to corresponding normal variables  $\omega_N$  via Gaussian anamorphosis (Wackernagel, 1998). This would implicitly define an exact PDF, which once again is an unjustifiably strong assumption. In contrast, the data-driven approach is based directly on the considered moments of the distributions or raw data of the uncertain input variables  $\omega$  in physical space. The left plot in Figure 6 corresponds to the conventional approach, i.e. involving Gaussian anamorphosis and Hermite polynomials. The right plot in Figure 6 corresponds to the data-driven approach without additional transformation. The convergence of both types of chaos expansions (dashed lines) was validated by traditional Monte Carlo simulations (solid lines) with 1270 realizations, where we sampled directly from the available data sets. We repeated the comparison study for different degrees of the chaos expansion, such as first order (4 evaluations), second order (10 evaluations), third order (20 evaluations) and fourth order (35 evaluations). Figure 6 demonstrates that the data-driven approach provides fast convergence. Even small degrees of the data-driven expansion (even the linear one in our specific case study) ensure adequate representation of the stochastic processes in the considered multiphase flow system. This fact can be very useful for fully resolved and complex real-world application challenges, where computational costs are very high even for a single model evaluation.

Convergence with the conventional approach (here: based on Hermite polynomials) strongly depends on the non-linearity of the required transformation from  $\omega_{ph}$  to  $\omega_N$  (Oladyskhin & Nowak, 2011). There are two sources of slow convergence (or errors) for chaos expansions

based on non-data-driven polynomial bases. First, the transformation from physical space to normal space introduces additional non-linear behavior into the overall problem, which will require a higher order of expansion to obtain comparable accuracy. Second, for non-intrusive methods that rely on numerical integration to obtain the coefficients  $c_i$ , the accuracy of numerical integration strongly depends on the choice of integration or collocation points. For example, in Gauss-Hermite integration (Abramowitz & Stegun, 1965), the polynomial basis of degree  $d$  defines the positions of integration points by the roots of the polynomial of degree  $d + 1$ . These integration points are optimal only if the weighting function (here: the probability measure) and the polynomial basis are in direct correspondence. Any non-linear transformation from  $\omega_{ph}$  to  $\omega_N$  destroys this direct relation. Thus, using a non-data-driven polynomial basis leads to a non-optimal placement of integration points, which causes a reduced accuracy of the integration (see Oladyskhin & Nowak (2011)). As a consequence, in comparison with the classical PCE, the same order of precision can be achieved with a smaller degree of expansion.

#### 4.4 Conclusions to section 4

In this Section, we presented a data-driven approach for uncertainty quantification and risk assessment, based on polynomial chaos expansion (PCE). The data-driven approach provides a response surface based on a global orthonormal polynomial basis for arbitrary distributions. Thus, the data-driven approach provides freedom for modeling even physical systems with unknown probability distribution functions, when only data sets of very limited size or only little prior knowledge is available. It offers a new approach to defining parameter uncertainty in stochastic analysis that avoids the subjectivity of assigning theoretical probability distributions. In a small fictitious example, we asked independent experts to select and fit parametric distributions to two raw data sets. The example demonstrated that subjectivity in data interpretation can easily lead to prediction bias. Modeling is a chain of many tasks. Like any chain, it is only as strong as its weakest link. Modeling results indicate clearly that the statistical treatment of input data is part of the chain, and that the subjectivity in assuming theoretical curves can weaken that link immensely.

