“Non-linearities and upscaling in porous media“

A comparison of different model reduction techniques for model calibration and risk assessment

Daniel Vogler

Supervisors:
Michael Sinsbeck
Sergey Oladyshkin
Wolfgang Nowak

Preprint 2013/5
Affidavit

Hereby I declare, that I prepared this thesis autonomously without illegitimate help and only with the sources duly noted and marked herein.

Daniel Vogler

Stuttgart, 15 August 2013
Topic: A comparison of different model reduction techniques for model calibration and risk assessment

Many engineering systems represent challenging classes of complex dynamic systems. Lacking information about their systems 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.

The current work will present recent approaches for these challenges based on response surface techniques, which reduce massively the initial complex model. The reduction is achieved by a regression-like analysis of model output with orthonormal polynomials that depend on the model input parameters. This way, the model response to changes in uncertain parameters, design or control variables is represented by polynomials for each model prediction 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: uncertainty quantification, model calibration and probabilistic risk assessment.

Obviously, a response surface can be constructed in different ways. Methods for constructing the response surface can demand only a minimum number of model evaluations, but as well may ask for many model evaluations to achieve a better quality of the involved projection integrals.

The scope of the current work is to test and compare different integration rules, i.e., methods to choose the sets of parameter values for which the model has to be evaluated. To test and compare the different methods, their accuracy in uncertainty quantification, model calibration and risk assessment will be measured against brute-force reference computations based on the original model. As illustrative example, we consider a study from the field of CO2 storage in the subsurface.

Connections to NUPUS:

- Application to a multiphase flow problem on CO2 storage in geological formations.
- Development and application of non-linear stochastic methods.

Supervision:

Jun.-Prof. Dr.-Ing. Wolfgang Nowak, M.Sc.
Dr.-Ing. Sergey Oladyshkin
Institut für Wasser- und Umwelt systemmodellierung /LH2 (SimTech)
Pfaffenwaldring 5a, 70569 Stuttgart, Germany
ACKNOWLEDGEMENTS

I want to thank my supervisors Wolfgang Nowak, Sergey Oladyshkin and Michael Sinsbeck for their guidance. Furthermore, I would like to thank my fellow coworkers for the relaxed and productive atmosphere.

I want to thank the NUPUS cooperation and the German National Academic Foundation for their support.

I would like to express my deepest appreciation to my parents and my family. They are awesome!
# TABLE OF CONTENTS

1 Introduction and Overview ........................................... 1

2 Materials and Methods .............................................. 4
   2.1 Arbitrary Polynomial Chaos Expansion ............................. 4
   2.2 Sampling Methods .................................................. 6
      2.2.1 Monte Carlo Sampling ....................................... 7
      2.2.2 Quasi-Monte Carlo Sampling ................................. 7
      2.2.3 Latin Hypercube Sampling .................................. 7
      2.2.4 Full Tensor Grid ............................................. 7
      2.2.5 Probabilistic Collocation Method ............................ 8
      2.2.6 OSC Method .................................................. 8
   2.3 Approach and Contribution ...................................... 8

3 Analysis of Sampling Strategies in the Light of Different Distribution and Problem Characteristics 9
   3.1 Data Distributions .................................................. 9
      3.1.1 Distributions with Small Tailing ............................ 9
      3.1.2 Distribution with Moderate Tailing .......................... 10
      3.1.3 Distribution with Strong Tailing ........................... 11
   3.2 Test Functions .................................................... 11
   3.3 Sampling .......................................................... 12
   3.4 Procedure Summary .............................................. 13
   3.5 Selected Test Functions ......................................... 13
      3.5.1 Distributions with Small Tailing ............................ 14
      3.5.2 Distributions with Moderate Tailing .......................... 22
      3.5.3 Distributions with Strong Tailing ........................... 27
      3.5.4 Summary ...................................................... 32
   3.6 Discussion ........................................................ 32

4 Numerical Study of a CO$_2$ Sequestration Setting 35
   4.1 Setup and Results of Numerical Simulations .................... 37
      4.1.1 Analysis of Degree of Expansion ............................ 37
      4.1.2 Computational costs constant for all methods ............... 39
   4.2 Discussion ........................................................ 41

5 Discussion .......................................................... 44

6 Conclusions ......................................................... 46

Bibliography .......................................................... 48
<table>
<thead>
<tr>
<th>Nomenclature</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendices</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.1</td>
<td>PDF for the selected data distributions</td>
</tr>
<tr>
<td>3.2</td>
<td>Resulting sampling points for two parameters that are log-Normally and log-Normally distributed</td>
</tr>
<tr>
<td>3.3</td>
<td>Resulting sampling points for two parameters that are Normally and Normally distributed</td>
</tr>
<tr>
<td>3.4</td>
<td>Beta Distribution - Exponential Function. Relative error of mean and standard deviation</td>
</tr>
<tr>
<td>3.5</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order for the 5th order polynomial function with a Beta distribution</td>
</tr>
<tr>
<td>3.6</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the 5th order polynomial function with a Beta distribution</td>
</tr>
<tr>
<td>3.7</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the trigonometric function with a Beta distribution</td>
</tr>
<tr>
<td>3.8</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the exponential function with a Beta distribution</td>
</tr>
<tr>
<td>3.9</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the runge function with a Uniform distribution</td>
</tr>
<tr>
<td>3.10</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the square root function with a Uniform distribution</td>
</tr>
<tr>
<td>3.11</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the 5th order polynomial function with a Normal distribution</td>
</tr>
<tr>
<td>3.12</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the trigonometric function with a Normal distribution</td>
</tr>
<tr>
<td>3.13</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the runge function with a Normal distribution</td>
</tr>
<tr>
<td>3.14</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the square root function with a Normal distribution</td>
</tr>
<tr>
<td>3.15</td>
<td>Log-Normal Distribution - Trigonometric Function. Relative error of mean and standard distribution</td>
</tr>
<tr>
<td>3.16</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the trigonometric function with a log-Normal distribution</td>
</tr>
<tr>
<td>3.17</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods investigating the exponential function with a log-Normal distribution</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.18</td>
<td>Comparison of the CDF of all approximating orders for the sampling methods</td>
</tr>
<tr>
<td></td>
<td>investigating the square root function with a log-Normal distribution.</td>
</tr>
<tr>
<td>4.1</td>
<td>Depiction of the scenario underlying the numerical test study investigated in this work.</td>
</tr>
<tr>
<td>4.2</td>
<td>Case 1a - CDF of $CO_2$ Leakage for Case 1 after 30 days with 10 sampling points for the QMC and LHS respectively.</td>
</tr>
<tr>
<td>4.3</td>
<td>Case 1b - CDF of $CO_2$ Leakage for Case 1 after 30 days with 20 sampling points for the QMC and LHS respectively.</td>
</tr>
<tr>
<td>4.4</td>
<td>Case 2a - CDF of $CO_2$ Leakage for Case 2 after 30 days with 20 sampling points for the QMC and LHS respectively.</td>
</tr>
<tr>
<td>4.5</td>
<td>Case 2b - CDF of $CO_2$ Leakage for Case 2 after 30 days with 40 sampling points for the QMC and LHS respectively.</td>
</tr>
<tr>
<td>4.6</td>
<td>Case 3 - CDF of $CO_2$ Leakage for Case 3 after 30 days with 20 sampling points for the QMC and LHS respectively.</td>
</tr>
<tr>
<td>4.7</td>
<td>CDF Bootstrap results of $CO_2$ Leakage after 30 and 100 days.</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.1</td>
<td>Color rating indicating the quality of performance for the sampling methods for different distributions</td>
</tr>
<tr>
<td>3.2</td>
<td>Performance Evaluation of Methods for all distributions</td>
</tr>
<tr>
<td>4.1</td>
<td>Simulation parameters</td>
</tr>
<tr>
<td>4.2</td>
<td>Degree of expansion (d) and number of sampling points (SP) for the sampling methods for all cases.</td>
</tr>
<tr>
<td>4.3</td>
<td>Degree of expansion (d) and number of sampling points (SP) for the bootstrap analysis.</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the trigonometric function with a Beta distribution.</td>
</tr>
<tr>
<td>2</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the exponential function with a Beta distribution.</td>
</tr>
<tr>
<td>3</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the 5th order polynomial function with a Normal distribution.</td>
</tr>
<tr>
<td>4</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the trigonometric function with a Normal distribution.</td>
</tr>
<tr>
<td>5</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the runge function with a Normal distribution.</td>
</tr>
<tr>
<td>6</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the square root function with a Normal distribution.</td>
</tr>
<tr>
<td>7</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the trigonometric function with a log-Normal distribution.</td>
</tr>
<tr>
<td>8</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the exponential function with a log-Normal distribution.</td>
</tr>
<tr>
<td>9</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the square root function with a log-Normal distribution.</td>
</tr>
<tr>
<td>10</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the runge function with a Uniform distribution.</td>
</tr>
<tr>
<td>11</td>
<td>Comparison of the CDF of all sampling methods for varying approximation order</td>
</tr>
<tr>
<td></td>
<td>for the square root function with a Uniform distribution.</td>
</tr>
</tbody>
</table>
1 Introduction and Overview

Many occurring natural and engineered systems are described by modeling physical processes in the system with mathematical models. These models depend on the physical properties that describe the system and their values. However, knowledge about the value range of these parameters can be scarce or highly uncertain. The evaluation of a system with numerical methods becomes more and more difficult with rising uncertainty of input parameters, especially for highly non-linear systems. This quickly leads to a state where evaluation of every possible combination via numerical methods becomes practically unfeasible. Therefore, this thesis investigates approaches to dealing with uncertainty quantification in complex, dynamic systems. As the systems are complex, non-intrusive methods are utilized, which have become very popular in recent decades.

