Uncertainty Quantification of Radionuclide Release Models using Non-Intrusive Polynomial Chaos.

Casper Hoogwerf

Foreword

This report presents the final thesis of the Master of Science programme in Applied physics at the Delft University of Technology. The project has been performed at the section of Physics of Nuclear Reactors of the faculty of Applied Sciences.

I would like to thank Jan Leen Kloosterman, Danny Lathouwers and Luca Gilli for their guidance, answering a lot of my questions and their time they invested in me during the project.
Also I would like to thank Thomas Schröder from NRG, whom did spend a lot of his time on checking, running calculation, post-processing data and for the extensive feedback on my work.
Furthermore I would like thank to Erik van der Linden with whom I had various pleasant discussions about various topics. I also would like to thank all the members of the PNR section, my closest friends, family and fellow students for all kinds of support the have given me along the journey I had to take to finish this project.

Abstract

Performance assessments of deep geological waste repositories need to be carried out in order to demonstrate the safety of a particular repository site. Radionuclide transport models are usually used for this purpose to mimic the transport of radionuclide's through the biosphere.

Uncertainty is an intrinsic part in these radionuclide transport models and in the presence of these uncertainties probabilistic safety assessments will need to be carried out in order to calculate an expected risk to the human population.
Commonly this is done by applying techniques such as Monte Carlo simulation or perturbation methods. Unfortunately for complex models (as these transport models) the computational costs when using Monte Carlo can be very high whereas perturbation methods are less reliable when there are large uncertainties involved.

In this work we present the application of a spectral technique based on a Polynomial Chaos Expansion (PCE) in order to quantify the uncertainties in stochastic radionuclide transport codes. The method is applied non-intrusively which means that the transport codes are sampled as a black-box in a deterministic set of points and sparse grids were used to reduce the number of deterministic sample points.
Firstly, the method is applied to an artificial compartment model for radionuclide release into the biosphere in which the retardation factor is considered to be uncertain. Secondly, PCE is applied to the *EMOS* code maintained by NRG. This code models the physics of a repository design in a salt rock formation and uncertainties in parameters modeling the salt plug, which is an engineered barrier, were considered.
In both cases the results obtained with PCE were compared with the results obtained with Monte Carlo simulations.

For the compartment model is shown that the Monte Carlo technique is roughly 2 to 4 times faster than the PCE technique when one is only interested in numerical results such as a mean value and variance of a certain quantity. If one is interested in a almost noise free pdf of this quantify then PCE is approximately 25 to 55 times faster.
For the *EMOS* case is it shown that with the PCE technique the central tendencies and statistical estimators are within 1% agreement of the Monte Carlo results and produces an almost noise free pdf and is approximately 12 times faster than Monte Carlo.

List of Figures

3.1 Two dimensional cubature grids. 26
3.2 Two dimensional sparse grids. 27
4.1 The radionuclide concentration in compartment 4 and 31 as a function of time. 32
4.2 pdf's of the first and fourth radionuclide in compartment 4 at 100 years. 35
4.3 pdf's of the first radionuclide in compartment 4 at 100 years, for different sparse grids. . . . 38
4.4 pdf's of the sixth radionuclide in compartment 4 at 100 years, for different sparse grids. . . . 39
5.1 pdf and cdf of the maximum dose rate based on 1000 realizations. 44
5.2 pdf's of the maximum dose rate computed with different sparse grids. 48
A.1 The pdf's of all radionuclide in compartment 4 at 100 years. 56
A.2 The pdf's of all the radionuclide in compartment 31 at 100 years. 57
B.1 pdf's of the first radionuclide in compartment 4 at 100 years. 62
B.2 pdf's of the second radionuclide in compartment 4 at 100 years. 63
B.3 pdf's of the third radionuclide in compartment 4 at 100 years. 64
B.4 pdf's of the fourth radionuclide in compartment 4 at 100 years. 65
B.5 pdf's of the fifth radionuclide in compartment 4 at 100 years. 66
B.6 pdf's of the sixth radionuclide in compartment 4 at 100 years. 67
B.7 pdf's of the seventh radionuclide in compartment 4 at 100 years. 68
B.8 pdf's of the eighth radionuclide in compartment 4 at 100 years. 69
C.1 cdf's calculated for different truncation order and different grids. 70