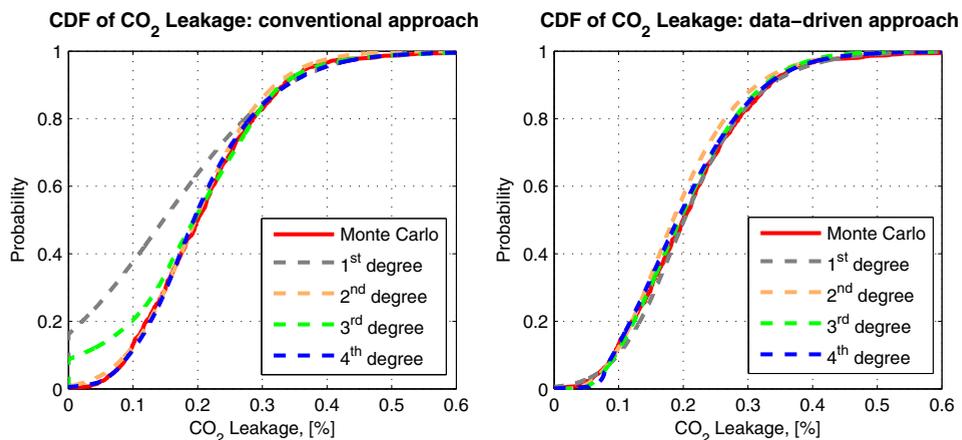


Fig. 6. Cumulative distribution function of CO<sub>2</sub> leakage rate after 30 days: conventional approach (left plot) and data-driven (right plot)

The data-driven stochastic approach was validated on the basis of Monte Carlo simulation using a common 3D benchmark problem. The proposed approach yields a significant computational speed-up compared with Monte Carlo, and provides faster convergence than conventional polynomial chaos expansions. Even for small degrees of expansion, the data-driven expansion can be very accurate, which can save a lot of computational power for probabilistic risk analysis.

Data-driven polynomial chaos expansion is based directly on raw data or other arbitrary sources of information without auxiliary assumptions. This increases the efficiency of chaos expansion and minimizes subjectivity, providing valuable support for risk-informed decision making as well as for robust design and control, allowing a better assessment and reduction of risk.

## 5. Robust design under uncertainty

The task of finding optimal system design while accounting for uncertainty leads to robust designs. In specific, risk can be controlled to stay below a given risk acceptance level, if robust design is performed under corresponding constraints. Obviously, the system performance and failure probability depends on uncertain parameters, but also depends on design variables. Evidently, the decision making will depend on the interplay between the response to design parameters, system uncertainty and, finally, the probability of failure or some statistical expectation of benefit. We propose to combine the model response to all design variables and uncertain parameters into a single approach based on an integrative (joint) response surface (obtained via PCE), which allows to reflect the non-linear dependence of the original model on all these parameters. This integrates the design task into the reduced stochastic model, allowing us to find robust designs with controlled risks at low computational costs. For example, designing system behaviors for maximum load at a guaranteed specific safety level can be achieved by quick Monte-Carlo methods on the resulting polynomials.

### 5.1 Integrative response surface

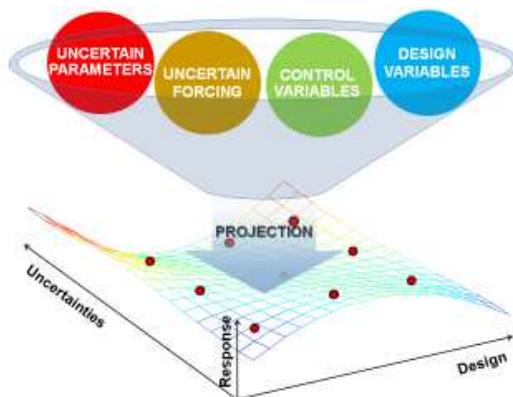


Fig. 7. Integrative response surface

Stochastic response surface methods (Isukapalli et al., 1998) deal with the characterization of uncertainties in systems describing the dependence of model output on the uncertain input parameters. Usually in many applied tasks, the model parameters can be classified in two classes: design or control variables that can be chosen by the operator of a system, and uncertain parameters that describe our (incomplete) knowledge of the system properties. On the other hand, the system's performance and failure probability will also depend on design parameters. Evidently, the decision-making for design parameters will depend on the interplay between the response to design parameters, system uncertainty and, finally, the probability of failure. In terms of system response, sensitivity and expansion, this distinction is artificial. Therefore, we drop the distinction between uncertain and design parameters. Instead, we project the model response to all design and uncertain parameters onto a single integrative model response surface. It is a multidimensional surface and contains the integral information about the system behavior under all possible conditions at all points in space and time. Thus, the notion of stochastic response surfaces introduced by Isukapalli et al. (1998) is expanded to integrative response surfaces forming an effective basis for robust design under uncertainty.