In the current work, the problem is tackled by the employment of a response surface that is constructed with the arbitrary Polynomial Chaos (aPC) method [Olayshkin et al., 2012, Olayshkin and Nowak, 2012]. A response surface gives a result for the system for every combination of input parameters. For constructing a response surface, the results of the system have to be known for a certain number of input parameter values. An essential part of any polynomial chaos expansion method is the utilization of efficient sampling approaches for the integration points where the model is to be evaluated. There are numerous works (e.g., [McKay et al., 1979], [Metropolis and Ulam, 1949], [Tatang et al., 1997]) that employ different techniques to obtain results by minimizing the number of sampling points or choose them according to the distribution underlying the system or similar features. The sampling methods examined in this work are among the most popular and include Full Tensor Grid (FTG), Monte Carlo Method (MCM), quasi-Monte Carlo (QMC), Latin Hypercube Sampling (LHS), Probabilistic Collocation Method (PCM) and OSC method (c.f. Section 2.2).

This work will give insight on how to choose the sampling method according to the problem under investigation. Other areas of interest include the order of the utilized expansion and the impact of the underlying distribution on performance results.

As test cases, select test functions were studied to systematically investigate the influence of data distribution, computational effort, characteristics of the test function and other factors on the performance and uncertainty of the result. To apply a similar strategy to an application of great practical interest, a model was chosen that requires numerical simulation and evaluated (c.f. Section 2).
A wide range of test functions was tested (c.f. Section 3.2) and then narrowed down to 5 sample functions that are known to be difficult to approximate or because the functions showcase interesting behavior. These test functions were approximated with polynomial chaos expansions of various degrees. The distributions of the data showed small, moderate and strong tailing, to investigate the affects of the distribution on the performance of the methods.

The numerical model employed simulates $CO_2$ Sequestration in the subsurface and the primary focus lies on the amount of $CO_2$ leakage from the system (c.f. Section 4). Analogue to Section 3.2, the impact of the degree of expansion was investigated. Especially for QMC and LHS, the impact of the number of sampling points (SP) was studied. To obtain an indicator of the reliability of Monte Carlo Simulations, a bootstrap analysis was performed.

Results of the analytical and numerical case studies are shown in Section 3.3. The analysis of the performance concentrates on the evaluation of the uncertainty prediction (e.g., in form of cumulative density functions - CDF) of the results.

Chapter 5 summarizes the findings of the results and poses a select questions that were of interest to the research group and that can help a modeler to characterize systems under investigation.

- How do distributions with strong tailing affect the performance of sampling approaches? Tailing strongly affects moments of higher order, which are used for some sampling methods.

- For complex systems, the minimum number of points (dependend on the degrees of freedom) often does not lead to accurate results. Using more integration points poses the question whether weighting should be used to solve this overdetermined system or if simple regression should be employed.

- Will approximations of higher order lead to more accurate results?

- How difficult are the individual methods to implement? Does the additional effort of more complicated methods lead to better performance?

- Some methods are strongly based on the probability density functions (PDF) of the data distributions. What are possible challenges posed by methods that are strongly based on probability density (QMC, MCM)?

- Is the evaluation of performance based on mean and variance sufficiently accurate? What can be said of very high/low values of CDF curves?

The most important conclusions of the study are summarized in Section 6. An increase of sampling points does not necessarily lead to a better approximation of the solution. As the order
of the approximating response surface become higher, it is more essential to employ more sam-
pling points than required by the degrees of freedom. High order approximation show instabilities
if the sampling points are not efficiently sampled.
Future work could examine the affect of weighted sampling points, higher order parameter dis-
tributions (e.g., four input parameters and more), more complex distributions (e.g., bimodal
distributions) and complex, non-linear, practical applications (c.f. Section 6).
2 Materials and Methods

For the construction of the response surface, the arbitrary Polynomial Chaos Expansion (aPC) was employed after Oladyshkin et al., 2012 (see Section 2.1). The methods for sampling the integration points for the aPC were the Monte Carlo Method, quasi-Monte Carlo method, Latin Hypercube Sampling, Full Tensor Grid, Probabilistic Collocation Method and OSC method (see Section 2.2). Three different categories of parameter distributions (small, moderate and strong tailing - see Section 3.1) and their affects on the capabilities to create an accurate response surface for the analytical problem functions (see Section 3.2) were investigated.

2.1 Arbitrary Polynomial Chaos Expansion

This section will provide an overview over the polynomial chaos extension (PCE) and illustrate the theory of the arbitrary Polynomial Chaos Expansion. Many theoretical variations and implementations of the PCE method have been developed (e.g., Xiu and Karniadakis, 2002, Xiu and Karniadakis, 2003, Augustin et al., 2008, Wan and Karniadakis, 2006 and Prempraneerach et al., 2010). The PCE has found application in many fields (e.g., Crestaux et al., 2009, Barthelmann et al., 2000, Li and Zhang, 2007 and Shi et al., 2009) and there are many works on the performance of PCE methods (e.g., Ernst et al., 2012, Barthelmann et al., 2000, Field and Grigoriu, 2007, Xiu, 2009 and Haro Sandoval et al., 2012).

The method used for this research is called arbitrary polynomial chaos (aPC) and was developed by Oladyshkin and Nowak, 2012. The advantage of the aPC is the possibility to use only the raw moments of the available input data. This makes implementation of probability density functions optional and allows to use various arbitrary distribution shapes. Let $Y$ be the output of the model under investigation. Then, the aPC leads to a stochastic analysis $Y$ of the model $f(\omega)$, which depends on the parameters $\omega \in \Omega$.

$$Y(\omega) \approx \sum_{i=1}^{M} c_i \Phi_i(\omega_1, \omega_2, ..., \omega_N).$$ (2.1)

Here, the function $\Phi_i$ contains the multi-variate polynomial basis $P^{(i)}$ of the input parameters $\omega$, $c_i$ are the determined expansion coefficients and $M$ depends on the order of the expansion $d$, and the number of input parameters, $N$. The number of coefficients $M$ can be calculated with
the combinatory formula
\[ M = \frac{(N + d)!}{(N!d)!}. \] (2.2)

The function \( \Phi_i \) is a product of the polynomials that form the basis \( P^{(0)}, P^{(1)}, ..., P^{(d)} \), which is orthogonal (or orthonormal) with respect to the measure \( \Gamma \).

\[ \Phi_i(\omega_1, \omega_2, ..., \omega_N) = \prod_{j=1}^{N} P_j^{(\alpha_j^i)}(\omega_1, \omega_2, ..., \omega_N), \] (2.3)

with
\[ \sum_{j=1}^{N} \alpha_j^i \leq M, \quad i = 1, ..., N, ..., M. \] (2.4)

Here, \( \alpha_j^i \) is a multivariate index and contains the information how to enumerate all possible products of individual univariate base functions. The polynomial \( P^{(k)}(\omega) \) of degree \( k \) can be constructed as follows
\[ P^{(k)}(\omega) = \sum_{i=0}^{k} p_i^{(k)} \omega^i, \quad k = 0, d, \] (2.5)

where \( p_i^{(k)} \) are the coefficients of \( P^{(k)}(\omega) \). The basis for \( P^{(k)}(\omega) \), consisting of \( P^{(0)}, P^{(1)}, ..., P^{(d)} \) needs to be orthonormal (orthogonality is also sufficient), which is defined as
\[ \int_{\omega \in \Omega} P^{(k)}(\omega) P^{(l)}(\omega) d\Gamma(\omega) = \delta_{kl}, \quad \forall k, l = 0, d, \] (2.6)

with the Kronecker delta \( \delta_{kl} \). With the raw moments \( \mu_k \) of the variable \( \omega \) a system of linear equations can be written as
\[
\begin{bmatrix}
\mu_{0,j} & \mu_{1,j} & \cdots & \mu_{k,j} \\
\mu_{1,j} & \mu_{2,j} & \cdots & \mu_{k+1,j} \\
\vdots & \vdots & \ddots & \vdots \\
\mu_{k-1,j} & \mu_{k,j} & \cdots & \mu_{2k-1,j} \\
0 & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
p_0^k \\
p_1^k \\
\vdots \\
p_{k-1}^k \\
p_k^k
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
\vdots \\
0 \\
1
\end{bmatrix}. \] (2.7)

