Description of the master's thesis

Many 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 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 [Oladyshkin et al., 2012, Oladyshkin 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 (see Figure 1).

This work gives 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, selected 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 that requires numerical simulation was chosen and evaluated.

This thesis examines modeling of highly non-linear systems with uncertain input parameters, which poses many challenges. The theoretical analysis and practical study lead to some conclusions, which are summarized briefly. The order of the polynomial chaos expansion tends 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 results. 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 misleading 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., 0.01-0.09) 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 by means of selected test functions. Monte Carlo and Latin Hypercube Sampling perform too uncertainly 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. This work shows that 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 (multivariate, 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.

![Figure 2: Examples of CDF for CO₂ sequestration risk assessment.](image)

References