We investigate the influence of uncertain and design parameters  $\omega$  on the model output  $\Omega$  (integrative response surface approach). This means that the considered model has a multivariable input  $\omega$  for the expansion Eq. (1):

$$\omega_1, \dots, \omega_{N_U}, \omega_{N_U+1}, \dots, \omega_{N_U+N_D} \quad (16)$$

where  $N_U$  is the number of uncertain parameters and  $N_D$  is the number of design parameters. The total number of input parameters is  $N = N_U + N_D$ .

Of course design parameters do not have probabilistic distributions, but suitable weighting functions for such parameters can be described by user-defined feasibility functions that select the feasible range or preferences of the designing engineer concerning the values of design parameters. Feasibility functions provide a freedom for scenario analysis and can be used as an entry point for monetary punishment terms.

## 5.2 Illustration: robust design for carbon dioxide storage

### 5.2.1 Problem formulation

We again consider the problem of CO<sub>2</sub> leakage (Class et al., 2009) presented in Section 4.3. In the current Section we will consider the design task of finding an optimal injection regime. As in the Section 4.3 we consider three uncertain parameters: reservoir absolute permeability, reservoir porosity and permeability of the leaky well. For the current case study we also included two design parameters for describing the injection strategy: the CO<sub>2</sub> injection rate (fluctuating around 8.87 kg/s) and the size of the screening interval (up to 30m), see details in the paper by Oladyshkin, Class, Helmig & Nowak (2011b). The choice of the design parameters in this study is only exemplary and serves to demonstrate how engineering decision-making can be supported by the approach presented here. Both the injection rate and the screening interval directly affect the ratio of forces in the reservoir during the injection. The choice of feasibility functions is arbitrary, and modelers have full freedom to introduce feasibility functions and to weight them according to their personal experience or preferences. For example, we chose a beta distribution for the size of the screening interval with lower

and upper bounds of  $z=0$ [m] and  $z=30$ [m], respectively, reflecting the physical bounds of the reservoir.

### 5.2.2 Design under uncertainty vs. conventional design

The greatest challenge to modeling consists of finding a healthy and reasonable compromise between an accurate system representation and computational efforts. Unfortunately, neglecting uncertainty in design tasks can be a strong simplification for modeling, and the consequences can be stronger than when neglecting several physical phenomena.

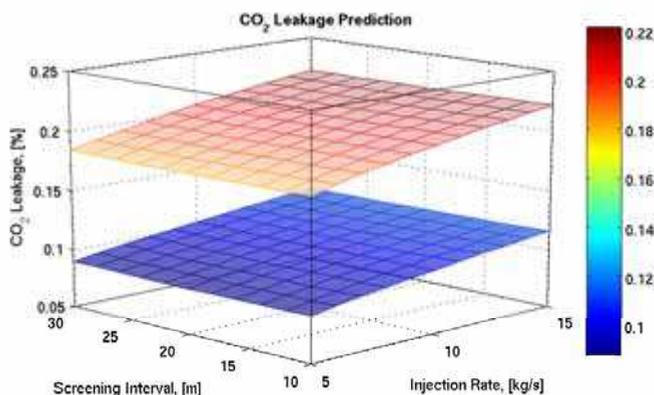


Fig. 8. Influence of design parameters on prediction of CO<sub>2</sub> leakage rate after 1000 days: top surface (robust design) - expected CO<sub>2</sub> leakage rate (average over uncertain parameters); bottom surface (conventional design) - CO<sub>2</sub> leakage rate evaluated deterministically with expected values of parameters