The authors [Oladyshkin and Nowak, 2012] further show that \( 2d - 1 \) moments of the parameter \( \omega_j \) are sufficient to construct an orthogonal polynomial basis up to order \( d \). The raw moments for a parameter \( \omega \) can be calculated as follows
\[ \mu_k = \int_{\omega \in \Omega} Y(t) \Psi_i(\omega) d\Gamma(\omega), \] (2.8)
with

$$\Psi^{k}(\omega) = \frac{1}{||P^{(k)}||} \sum_{i=0}^{k} p^{(k)}_{i} \omega^{i}. \quad (2.9)$$

If the matrix of moments in Eq. 2.7 is not singular, this enables simple computation of the coefficients $p^{(k)}_{i}$, based only on the input of the raw data available. The next step is the computation of the coefficients $c_{i}$ to be able to compute the output of the model (see Eq. 2.1) based only on the values of the input parameters $\omega_{j}$ and the polynomials calculated earlier. Common statistical quantities to investigate the accuracy of the aPC approximation of $Y$ are the mean and variance of $Y(\omega)$ according to [Li and Zhang, 2007]

$$\mu_{Y} = c_{1}, \quad \sigma_{Y}^{2} = \sum_{i=2}^{N} c_{i}^{2}. \quad (2.10)$$

Once $c_{i}$ has been calculated, it is possible to use data points to construct $P^{(k)}$ (Eq. 2.5) and then evaluate the function $\Phi_{i}$ (Eq. 2.3). With $\Phi_{i}$ and the coefficients $c_{i}$, it is possible to calculate the model output $Y(\omega)$, which depends on the individual data points for which a model response is desired. These computational steps are not costly, which allows for evaluation of large data sets.

### 2.2 Sampling Methods

If the input parameters of a model or function have a large range of values, it is difficult to obtain accurate knowledge about the global behavior of the system in question. This becomes more difficult as the number of input parameters that influence the output of the investigated system rises. Obtaining results for all possible combinations of parameter values is normally not feasible because calculations of complex systems can require extensive computational effort. Therefore, many methods have been developed that help to choose the parameter combination points for which the system will be investigated. A selected few of these methods have been investigated in the scope of this work. The arbitrary Polynomial Chaos Expansion requires points in which the model or function is to be evaluated.

There are numerous methods to sample the integration points to calculate the coefficients $c_{i}$ for the aPC. Notable examples include the Monte Carlo Method [Metropolis and Ulam, 1949], quasi-Monte Carlo Methods [Niederreiter, 1992], Latin Hypercube Sampling [McKay et al., 1979], Full Tensor Grid and Probabilistic Collocation Method [Tatang et al., 1997].
2.2.1 Monte Carlo Sampling

The Monte Carlo Method (MCM), which was described by [Metropolis and Ulam, 1949], is a very simple sampling method. The MCM chooses the sampling points at random and independently. In most technical applications, the points are chosen pseudorandomly, as a pseudorandom number generator can be used to compute a set of seemingly random numbers in a closed interval. Choosing random points prevents a biasing of the evaluated sampling points by focusing on certain, individually chosen, regions of interest. As points are distributed completely at random, this can also cause a distribution of sampling points that does not cover the entire output space evenly, or it does not cover regions that might be especially interesting.

The convergence rate of the Monte Carlo Method is $O(M^{1/2})$. This is a very slow convergence rate, but the convergence rate is independent of the number of parameters, which is a unique feature of the MCM.

2.2.2 Quasi-Monte Carlo Sampling

The quasi-Monte Carlo Method (QMC) is described in [Niederreiter, 1992] and [Morokoff and Caflisch, 1995]. QMC tries to cover the variable space more evenly. To achieve this, points are distributed with a low discrepancy. The discrepancy is low if the number of points that fall into an interval stands in the same relation to the number of points in the total space as the interval size stands to the total space. If the interval sizes tend to zero, and the number of total points tends towards infinity, the total space will be homogeneously covered.

2.2.3 Latin Hypercube Sampling

In Latin Hypercube Sampling (LHS), which was first introduced by [McKay et al, 1979], the parameter values from multidimensional distributions $\omega_j$ are each divided into $N$ strata of equal marginal probability $1/N$. From each of the $N$ strata, one value is then used, which leads to exactly $N$ values for each of the $N$ parameters. The values of the parameters $\omega$ are then matched randomly. Other studies on the performance and properties of LHS were discussed in [Helton and Davis, 2003], [Stein, 1987] and [Sloan and Kachoyan, 1987].

2.2.4 Full Tensor Grid

In the Full Tensor Grid (FTG) method, the sampling points are derived from Gaussian quadrature approximation of a polynomial of one order higher than the approximation order. The points of all
parameters are then combined as such, so that all possible combinations are evaluated. This makes the method very costly, but can also cover the multidimensional parameter space quite evenly. Therefore, the FTG shares many points with the PCM if the same order of approximation is chosen. Especially, for distributions with strong tailings, the FTG tends to sample points that are far from regions of high probability.

2.2.5 Probabilistic Collocation Method

The Probabilistic Collocation Method (PCM) was introduced by Tatang et al. [1997] among others and is based on the Gaussian quadrature approximation. It is designed to cover the high probability domain of the data distribution. The collocation points are the roots of the orthogonal polynomial of the input parameters of the order \( d + 1 \), if \( d \) is the desired order of polynomial approximation. The relative error of the method can be estimated with the orthogonal polynomials of order \( d + 2 \) [Tatang et al., 1997]. This work is strongly based on the works of [Olayshkin et al., 2012] and [Olayshkin and Nowak, 2012].

2.2.6 OSC Method

The OSC method was developed by Sinsbeck and Nowak at the University of Stuttgart. The method was included in this work for select problem settings to examine performance, limitations and other characteristics of the method. A detailed description of the OSC method has been submitted for publication to the International Journal for Uncertainty Quantification and is omitted here as it is out of the scope of this work.

2.3 Approach and Contribution

To evaluate the sampling methods introduced above and the accuracy of response surfaces created with aPC, various computational experiments were performed. To achieve a broad overview over the characteristics of the individual sampling methods and the PCE in general, multiple input properties were varied.

- Approximation of multiple test functions.
- Approximation of a numerical problem dealing with \( CO_2 \) sequestration.
- Investigation of the required order of the polynomials constructed by the PCE.
- Affect of the number of sampling points utilized on accuracy and computational cost.
- Affect of input distribution (Three classes with small, moderate and strong tailing).
3 Analysis of Sampling Strategies in the Light of Different Distribution and Problem Characteristics

3.1 Data Distributions

To judge the performance of the methods introduced above (c.f. Section 2.2), three commonly occurring classes of distributions were used. The distributions utilized displayed small, moderate and strong tailing. To present each class with an exemplary distribution, four distributions were chosen. Exemplary configuration of the sample distributions are shown in Figure 3.1. The employed data sets were normally, log-normally, beta and uniform distributed. These distributions can be classified as follows:

- Small Tailing: Uniform and Beta distribution
- Moderate Tailing: Normal distribution
- Strong Tailing: Log-normal distribution

3.1.1 Distributions with Small Tailing

Beta Distribution  The beta distribution employed for this work is symmetric around the mean \( \mu = 0 \) and has a standard deviation \( \sigma = 1 \), which results from the shape parameters of the beta distribution \( \alpha_{BD} = 1.5 \) and \( \beta_{BD} = 1.5 \). The distribution was stretched by 4 to fit the standard deviation \( \sigma \). With the distribution parameters, the distribution is obtained by

\[
f(x, \alpha_{BD}, \beta_{BD}) = 4 \times x^{\alpha_{BD} - 1}(1 - x)^{\beta_{BD} - 1}.
\]

(3.1)

Afterwards, the obtained values were shifted by the mean value of the distribution, thereby creating the mean value mentioned above.

Uniform Distribution  The uniform distribution was designed to have similar statistical properties (mean, standard deviation) as the normal and beta distributions. Hence, the distribution was bound by \( a = -1.73 \) and \( b = -a \) with the variable \( x \) uniformly randomly distributed
$x \in [-1.73, 1.73]$. The uniform distribution has no distinct region of high probability.

$$f(x, \alpha_{BD}, \beta_{BD}) = a + (b - a)x.$$ (3.2)

### 3.1.2 Distribution with Moderate Tailing

**Normal Distribution** The normally distributed data has a mean of 0, is symmetric around 0 and has unit standard deviation, to make it more comparable to the other distributions. The distribution is characterized by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$ (3.3)
3.1.3 Distribution with Strong Tailing

**Log-normal Distribution**  The log-normal distribution is characterized by the mean $\mu = 1$ and the standard deviation $\sigma = 1$. The distribution is then given by

$$f(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{\ln(x) - \mu}{2\sigma^2}}, x > 0. \quad (3.4)$$

3.2 Test Functions

To evaluate the performance of the introduced methods (see Section 2.2) and distributions (see Section 3.1), selected test functions were employed. The functions were chosen to cover a range of possible system behaviors. Each function depends on three parameters. The reference for each function was determined with Monte Carlo Simulations utilizing 100000 function evaluations.

**Exponential Function**  Exponential behavior of a system response can lead to very large values at one end of the parameter spectrum. This can significantly affect the fitted polynomials, especially when a least square error fit is used.

$$f(x_1, x_2, x_3) = e^{(1.5*(x_1+x_2+x_3))}. \quad (3.5)$$

**Polynomial of 5th Order**  Polynomials of various order were employed to test the convergence of the polynomial chaos expansion. The polynomial of 5th order was chosen to represent this group as the order of approximation used for the analysis never exceeded the 5th order.

$$f(x_1, x_2, x_3) = x_1 * x_2^2 * x_3^2 + x_3^3 * x_1^3 + x_1 * x_2^4. \quad (3.6)$$

**Runge Function**  The runge function is well known for being difficult to approximate with polynomials. This observation holds especially true for the outer bounds of the function.

$$f(x_1, x_2, x_3) = \frac{1}{(1 + x_1^2 + x_2^2 + x_3^2)}. \quad (3.7)$$

**Square Root Function**  Often, it is difficult to approximate functions that show abrupt behavior in specific points. A function based on a square root behavior was used to test the approximation of values that can not present negative values.

$$f(x_1, x_2, x_3) = \sqrt{|(x_1 + x_2 + x_3)|}. \quad (3.8)$$
Trigonometric Function  Many systems behave in a periodic fashion. A trigonometric function such as the cosine is well suited to investigate such characteristics.

\[ f(x_1, x_2, x_3) = \cos(x_1 + x_2 + x_3). \]  \hspace{1cm} (3.9)

3.3 Sampling

Figures 3.2 and 3.3 show the distribution of 36 sampling points for the five sampling methods. The OSC is not included in these Figures as the selection of sampling points is currently being submitted (see Section 2.2.6). In Figure 3.2 the two parameters are log-normally distributed while the parameters in Figure 3.3 are both normally distributed. Combinations that have different distributions for each of the parameters are also possible and occur often. For these examples, a polynomial base of higher order was chosen to obtain the sampling points for the PCM method to achieve an equal number of points for all sampling methods. The Figures 3.2 and 3.3 illustrate which affect, e.g., distributions with strong tailing have on the sampling approaches by the methods under investigation. The MCM, QMC and LHS place most of their sampling points into regions of high probability for the PDF of the respective distributions. As the FTG, PCM and OSC employ roots of higher order polynomials, the points are usually farther away from high probability regions and thereby account for moments of higher order.

Figure 3.2: Resulting sampling points for two parameters that are log-Normally and log-Normally distributed.
3.4 Procedure Summary

All test functions introduced above (Section 3.2) were approximated via arbitrary Polynomial Chaos Expansion (Section 2.1) while combining each test function with each data distribution (Section 3.1). Given the data set, the degree of the expansion and the number of sampling points, the FTG, QMC and PCM always result in the same sampling points, which does not occur for MCM and LHS. Therefore, MCM and LHS sampling was repeated 100 times for each combination of data distribution and test function. This allows to obtain insight into the variability of the results. The confidence interval of the 5% and 95% values of these 100 simulations is plotted as MCM SIG in the Figures sorting the results of the test functions by degree of approximating polynomial. The results for MCM and LHS depict the average of 100 performed simulations, denoted as MCM AVG and LHS AVG.

3.5 Selected Test Functions

Presenting all results and areas of investigation would exceed the scope of this work. Therefore, this study focuses on a few select points of interest and illustrates them with the according simulation results. Section 3.5.4 provides an overview over the performance of all methods. Two kind of plots are employed to illustrate the results: (1) The cumulative density function (CDF) for all approximation orders are plotted for each method individually, (2) The CDF of all methods are plotted together for each order of approximation. It is essential that all plots contain the
average of the MCM and LHS simulations (MCM AVG and LHS AVG), which were performed 100 times for each order of expansion, respectively. The figures sorted by order of expansion (e.g., Figure 3.5) contain the confidence interval of the 100 MCM runs. The confidence interval (MCM SIG) represents the bounds of the 5th highest and 95th highest results for the CDF, respectively, thereby giving the range of results for 90% of MCM runs.

3.5.1 Distributions with Small Tailing

**Beta Distribution - 5th Order Polynomial**  
The approximation of the 5th order polynomial (Figure 3.6) serves as an example to take a closer look at convergence behavior of the polynomial chaos expansion. Starting at a low order, the sampling methods do not provide an accurate fit of the CDF. However, almost all methods stay within the confidence range that is bordered by the 5% and 95% value for the 100 Monte Carlo Simulations that were run for the given number of sampling points. Accuracy of the predicted CDF increases and 3rd order polynomials provide a very accurate fit of the CDF, which leads to a perfect matching of the reference curve once the order of the approximating polynomials reaches the order of the test function. This case illustrates the possible implications of a steadily increasing approximation order with the corresponding increase in the number of sampling points required. A good representation of the CDF is achieved with a polynomial of 3rd order and 4th order. The approximation of 4th order requires a significantly larger number of sampling points and does not lead to a strong performance increase.

**Beta Distribution - Trigonometric Function**  
For the trigonometric function (Figure 3.7), all methods tend to stay within the confidence bounds marked by repeated Monte Carlo Simulations. This case shows very inaccurate fits of the CDF for small orders of approximation (e.g., approximation orders 1 and 2). Very accurate representations of the CDF are achieved for an approximation order of 4. This case also illustrates another difficulty of finding a balance between computational cost and accuracy. While the OSC method leads to accurate results for polynomials of order 4, the method fails to converge further and leads a vastly less accurate result for a polynomial of order 5. On this particular case, one observes the difficulty of determining whether a result is not satisfactory due to a small number of sampling points or if the behavior of the system favors a certain order of approximation.

**Beta Distribution - Exponential Function**  
Approximation of the reference CDF is reasonably accurate, even for low order approximation (see Figure 3.8). This result stands in stark contrast to the approximation of the exponential function with a log-Normal distribution (Figure 3.17), which
proves to be very difficult. This illustrates the strong dependency of the results and performance of the methods on the underlying distribution. The relative error of mean and standard deviation (Figure 3.4) converge steadily. However, the results for MCM and LHS show uncertainty.

Figure 3.4: Beta Distribution - Exponential Function. Relative error of mean and standard deviation. The MCM and LHS variability is illustrated with an error-bar.

**Uniform Distribution - Runge Function** For a uniform distribution (Figure 3.9), the FTG performs well, as there is no distinct region of high probability and the sampling points stabilize the polynomials by lying farther from the median of the distribution. PCM has difficulties, as the selection of the sampling points according to regions of high probability does not cope well with the absence of a distinct peak of the PDF. MCM, QMC and LHS display difficulties to capture very low and high values of the reference CDF. This behavior is typical for bounded functions.

**Uniform Distribution - Square Root** As with the runge function for uniform distributions (Figure 3.9), the square root function output (Figure 3.10) is fit acceptably well for CDF values from 0.2 to 0.8. The PCM does not lead to reliable results for uniform distributions.
Figure 3.5: Comparison of the CDF of all sampling methods for varying approximation order for the 5th order polynomial function with a Beta distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.6: **Beta Distribution - 5th Order Polynomial** Comparison of the CDF of all approximating orders for the sampling methods investigating the 5th order polynomial function with a Beta distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.7: **Beta Distribution - Trigonometric Function** Comparison of the CDF of all approximating orders for the sampling methods investigating the trigonometric function with a Beta distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.8: Beta Distribution - Exponential Function Comparison of the CDF of all approximating orders for the sampling methods investigating the exponential function with a Beta distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.9: **Uniform Distribution - Runge Function** Comparison of the CDF of all approximating orders for the sampling methods investigating the runge function with a Uniform distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.10: **Uniform Distribution - Square Root Function** Comparison of the CDF of all approximating orders for the sampling methods investigating the square root function with a Uniform distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
3.5.2 Distributions with Moderate Tailing

Normal Distribution - 5th Order Polynomial  The 5th order polynomial shows similar features for the normal distribution (Figure 3.11) as for the beta distribution. Convergence is most prominent for the first three orders of approximation, after which a reasonably good result has been obtained. MCM, QMC, LHS and PCM approximate the region of higher probability (e.g.: 0.1 – 0.9 of CDF) with an approximation order of 1 and 2 respectively, which predominantly lead to values between −5 and 5. Smaller and larger values are not captured with polynomial approximations of low orders, which is to be expected due to the stronger non-linearities displayed by a polynomial of 5th order.

Normal Distribution - Trigonometric Function  The trigonometric function with a normal distribution (Figure 3.12) displays good convergence behavior. In most cases, a first order approximation barely captures the rough shape of the CDF. Approximations of 2nd and 3rd order manage to capture the high probability region without exact representation of the bounded nature of the trigonometric function. This characteristic is displayed more exactly for approximation of high orders (e.g.: 4th and 5th order).