Figure 8 demonstrates how the injection rate and screening interval influence the leakage rate of CO<sub>2</sub>. An important advantage of the integrative response surface approach is that parameter uncertainty is easily included in such predictions. The top surface in Figure 8 is the CO<sub>2</sub> leakage rate expected after 100 days as a function of the design parameters, averaged over the uncertain parameters (robust design). The bottom surface in Figure 8 is the CO<sub>2</sub> leakage rate using the expected values of the uncertain parameters, i.e. as in deterministic simulations (conventional design). It is easy to see that the impact can be extremely important for non-linear systems (here, a factor of about two), especially in long-term simulations. Instead of looking at mean values, robust design can also look at failure probability (see below). Thus, designs that admit uncertainty are much more robust than design under deterministic assumptions.

### 5.2.3 Robust design of failure probability

The integrative response surface approach provides a constructive solution to the problem of robust design under uncertainty and provides valuable support for risk-informed decision making.

In a similar fashion as in the previous Section 5.2.2, the dependence of the leakage probability or any other statistical characteristics on design parameters can be evaluated, so that the

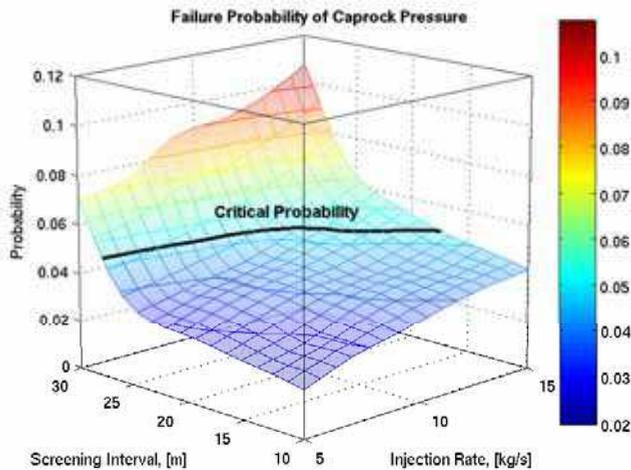


Fig. 9. Choice of design parameters based on caprock pressure after 1000 days: critical pressure 330 bar at a significance level of 5 %

injection regime can be chosen according to a maximum allowable failure probability. Figure 9 illustrates the choice of design parameters based on the caprock pressure after 1000 days. In this test case, a critical caprock pressure equal to 330 bar was chosen at a significance level of 5%, i.e. the maximal acceptable probability of failure is set to 0.05 (solid black line on surface). Figure 9 demonstrates acceptable strategies of injection where the caprock pressure does not exceed the limit of 330 bar, which corresponds to an injection-induced pressure build-up of about 40 bar.

### 5.3 Conclusions to section 5

This Section explores a massive stochastic model reduction via polynomial chaos expansion for robust design. This approach offers fast evaluation for statistical quantities and their dependence on design or control parameters. In particular, we map the response of a model to changes in all the design parameters and uncertain parameters onto one single integrative response surface. Based on this integrative concept, the design task explicitly includes uncertainty, which leads to robust designs with minimum failure probability.

We demonstrated that neglecting parametric uncertainty in design can be a strong simplification for modeling. Due to the non-linearity of processes, including uncertainty can lead to a systematic and significant shift of the predicted values, affecting both risk estimates and the design of injection scenarios. Thus, design under uncertainty, that openly admits uncertainty and seeks for robust solutions is much reliable in comparison to conventional design.

### 6. Acknowledgement

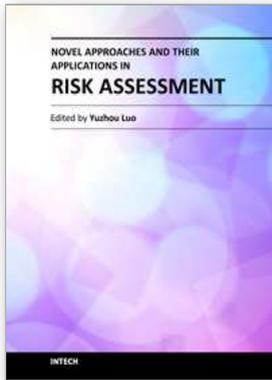
The authors acknowledge the German Research Foundation (DFG) for its financial support of the project within the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart.

## 7. References

- Abramowitz, M. & Stegun, I., A. (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover.
- Anderson, M. & Burt, T. P. (1985). Modelling strategies, in Anderson, M.G. and Burt, T.P. (eds.), *Hydrological Forecasting* pp. 1–13.
- Askey, R. & Wilson, J. (1985). *Some basic hypergeometric polynomials that generalize Jacobi polynomials*, *Memoirs of the American Mathematical Society*, AMS, Providence.
- Class, H., Ebigbo, A., Helmig, R., Dahle, H., Nordbotten, J. N., Celia, M. A., Audigane, P., Darcis, M., Ennis-King, J., Fan, Y., Flemisch, B., Gasda, S., Jin, M., Krug, S., Labregere, D., Naderi, A., Pawar, R. J., Sbai, A., Sunil, G. T., Trenty, L. & Wei, L. (2009). A benchmark-study on problems related to CO<sub>2</sub> storage in geologic formations, *Computational Geosciences* 13: 451–467.
- Crestaux, T., Le Maitre, O. & Martinez, J.-M. (2009). Polynomial chaos expansion for sensitivity analysis, *Reliability Engineering and System Safety* 94(7): 1161–1172.
- Dagan, G. (1988). Time-dependent macrodispersion for solute transport in anisotropic heterogeneous aquifers, *Water Resour. Res.* 24(9): 1491–1500.
- Dagan, G. (1989). *Flow and Transport in Porous Formations*, Springer Verlag, Berlin.
- de Barros, F. P. J., Ezzedine, S. & Rubin, Y. (2011). Impact of hydrogeological data on measures of uncertainty, site characterization and environmental performance metrics, *Advances in Water Resources*.
- de Barros, F. P. J., Oladyshkin, S. & Nowak, W. (2011). An integrative data-adaptive approach for global sensitivity analysis: application to subsurface flow and transport, *Geophys. Res. Abstr.*, Vol. 13, EGU General Assembly 2011.
- Efron, B. (1987). *The Jackknife, the Bootstrap, and Other Resampling Plans*, Society for Industrial Mathematics.
- Fajraoui, N., Ramasomanana, F., Younes, A., Mara, T. A. and Ackerer, P. & Guadagnini, A. (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.* (doi:10.1029/2010WR009639).
- Flemisch, B., J., F., Helmig, R., Niessner, J. & Wohlmuth, B. (2007). Dumux: a multi-scale multi-physics toolbox for flow and transport processes in porous media, In A. Ibrahimbegovic and F. Dias, editors ECCO3MAS Thematic Conference on Multi-scale Computational Methods for Solids and Fluids, Cachan, France, November 28–30, 2007.
- Foglia, L., Mehl, S., W., Hill, M., C., Perona, P. & Burlando, P. (2007). Testing alternative ground water models using cross validation and other methods, *Ground Water* 45(5): 627–641.
- Foo, J. & Karniadakis, G. (2010). Multi-element probabilistic collocation method in high dimensions, *Journal of Computational Physics* 229(5): 1536–1557.
- Ghanem, R. & Doostan, A. (2006). On the construction and analysis of stochastic models: Characterization and propagation of the errors associated with limited data, *Journal of Computational Physics* 217: 63–81.
- Ghanem, R. G. & Spanos, P. D. (1991). *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York.
- Ghanem, R. & Spanos, P. (1993). A stochastic galerkin expansion for nonlinear random vibration analysis, *Probabilistic Engineering Mechanics* 8: 255–264.
- Ghanem, R. & Spanos, P. D. (1990). Polynomial chaos in stochastic finite elements, *Journal of Applied Mechanics* 57: 197–202.