Normal Distribution - Runge Function  For the runge function (Figure 3.13), all methods display difficulties to exhibit precise values for extreme CDF values, especially for the low CDF values. Despite a large number of computations, MCM and LHS do not capture low CDF values, which remain prevalent even for high orders of approximations. The OSC displays effects commonly encountered with runge functions. While the aPC of 4th order captures the CDF reasonably well, the approximation orders of one degree lower and higher respectively (e.g., 3rd order and 5th order). Phenomena as displayed here make it inherently difficult to choose the desired order for the aPC, as performance of the aPC can strongly depend on the characteristics displayed by the system under investigation.

Normal Distribution - Square Root Function  The square root function (Figure 3.14) leads to comparable results as the runge function for a normal distribution. MCM and LHS show related difficulties to depict the boundedness of the function values precisely. Furthermore, 1st order approximations of QMC and PCM lead to results whose performance are competitive with approximations of high order (e.g., 4th order), which is noteworthy, given the difference in computational effort of 8 sampling points to 70 sampling points (for QMC) and 4 sampling points to 35 sampling points (PCM).
Figure 3.11: **Normal Distribution - 5th Order Polynomial** Comparison of the CDF of all approximating orders for the sampling methods investigating the 5th order polynomial function with a Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.12: **Normal Distribution - Trigonometric Function** Comparison of the CDF of all approximating orders for the sampling methods investigating the trigonometric function with a Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.13: Normal Distribution - Runge Function Comparison of the CDF of all approximating orders for the sampling methods investigating the runge function with a Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.14: **Normal Distribution - Square Root Function** Comparison of the CDF of all approximating orders for the sampling methods investigating the square root function with a Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
3.5.3 Distributions with Strong Tailing

Log-Normal Distribution - Trigonometric Function  The trigonometric function (Figure 3.16) is a good example of a function that does not cause inaccurate approximations with a log-normal distribution. Non-bounded functions can be difficult to approximate for log-normal distributions. This holds especially true for low approximation orders. Here, points drawn from tailing regions of the log-normal distribution can be magnitudes larger than points from high probability regions, which can strongly skew coefficients calculated by least square fitting. Looking at classical methods to compare performance, Figure 3.15 shows the relative error of the mean and standard deviation. While the relative error of the mean value of the FTG seems to converge until an expansion order of 4, the results for MCM, QMC and LHS show no convergence, with approximation of mean and standard deviation becoming worse for higher orders. However, Figure 3.16 shows good approximation of the reference CDF for all those three methods for expansion order of 2 and higher. This illustrates the difficulty of judging simulation results solely by mean and standard deviation.

Log-Normal Distribution - Exponential Function The exponential function (Figure 3.17) exhibits combinations (exponential function / log-normal distribution) that causes great approximation difficulties for some methods. Large values in the tailing of the distribution can instigate strongly diverging values. Especially FTG and PCM, whose sampling methods are greatly influenced by moments of higher order, can thereby display mostly inaccurate CDF.

Log-Normal Distribution - Square Root Function Approximation of the square root function with a log-normal distribution (Figure 3.18) enables an interesting comparison to the approx-
imation of the same function with an underlying normal distribution (Figure 3.14). The impact of the distribution is striking. Here, the input variables drawn from (or estimated and drawn from) a log-normal distribution do not contain negative values, which enhances simulation results as the output values are non-negative as well.
Figure 3.16: **Log-Normal Distribution - Trigonometric Function** Comparison of the CDF of all approximating orders for the sampling methods investigating the trigonometric function with a log-Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.17: Log-Normal Distribution - Exponential Function
Comparison of the CDF of all approximating orders for the sampling methods investigating the exponential function with a log-Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3.18: **Log-Normal Distribution - Square Root Function** Comparison of the CDF of all approximating orders for the sampling methods investigating the square root function with a log-Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
3.5.4 Summary

While Section 3.5 deals with some select phenomena and observations, this section attempts to provide an overview of the performance of the sampling methods. For this purpose, a color table is employed, which enables a quick qualitative impression. The results are grouped by the investigated test functions (see Section 3.2). It is essential to point out that the displayed performance for MCM and LHS is based on the average value derived from all calculations using bootstrapping. This means that the actual computational costs necessary to achieve the results for MCM and LHS far outweigh those of all other methods. As the sampling points for each realization of MCM and LHS are never the same, it proves difficult to compare MCM and LHS to other methods as the results strongly vary for each realization. While the computational costs required for MCM and LHS may not be competitive in terms of performance, the average values are included in this table nonetheless, for comparison.

For a long tailing distribution like the log-Normal distribution, a polynomial of low order can give a very accurate representation for the FTG and the PCM, as the roots of the polynomial for low orders are in the main probability region. For higher order polynomials, the roots move to the low probability regions and therefore, the CDF representation becomes inaccurate. For the QMC and the LHS, the points are chosen according to the CDF, which leads to the selection of the points from the high probability region.

3.6 Discussion

Investigation of the results obtained for the test functions shows that the FTG does not achieve good results despite the comparatively large computational effort due to the large number of sampling points. Performance of the FTG is especially unreliable for distributions with large tailings, where the sampling points can be distributed well outside the high probability region of the PDF (see Figure 3.2). The QMC gives the most accurate results on average. For problem

<table>
<thead>
<tr>
<th>Performance</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insatisfactory</td>
<td></td>
</tr>
<tr>
<td>Insatisfactory - Acceptable</td>
<td></td>
</tr>
<tr>
<td>Acceptable</td>
<td></td>
</tr>
<tr>
<td>Acceptable - Good</td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Color rating indicating the quality of performance for the sampling methods for different distributions
Table 3.2: Performance Evaluation of Methods for all distributions
settings of comparatively low complexity, the sampling approach of the QMC seldomly fails to capture the system behavior completely. The results for MCM and LHS only serve as an estimate. As multiple runs are needed respectively, it is possible to evaluate best and worst case scenarios. All methods generally lie within the confidence interval of the MCM for low orders of approximation. For higher orders, the confidence interval narrows and converges toward the reference solution.

Prediction of large or small CDF values (e.g. \( > 0.97 \) and \( < 0.03 \)) proves to be very difficult. While extreme values are sometimes approximated accurately without a good fit of the main probability region of the CDF (e.g. low CDF values for FTG in Figure 3.13 and high CDF values for FTG in Figure 3.14), it remains difficult to make certain estimates for which problem a scenario like this may occur. Without a reference, it is difficult to say whether the fluctuating side of the CDF converges towards the correct extreme CDF value with a higher degree of expansion and larger number of sampling points (see FTG in Figure 3.18), or if no convergence is evident (see FTG in Figures 3.13 and 3.14).

The OSC method performs unstable at times. It has been found that the OSC method requires weighting of the individual sampling points. As the other methods (MCM, QMC, LHS, FTG) were performed without weighting, the results of the unweighted OSC sampling points were examined for this study while the results of the weighted sampling points have been included in the submission for publication (see Section 2.2.6).
4 Numerical Study of a \( \text{CO}_2 \) Sequestration Setting

To test the methods employed (see Section 2.2), the problem setting of a \( \text{CO}_2 \) Sequestration site was investigated. \( \text{CO}_2 \) Sequestration is a topic of ongoing research, that is linked with uncertainty. Knowledge of subsurface geometries and hydraulic properties is often incomplete or highly uncertain. Polynomial chaos can help to examine the risks associated with individual storage sites. To implement PCE, the main parameters of interest have to be singled out. Then, the question of sampling the integration points out of the distributions of the parameters of interest for polynomial chaos arises. To study this question, the affect of different sampling strategies (see Section 2.2) on the simulation of the system depicted in Figure 4.1 is examined in this section.

To measure the performance of the model predictions, focus lay on the amount of \( \text{CO}_2 \) that leaked out of the system within a given time. The geometry of the system under consideration is depicted in Figure 4.1 and has been investigated by [Class et al., 2009] and [Oladyshkin et al., 2011].

The main hydraulic properties are listed in Table 4.1. The physical processes occurring in the system (Figure 4.1) can be described with mass balances for the two phases and the multiphase equation for Darcy’s Law. This gives the following coupled differential equations:

\[
\phi \frac{\partial S_w}{\partial t} - \nabla \cdot \left\{ \frac{k_{rw}}{\mu_w} \mathbf{K} \cdot (\nabla p - \varrho_w \mathbf{g}) \right\} - q_w = 0,
\]

\[
\phi \frac{\partial S_g}{\partial t} - \nabla \cdot \left\{ \frac{k_{rg}}{\mu_g} \mathbf{K} \cdot (\nabla p - \varrho_g \mathbf{g}) \right\} - q_g = 0,
\]

which are coupled by the relation

\[
S_w + S_g = 1. \tag{4.2}
\]

The subscripts \( w \) and \( g \) denote the brine (water) phase, and the \( \text{CO}_2 \)-rich (gas) phase, respectively. The variables \( S_g, S_w \) and pressure \( p \) denote the gas phase saturation, water phase saturation and the pressure. The relative permeabilities \( k_{rw} \) and \( k_{rg} \) are variables depending on \( S_w \) and \( S_g \) in linear fashion \((k_{rw} = S_w = 1 - S_g; k_{rg} = S_g)\). The gravity vector is denoted by \( \mathbf{g} \) and \( \mathbf{K} \) is the absolute permeability tensor for the geometry. The viscosity is denoted by \( \mu \). Finally, \( q_w, q_g \) are sources/sinks.

Figure 4.1 shows an underground storage site for \( \text{CO}_2 \) storage. \( \text{CO}_2 \) is injected into the
subsurface at the injection well. In the aquifer, the $CO_2$ spreads from the injection well to areas of lower saturation. Depending on the geometry of the system, the $CO_2$ saturation can reach regions in which the cap rock that sits on top of the $CO_2$ plume, allows leaking of the $CO_2$ to

Figure 4.1: Depiction of the scenario underlying the numerical test study investigated in this work.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$CO_2$ density, $\varrho_g$</td>
<td>479kg/m$^3$</td>
</tr>
<tr>
<td>Brine density, $\varrho_w$</td>
<td>1045kg/m$^3$</td>
</tr>
<tr>
<td>$CO_2$ viscosity, $\mu_g$</td>
<td>$3.950\times10^{-5}$Pa s</td>
</tr>
<tr>
<td>Brine viscosity, $\mu_w$</td>
<td>$2.535\times10^{-4}$Pa s</td>
</tr>
<tr>
<td>Aquifer permeability, $K_A$</td>
<td>$2\times10^{-14}$m$^2$</td>
</tr>
<tr>
<td>Aquifer thickness</td>
<td>30m</td>
</tr>
<tr>
<td>Aquitard thickness</td>
<td>100m</td>
</tr>
<tr>
<td>Leaky well permeability, $K_L$</td>
<td>$1\times10^{-12}$m$^2$</td>
</tr>
<tr>
<td>Porosity, $\phi$</td>
<td>0.15</td>
</tr>
<tr>
<td>Leaky &amp; injection well radius</td>
<td>0.15m</td>
</tr>
<tr>
<td>Injection rate</td>
<td>8.87kg/s (1600m$^3$/d)</td>
</tr>
<tr>
<td>Distance between wells</td>
<td>100m</td>
</tr>
<tr>
<td>Dimensions of model domain</td>
<td>1000m×1000m×160m</td>
</tr>
<tr>
<td>Simulation time, $t$</td>
<td>100 days</td>
</tr>
</tbody>
</table>

Table 4.1: Simulation parameters
the surface. This is inherently dangerous for the environment and jeopardizes the safe conduction of CO$_2$ storage. The case examined in this work studies leakage of CO$_2$ to the surface through an abandoned well, notated as “leaky well” in Figure 4.1. The leaked mass of CO$_2$ is the primary output variable investigated in this study.

The input parameters are constituted of three parameters. The data distributions representing the parameters were obtained by moment matching measured data from a CO$_2$ storage site. The moment matching leads to a smoothing of the input distributions. The variable input parameters for the model input are the permeability $K$, the porosity of the medium $\phi$ and the permeability of the leakage well. To obtain a reference, to which the results of the polynomial chaos expansion can be compared, a Monte Carlo simulation with 1260 points was performed. The results of these computations were used to perform aPC to obtain the best result achievable via polynomial basis with the given number of runs. Response surfaces were created by utilizing the arbitrary polynomial chaos expansion and sampling the integration points with FTG, MCM, QMC, LHS and PCM (see Section 2.2). Three different cases were studied to achieve comparable results and investigate specific points of interest. As numerical evaluation of the system is computationally expensive, a few select problems were investigated.

The simulation results were compared by investigating the amount of leaked CO$_2$ predicted by the numerical simulations. The results of the bootstrap comparison were investigated in form of the average of the 100 simulation runs (AVG) and their standard deviation (STD). The setup for the bootstrap simulations is summarized in Table 4.3.

4.1 Setup and Results of Numerical Simulations

4.1.1 Analysis of Degree of Expansion

Analogue to the investigation in Section 3.4, the influence of the degree of the expansion of the PCE method was studied. The two test cases include expansions of degree 2 and 3, as Chapter 3 showed that performance improvement for higher order expansions mostly does not outweigh the additional computational costs. The degree of expansion is constant for each case respectively, while the number of sampling points varies for the different sampling methods.

Case 1 - Order of expansion $d=2$ The order of the expansion $d$ was kept constant at 2 for all sampling methods. For Case 1a, all methods used the minimum number of sampling points $P$ (see Table 4.2). For Case 1b, the number of sampling points for QMC and LHS was raised to twice the minimum number, resulting in 20 sampling points (see Table 4.2). The results shown in Figure 4.2 depict the CO$_2$ leakage after 30 days. Utilizing the minimum
number of points for degree 2 (sampling points = 10) leads to a very inaccurate fit of the reference CDF by LHS (c.f. Figure 4.2). Only PCM leads to an accurate fit, while QMC achieves acceptable results. The average CDF of the two bootstrap analyses performed with 10 and 20 sampling points respectively are almost the same, while the smaller number of points leads to a larger standard deviation.

Increasing the number of points for the QMC and LHS improves their accuracy (see Figure 4.3).

Figure 4.2: Case 1a - CDF of CO$_2$ Leakage for Case 1 after 30 days with 10 sampling points for the QMC and LHS respectively.

Figure 4.3: Case 1b - CDF of CO$_2$ Leakage for Case 1 after 30 days with 20 sampling points for the QMC and LHS respectively.

Case 2 - Order of expansion d=3 This setup also compared the performance of all methods with the same order of the approximating polynomials d=3. Similar to Case 1, the minimum number of sampling points $P$ were used for Case 2a while the number of sampling points was doubled for QMC and LHS in Case 2b. A detailed description of the setup can be found in Table
In Case 2, polynomials of 3rd order were used to calculate the response surface of the system. The resulting CDF can be seen in Figures 4.4 and 4.5.

As for Case 1 (c.f. Figure 4.2), FTG sampling does not lead to satisfying results even though the number of sampling points, and thereby the computational costs, for FTG are over three times larger than for PCM.

![Figure 4.4: Case 2a - CDF of CO₂ Leakage for Case 2 after 30 days with 20 sampling points for the QMC and LHS respectively.](image)

![Figure 4.5: Case 2b - CDF of CO₂ Leakage for Case 2 after 30 days with 40 sampling points for the QMC and LHS respectively.](image)

### 4.1.2 Computational costs constant for all methods

**Case 3 - Computational costs constant for all methods**  
This setup allows varying orders of expansion. The purpose of this constellation is the importance of computational costs. The calculation of the output of the numerical model for the given input parameters is the most costly.
part of the construction of a response surface. Therefore, the total number of evaluations of a model often gives an upper limit to the order of the expansion. Therefore, the number of sampling points and the order of expansion was chosen as such to achieve a comparably similar number of sampling points. This led to a number of sampling points that varied between 56 (PCM), 60 (QMC and LHS) and 64 (FTG). The degree of the expansion was chosen in accordance to the highest degree of expansion possible with the number of points available. The number of points for PCM and FTG dictates the according degree of the expansion, while QMC and LHS are more flexible. Therefore, the number of sampling points was chosen to lie between PCM and FTG, which gives exactly 60 sampling points. This would allow any degree of expansion for QMC and LHS between 1 and 5. To obtain a balance between an overdetermined system and a high degree of expansion, QMC and LHS were approximated with a 4th order polynomial.

Case 3 attempts to compare all methods with the same computational effort. The only method achieving a good representation of the system behavior is the PCM. The results of QMC, LHS and FTG are not achieving a close fit. Similar to Case 1 and 2, the FTG shows a very steep CDF and only captures the mean of the CDF. It is important to point out that, besides FTG, all other sampling methods employ significantly more sampling points than in Cases 1 and 2. However, this does not improve accuracy for QMC and LHS. The means of the MCM bootstrapping simulations for 60 and 120 sampling points are significantly worse than in Case 1, where only 10 and 20 sampling points were employed, respectively. This result indicates that the ideal number of sampling points to employ might not increase proportional to the degree of expansion. Even 120 sampling points are not sufficient to achieve more reliable MCM response surfaces than with 10 points in Case 1.

Figure 4.6: Case 3 - CDF of CO₂ Leakage for Case 3 after 30 days with 20 sampling points for the QMC and LHS respectively.
4.2 Discussion

As the numerical test study requires significant computational effort, it is normally not technically feasible to construct a large number of independent Monte Carlo runs. Therefore, the variability of the results of the MCM was examined with a bootstrap method. Response surfaces were constructed with the minimum number of points required for a given case (see Section 4.1) and twice the minimum number of sampling points (see Table 4.2). This procedure was repeated 100 times to obtain an estimate of the uncertainty regarding MCM. This procedure keeps the number of sampling points small for each application, but allows testing of the variability of the results. Therefore, the bootstrapping enables an estimate on how sensitive the model is to the sampling of input data.