- Grigoriu, M. (2002). *Stochastic Calculus: Applications in Science and Engineering*, Birkhauser, Boston.
- Homma, T. & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models, *Reliability Engineering and System Safety* 52(1): 1–17.
- Isukapalli, S., S., Roy, A. & Georgopoulos, P., G. (1998). Stochastic response surface methods (srsms) for uncertainty propagation: Application to environmental and biological systems, *Risk Analysis* 18(3): 351–363.
- Jang, Y., S., Sitar, N. & Kiureghian, A., D. (1994). Reliability analysis of contaminant transport in saturated porous media., *Water Resources Research* 30(8): 2435–2448.
- Keese, A. & Matthies, H. G. (2003). Sparse quadrature as an alternative to mc for stochastic finite element techniques, *Proc. Appl. Math. Mech.* 3: 493–494.
- Kopp, A., Class, H. & Helmig, H. (2009). Investigations on CO<sub>2</sub> storage capacity in saline aquifers - part 1: Dimensional analysis of flow processes and reservoir characteristics, *Int. J. of Greenhouse Gas Control* 3: 263–276.
- Li, H., Sarma, P. & Zhang, D. (2011). A comparative study of the probabilistic collocation and experimental design methods for petroleum reservoir uncertainty quantification, *SPE Journal* pp. SPE-140738-PA-P.
- Li, H. & Zhang, D. (2007). Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods, *Water Resources Research* 43: 44–48.
- Matthies, H., G. & Keese., A. (2005). Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations, *Comp. Meth. Appl. Mech. Engrg.* 194: 1295–1331.
- Maxwell, R. M. & Kastenber, W. E. (1999). Stochastic environmental risk analysis: An integrated methodology for predicting cancer risk from contaminated groundwater., *Stochastic Environmental Research and Risk Assessment* (13(1-2)): 27–47.
- Morris, M., D. (1991). Morris factorial sampling plans for preliminary computational experiments, *Technometrics* (33): 161–174.
- Nowak, W. (2009). Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator, *Water Resour. Res.* 45(W04431).
- Oladyshkin, S., Class, H., Helmig, R., Nowak, W., de Barros, F. P. J. & Ashraf, M. (2011). Data-driven polynomial response surfaces as efficient tool for applied tasks under uncertainty, SAMSI Geosciences Applications Opening Workshop 2011.
- Oladyshkin, S., Class, H., Helmig, R. & Nowak, W. (2011a). A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations, *Advances in Water Resources* 34: 1508–1518.
- Oladyshkin, S., Class, H., Helmig, R. & Nowak, W. (2011b). An integrative approach to robust design and probabilistic risk assessment for CO<sub>2</sub> storage in geological formations, *Computational Geosciences* 15(3): 565–577.
- Oladyshkin, S., de Barros, F. P. J. & Nowak, W. (2011). Global sensitivity analysis: a flexible and efficient framework with an example from stochastic hydrogeology, *Advances in Water Resources* DOI: 10.1016/j.advwatres.2011.11.001.
- Oladyshkin, S. & Nowak, W. (2011). Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion, *Reliability Engineering and System Safety* .
- Plischke, E. (2010). An effective algorithm for computing global sensitivity indices (easi), *Reliability Engineering and System Safety* 95(4): 354–360.
- Rubin, Y. (2003). *Applied Stochastic Hydrogeology*, Oxford University Press, Oxford.