The results of the bootstrapping analysis show that MCM cannot be reliably used for a small number of sampling points. The standard deviation (c.f. Figure 4.7 - MCM BS STD) of the results shows large divergence and often exceeds 100% $CO_2$ leakage which is physically not feasible. Figure 4.7 shows the affect of the approximation order on the bootstrap results. Despite the increasing number of sampling points from Case 1 to Case 3, the results do not converge towards the reference. For later times (100 days), the CDF diverges even from the results which can be observed after 30 days. This stresses the importance of accurate sampling methods. Results which are inaccurate or highly uncertain after a small number of time steps can propagate.
Figure 4.7: CDF Bootstrap results of $CO_2$ Leakage for Case 1 after 30 and 100 days (top left and right), Case 2 after 30 and 100 days (middle left and right) and Case 3 after 30 and 100 days (bottom left and right).

this uncertainty through the simulation, making decision finding for large time steps very risky. Except for Case 1, the average obtained from MCM with the minimum number of sampling points required does not result in acceptable results.

The results for the $CO_2$ Leakage problem show that QMC and LHS require a higher amount of points than the minimum number $P$ (see Figures 4.2 to 4.6). This is yet more pronounced for LHS, as can be expected after the results of the test functions (c.f. Section 3.5). For problems
of comparatively low dimensionality (e.g. 3), LHS does not manage to fully utilize its advantage obtained form dividing each parameter space into strata. At dimensionality of this order (e.g. 3 parameters), spacing of sampling points with low discrepancy via QMC is achievable without large computational costs (c.f. Section 2.2.2). Variability and model dependency on well chosen sampling points is very pronounced, as can be seen in Figure 4.7. QMC and LHS perform significantly worse for Case 3, despite using up to 6 times as many points as in the Cases 1 and 2. This illustrate the consequences that a high order of expansion (degree 4 in Case 3) in combination with a comparatively small number of sampling points can have. This shows how crucial appropriate methods and a carefully chosen balance between the degree of the expansion and the number of sampling points is.
5 Discussion

**Selected Areas of Interest**  Complex, non-linear systems with uncertain input parameters demand a high level of expertise concerning the simulation methods employed. Therefore, this Section poses various questions that a modeler can use to characterize the system under investigation.

**How do distributions with strong tailing affect the performance of sampling approaches?**  Strong tailing can lead to larger moments of higher order. This affects the sampling points of methods based on polynomial bases (FTG and PCM) significantly more than methods that strongly focus on high probability regions of the PDF. Without weighting, this often leads to inaccurate results by the FTG that only capture the mean of the CDF.

**For complex systems, the minimum number of points (depending on the degrees of freedom) often does not lead to accurate results. Using more integration points poses the question whether weighting should be used to solve this overdetermined system or if simple regression should be employed?**  Weighting gives the individual sampling points different weight. This means that certain points will contribute more strongly to the response surface than others. For the newly developed OSC method, it has already been found that weighting of the individual sampling points is crucial for accurate representation of the response surface. For MCM, QMC and LHS, weighing was not investigated. However, the results of the CO$_2$ Leakage study give reason to believe that weighting could lead to more accurate results. The number of additional sampling points between Case 1a and 1b as well as Case 2a and 2b is only doubled, yet a significant improvement in prediction accuracy is obtained.

**Will approximations of higher order lead to more accurate results?**  A general answer can not be given to this question, as it strongly depends on the order of non-linearity of the problem under investigation. For most problems investigated, the additional costs past order 3 do not seem to pay off with significantly more accurate results. As higher-order polynomials can lead to more instability, this can cause results to diverge strongly from the reference CDF. Another important factor is the symmetry of the function that is to be approximated. An even function tends to be easier to approximate with a polynomial of an even order while an uneven function is easier to approximate with an uneven order of the approximating polynomial. Especially for the
FTG and PCM, where the sampling points are sampled from the roots of the polynomial base describing the sample data, the accuracy of the results can oscillate in relation to the order of the polynomials (see Section 3.5).

How difficult are the individual methods to implement? Does the additional effort of more complicated methods lead to better performance? This issue is of great importance, as implementation and accuracy of the methods do depend on the data available and the characteristics of the problem itself. For FTG and PCM, a large data sample is required to ensure efficient sampling points for moments of higher order. For small data sets, individual points contribute stronger to sampling based on moments, which is especially true for distributions like the log-Normal distribution. MCM, QMC and LHS draw their SP directly from the data points given by the original distribution. Therefore, these methods performance can suffer greatly if the available data is sparse and a small number of SP is employed, as this can lead to an inaccurate representation of the PDF by the SP.

Some methods are strongly based on the probability density functions (PDF) of the data distributions. What are possible affects What disadvantages pose methods that are strongly based on probability density (QMC, MCM)? Initially, it was a concern whether methods that mainly sample from high-probability regions might not perform strongly in regions of very high or low CDF values. However, no clear evidence was found that supports the finding that some methods might not accurately represent high probability regions (e.g., 0.1–0.9) while leading to reliable results for extreme CDF values. Methods that lead to accurate extreme values (e.g. < 0.1 or > 0.9) but do not represent high probability regions accurately raise the question whether accurate extreme value prediction was due to efficient sampling of integration points or chance. Once a method is unreliable in high probability regions, it is hard to determine how much certainty can be associated with results for low probability regions.

Is the evaluation of performance based on mean and variance sufficiently accurate? What can be said of very high/low values of CDF curves? The FTG often predicts very accurate results for the mean of the response surface CDF without accomplishing an accurate representation of high or low CDF value regions (see Figures 3.12 and 3.14). While capturing the mean of the CDF without capturing low and high CDF values (see low order approximations in Figure 3.5) often occurs, a high uncertainty is connected to such results, which can be misleading (see FTG in Figures 3.13, 3.14 and 3.16).
6 Conclusions

This thesis examines modeling of highly non-linear systems with uncertain input parameters, which poses many challenges. The theoretical analysis (c.f. Section 3) and practical study (c.f. Section 4) lead to some conclusions, which are summarized below.

The order of the polynomial chaos expansion tend to converge more strongly for expansions up to 3rd order. For higher order approximations, instabilities can be caused by a small number of sampling points in comparison to the degrees of freedom. For expansions of higher order (e.g., higher than 3), increasing the number of sampling points without increasing the order of expansion generally leads to more stable result. Note, that the PCM can also sample more points than the minimum number of sampling points \( P \) without utilizing a higher order approximation. This can be achieved by using higher order moments, solely for the calculation of sampling points.

The mean and standard deviation can be misleadig if used for comparison of simulation results or convergence analysis. This is especially true for risk analysis, where extreme system responses that occur with a low probability are of interest.

Sacrificing accuracy in the main probability regions (e.g., 0.1 – 0.9) and focus on good estimates in low probability regions is not feasible, as it is hard to estimate whether these approaches will lead to accurate estimates in both low probability regions (e.g., < 0.1 and > 0.9) or only seemingly converge for one of these regions. Low probability regions in general are harder to approximate than, e.g., the mean. This can be observed in Section 3.5.

Monte Carlo and Latin Hypercube Sampling perform too uncertain for small numbers of sampling points. However, as they both allow examination of the impact, that the choice of sampling points can have on system performance for a given problem, MCM and LHS can be employed to make worst or best case scenario estimates.

Quasi-Monte Carlo Sampling and the Probabilistic Collocation Method were found to produce the most reliable results while employing a comparatively small number of sampling points.

Distributions with strong tailing require specific care, especially for systems with strong non-linearity (e.g., exponential behavior).

Using a small number of sampling points (a number of sampling points close to the minimal value \( P \), that is), is more prone to lead to strongly diverging results for higher orders. This is due to polynomials oscillating more strongly near the bounds of the parameter value interval for higher orders. Inaccurate results can then diverge faster.

Time dependency is a great problem for risk assessment. While estimates for early times might be
accurate, results for later times can diverge from the actual result. However, knowing for which time scales the results are accurate is difficult, unless results start to give unphysical estimates. It is important to stress the fact that increasing of the degree of the expansion and or the sampling points does not guarantee accurate representation of system behavior. Understanding the method employed as well as the underlying input parameters and sampling methods is essential for efficient sampling and satisfying results. As Section 4.1 shows, acceptable results can be obtained for a small number of sampling points if the chosen setup is suited for the system behavior.

For future work, a point of great interest are distributions. While this work has focused on beta, normal, log-normal and uniform distribution, there are significantly more complex distributions (multivariat, bimodal and others) which can be readily found in natural settings and applications.

Another research area of great significance is the influence of the dimensionality of the input parameters. While this study focused on three dimensional input parameter configurations, practical application may require much higher dimensionality to obtain uncertainty estimates for all input parameters.