- Saltelli, A., Ratto, M. & Andres, T. (2008). *Global Sensitivity Analysis: The Primer*, John Wiley & Sons.
- Siebert, W. M. (1986). Circuits, signals, and systems, *MIT Press: Cambridge, MA* pp. 410–411.
- Siirila, E., R., Navarre-Sitchler, A., K., Maxwell, R., M. & McCray, J., E. (2011). A quantitative methodology to assess the risks to human health from co2 leakage into groundwater., *Advances in Water Resources* .
- Sobol, I. M. (1990). On sensitivity estimation for nonlinear mathematical models, *Mathem. Mod.* 2(1): 112–118.
- Soize, C. & Ghanem, R. (2004). Physical systems with random uncertainties: chaos representations with arbitrary probability measure, *SIAM J. Sci. Comput.* 26(2): 395–410.
- Stieltjes, T., J. (1884). Quelques recherches sur la théorie des quadratures dites mécaniques., *Oeuvres I* pp. 377–396.
- Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions, *Reliability Engineering and System Safety* 93(7): 964–979.
- USEPA (December 1989). Risk assessment guidance for superfund volume i: Part a, human health manual., *Tech. Rep. Rep.EPA/540/1-89/002* .
- USEPA (December 2001). Risk assessment guidance for superfund: Volume iii – part a, process for conducting probabilistic risk assessment, *Tech. Rep. Rep.EPA 540/R-02/002*, .
- Villadsen, J. & Michelsen, M. L. (1978). *Solution of differential equation models by polynomial approximation*, Prentice-Hall.
- Wackernagel, H. (1998). *Multivariate Geostatistics, An Introduction With Applications*, Second, Springer, Berlin.
- Wan, X. & Karniadakis, G., E. (2006). Multi-element generalized polynomial chaos for arbitrary probability measures, *SIAM Journal of Scientific Computing* 28(3): 901–928.
- Wiener, N. (1938). The homogeneous chaos, *Am. J. Math* 60: 897–936.
- Winter, C. L., Guadagnini, A., Nychka, D. & Tartakovsky, D. M. (2006). Multivariate sensitivity analysis of saturated flow through simulated highly heterogeneous groundwater aquifers, *Journal of Computational Physics* 217(1): 166–175.
- Witteveen, J. A. S. & Bijl, H. (2006). Modeling arbitrary uncertainties using gram-schmidt polynomial chaos, *44th AIAA Aerospace Sciences Meeting and Exhibit Reno, Nevada: AIAA–2006–896*.
- Witteveen, J. A. S., Sarkar, S. & Bijl, H. (2007). Modeling physical uncertainties in dynamic stall induced fluid-structure interaction of turbine blades using arbitrary polynomial chaos, *Computers and Structures* 85: 866–878.
- Xiu, D. & Karniadakis, G. E. (2003). Modeling uncertainty in flow simulations via generalized polynomial chaos, *Journal of Computational Physics* 187: 137–167.
- Zhang, D. & Lu, Z. (2004). An efficient, high-order perturbation approach for flow in random media via karhunen-loeve and polynomial expansions, *Journal of Computational Physics* 194: 773–794.



## **Novel Approaches and Their Applications in Risk Assessment**

Edited by Dr. Yuzhou Luo

ISBN 978-953-51-0519-0

Hard cover, 344 pages

**Publisher** InTech

**Published online** 20, April, 2012

**Published in print edition** April, 2012

Risk assessment is a critical component in the evaluation and protection of natural or anthropogenic systems. Conventionally, risk assessment is involved with some essential steps such as the identification of problem, risk evaluation, and assessment review. Other novel approaches are also discussed in the book chapters. This book is compiled to communicate the latest information on risk assessment approaches and their effectiveness. Presented materials cover subjects from environmental quality to human health protection.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Sergey Oladyshkin and Wolfgang Nowak (2012). Polynomial Response Surfaces for Probabilistic Risk Assessment and Risk Control via Robust Design, Novel Approaches and Their Applications in Risk Assessment, Dr. Yuzhou Luo (Ed.), ISBN: 978-953-51-0519-0, InTech, Available from: <http://www.intechopen.com/books/novel-approaches-and-their-applications-in-risk-assessment/polynomial-response-surfaces-for-probabilistic-risk-assessment-and-risk-control-via-robust-design>

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.