Investigating possible combinations of high probability region sampling (e.g., QMC) with weighted sampling points poses an interesting area of further research. Especially for small number of sampling points, weighting could lessen the impact of points that lead to instable response surface results.
Bibliography


## Nomenclature

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha^i_j )</td>
<td>Multivariate index</td>
</tr>
<tr>
<td>( \alpha_{BD} )</td>
<td>Shape parameter of the beta distribution</td>
</tr>
<tr>
<td>( \beta_{BD} )</td>
<td>Shape parameter of the beta distribution</td>
</tr>
<tr>
<td>( \delta_{kl} )</td>
<td>Kronecker delta</td>
</tr>
<tr>
<td>( \mathcal{N} )</td>
<td>Number of strata in each parameter for LHS</td>
</tr>
<tr>
<td>( \omega )</td>
<td>Parameter</td>
</tr>
<tr>
<td>( \Phi_i )</td>
<td>Function that combines the polynomials of different parameters</td>
</tr>
<tr>
<td>( r_{w} )</td>
<td>Relative permeability for gas phase</td>
</tr>
<tr>
<td>( r_{w} )</td>
<td>Relative permeability for water phase</td>
</tr>
<tr>
<td>( c_i )</td>
<td>Expansion coefficients of the PCE</td>
</tr>
<tr>
<td>( d )</td>
<td>Order of the PCE</td>
</tr>
<tr>
<td>( f(\omega) )</td>
<td>Model Output</td>
</tr>
<tr>
<td>( g )</td>
<td>Gas phase</td>
</tr>
<tr>
<td>( g )</td>
<td>Gravity vector</td>
</tr>
<tr>
<td>( M )</td>
<td>Combinatoric coefficient, calculated from ( d ) and ( N )</td>
</tr>
<tr>
<td>( \mathbf{K} )</td>
<td>Absolute permeability tensor</td>
</tr>
<tr>
<td>( N )</td>
<td>Number of different input parameters</td>
</tr>
<tr>
<td>( p )</td>
<td>Pressure</td>
</tr>
<tr>
<td>( P^{(i)} )</td>
<td>Polynomial basis for expansion of order ( d )</td>
</tr>
<tr>
<td>( \beta_{(k)} )</td>
<td>Coefficients of the polynomial basis</td>
</tr>
<tr>
<td>( q_g )</td>
<td>Source/sink</td>
</tr>
<tr>
<td>( q_w )</td>
<td>Source/sink</td>
</tr>
<tr>
<td>( S_g )</td>
<td>Gas-phase saturation</td>
</tr>
<tr>
<td>( S_w )</td>
<td>Water-phase saturation</td>
</tr>
<tr>
<td>( w )</td>
<td>Brine (water) phase</td>
</tr>
</tbody>
</table>
Figure 1: Comparison of the CDF of all sampling methods for varying approximation order for the trigonometric function with a Beta distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 2: Comparison of the CDF of all sampling methods for varying approximation order for the exponential function with a Beta distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 3: Comparison of the CDF of all sampling methods for varying approximation order for the 5th order polynomial function with a Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 4: Comparison of the CDF of all sampling methods for varying approximation order for the trigonometric function with a Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 5: Comparison of the CDF of all sampling methods for varying approximation order for the runge function with a Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 6: Comparison of the CDF of all sampling methods for varying approximation order for the square root function with a Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 7: Comparison of the CDF of all sampling methods for varying approximation order for the trigonometric function with a log-Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 8: Comparison of the CDF of all sampling methods for varying approximation order for the exponential function with a log-Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 9: Comparison of the CDF of all sampling methods for varying approximation order for the square root function with a log-Normal distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 10: Comparison of the CDF of all sampling methods for varying approximation order for the runge function with a Uniform distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Figure 11: Comparison of the CDF of all sampling methods for varying approximation order for the square root function with a Uniform distribution. MCM AVG and LHS AVG denote averaged CDFs over all individual MCM and LHS runs. MCM SIG denotes the 0.05 and 0.95 value of the total runs of MCM.
Published Preprints

http://www.nupus.uni-stuttgart.de

2007/1  Cao, Y. / Eikemo, B. / Helmig, R.: Fractional flow formulation for two-phase flow in porous media

2008/1  Helmig, R. / Weiss, A. / Wohlmuth, B.: Variational inequalities for modeling flow in heterogeneous porous media with entry pressure

2008/2  Cao, Y. / Helmig, R. / Wohlmuth, B.: Convergence study and comparison of the multipoint flux approximation L-method

2008/3  van Duijn, C.J. / Pop, I.S. / Niessner, J. / Hassanizadeh, S.M.: Philip’s redistribution problem revisited: the role of fluid-fluid interfacial areas


2008/6  Cao, Y. / Helmig, R. / Wohlmuth, B.: Geometrical interpretation of the multipoint flux approximation L-method

2008/7  Vervoort, R.W. / van der Zee, S.E.A.T.M.: Simulating the effect of capillary flux on the soil water balance in a stochastic ecohydrological framework


2008/10 Wolff, M.: Comparison of mathematical and numerical models for two-phase flow in porous media


2008/12 Cao, Y. / Helmig, R. / Wohlmuth, B.: Convergence of the multipoint flux approximation L-method for homogeneous media on uniform grids

2008/13 Ochs, S.O.: Development of a multiphase multicomponent model for PEMFC

2008/14 Walter, L.: Towards a model concept for coupling porous gas diffusion layer and gas distributor in PEM fuel cells

2009/1  Hægland, H. / Asstreerawatt, A. / Helmig, R. / Dahle, H.K.: Streamline approach for a discrete fracture-matrix system


2009/3  Heimann, F.: An unfitted discontinuous Galerkin method for two-phase flow

2009/4 Hilfer, R. / Doster, F.: Percolation as a basic concept for macroscopic capillarity
2009/5 van Noorden, T.L. / Pop, I.S. / Ebigbo, A. / Helmig, R.: An effective model for biofilm growth in a thin strip

2009/6 Baber, K.: Modeling the transfer of therapeutic agents from the vascular space to the tissue compartment (a continuum approach)

2009/7 Faigle, B.: Two-phase flow modeling in porous media with kinetic interphase mass transfer processes in fractures

2009/8 Fritz, J. / Flemisch, B. / Helmig, R.: Multiphysics modeling of advection-dominated two-phase compositional flow in porous media

2009/9 Støverud, K.: Modeling convection-enhanced delivery into brain tissue using information from magnetic resonance imaging


2010/1 Rosenbrand, E.: Modelling biofilm distribution and its effect on two-phase flow in porous media

2010/2 Schöniger, A.: Parameter estimation by ensemble Kalman filters with transformed data


2011/1 Linders, B.: Experimental investigations on horizontal redistribution

2011/2 Rau, M.T.: Geostatistical analysis of three-dimensional hydraulic conductivity fields by means of maximum Gauss copula

2011/3 Kraus, D.: Two phase flow in homogeneous porous media - The role of dynamic capillary pressure in modeling gravity driven fingering

2011/4 Brugman, R.: Dimensionless analysis of convection enhanced drug delivery to brain tissues

2011/5 Sinsbeck, M.: Adaptive grid refinement for two-phase flow in porous media


2012/1 Köppl, T. / Wohlmuth, B. / Helmig, R.: Reduced one-dimensional modelling and numerical simulation for mass transport in fluids

2012/2 Kumar, K. / Pop, I.S. / Radu, F.A.: Convergence analysis for a conformal discretization of a model for precipitation and dissolution in porous media

2012/3 Hommel, J.: Modelling biofilm induced calcite precipitation and its effect on two phase flow in porous media
2012/4 Estrella, D.: Experimental and numerical approximation methods for zero-valent iron transport around injection wells

2012/5 Heimhuber, R.: Efficient history matching for reduced reservoir models with PCE-based bootstrap filters

2012/6 Kissling, F. / Karlsen, K.H.: On the singular limit of a two-phase flow equation with heterogeneities and dynamic capillary pressure

2012/7 Fritz, S.: Experimental investigations of water infiltration into unsaturated soil - Analysis of dynamic capillarity effects

2012/8 Strohmer, V.: Numerische Analysis von nahezu parallelen Strömungen in porösen Medien

2012/9 Kissling, F. / Rohde, C.: The computation of nonclassical shock waves in porous media with a heterogeneous multiscale method: The multidimensional case

2012/10 Fetzer, T.: Numerical analysis of the influence of turbulence on exchange processes between porous-medium and free flow

2012/11 Schröder, P.: A response surface bootstrap filter to calibrate CO₂ injection models


2013/2 Köppel, M.: Flow modelling of coupled fracture-matrix porous media systems with a two mesh concept

2013/3 van Helvoort, M.: Upscaling of processes involving rough boundaries

2013/4 Redeker, M. / Haasdonk, B.: A POD-EIM reduced two-scale model for crystal growth

2013/5 Vogler, D.: A comparison of different model reduction techniques for model calibration and risk assessment