List of Tables

3.1 Families of probability laws and corresponding families of orthogonal polynomials 17
3.2 2 Dimensional shifted-Legendre polynomials of maxmum order 2 27
4.1 Statistics of the concentrations of the first and sixth radionuclides in compartment 4 at 100 years. 34
4.2 Number of nodes and cputime for given quadrature rules 36
4.3 Statistics of the first and sixth radionuclide in compartment 4 at time 100 years. 37
5.1 Distribution and values of the 6 random variables used 42
5.2 The different measures of the dose rate in the biosphere calculated by Monte Carlo sampling for 1000 realizations. Taken at the dose rate peak-time. Results have been taken from [1] . . 43
5.3 Statistics of the maximum dose rate in the Biosphere for the wet scenario, with various numerical integration techniques. 46
5.4 Quantiles for the maximum dose rate in the Biosphere for the wet scenario, with various numerical integration techniques. 47
A.1 MC results for compartment 4 54
A.2 MC results for compartment 4 54
A.3 MC results for compartment 31 55
A.4 MC results for compartment 31 55
B.1 Statistics of compartment 4 58
B.2 Statistics of compartment 4 59
B.3 Statistics of compartment 31 59
B.5 Statistics of compartment 4 59
B.4 Statistics of compartment 31 60
B.6 Statistics of compartment 4 60
B.7 Statistics of compartment 31 60
B.8 Statistics of compartment 31 61

Nomenclature

C_k	Molar concentration in water of radionuclide k at position $\mathbf{r}$ and time t	[Moles·m^{-3}]
$C_{i,n}^k(t)$	PC coefficients	[-]
$D(\mathbf{r},t)$	Dispersion coefficient of water at position $\mathbf{r}$ and time t	[m^2·y^{-1}]
F_k	Adsorbed molar concentration on solids of radionuclide k at position $\mathbf{r}$ and time t	[Moles·m^{-3}]
$K_{d,k}$	Distribution coefficient of the kth radionuclide	[m^3·kg^{-1}]
N_c	Number of compartments	[-]
N_r	Number of radionuclides in the decay chain	[-]
$P_i()$	i^{th} Polynomial	[-]
$Q_l^{(1)}$	one-dimensional quadrature rule, in direction l	[-]
R_k	Retardation coefficient of the kth radionuclide	[-]
$S_k(\mathbf{r},t)$	External source of radionuclide k at position $\mathbf{r}$ and time t	[Moles·y^{-1}]
U	Random Process	[-]
V_i	Volume of compartment i	[m^3]
$\dot{D}_B$	Dose rate in the Biosphere	[Sv·Year^{-1}]
$\mathbf{U}(\mathbf{r},t)$	Advective flow of water at position $\mathbf{r}$ and time t	[m·y^{-1}]
$a_{i,j}$	Rate transfer coefficient into and out compartment i over a boundary which is shared with compartment j	[y^{-1}]
c_i	PC coefficient of the ith Polynomial	[-]
$f(F,C)$	Adsoprtion isotherm	[Moles·m^{-3}·y^{-1}]
k	Index for radionuclide species k, $k = 1,2,\cdots,N_r$	[-]
$r_{i,j}$	Rate transfer coefficient into and out compartment i and compartment j	[y^{-1}]
Θ	Sample space	[-]

$\bar{\theta}$	Stochastic process with known distribution depending on ξ	[-]
ε	porosity of rock	[-]
λ_k	Decay constant of radionuclide k	[y^{-1}]

Contents

1 Introduction **10**

2 Uncertainty Quantification **12**
2.1 What is UQ . . . 12
2.2 Techniques for UQ . . . 13
2.2.1 Monte-Carlo . . . 14
2.2.2 Spectral Methods . . . 15

3 Polynomial Chaos **16**
3.1 Introduction . . . 16
3.2 Orthogonal Polynomials of the Askey scheme . . . 16
3.3 Properties of orthognal Polynomials . . . 17
3.4 Definitions and Notations . . . 17
3.4.1 One Dimensional PC Basis . . . 20
3.4.2 Multidimensional PC Basis . . . 20
3.5 Truncation of the PC Expansion . . . 21
3.6 First- and Second-Order Estimates . . . 21
3.7 Spectral Projection . . . 22
3.8 Non-Intrusive Spectral Projection . . . 23
3.8.1 Orthogonal Projection . . . 23
3.8.2 Monte Carlo Integration . . . 24
3.8.3 Deterministic Integration . . . 24
3.8.3.1 Gauss Quadrature . . . 25
3.8.3.2 Tensor Product Formulas . . . 25
3.8.3.3 Sparse Grids . . . 26

4 Compartment Model for radionuclide release into the Biosphere **29**
4.1 Introduction . . . 29
4.2 The Transport Equation . . . 29
4.3 Derivation of the Compartment Model . . . 30
4.4 Solving the Compartment Model . . . 32
4.5 Uncertainties in the Compartment Model . . . 32
4.6 UQ of the Compartment Model with Monte Carlo . . . 34
4.6.1 Results of UQ with Monte Carlo . . . 34
4.7 UQ of the Compartment Model with Polynomial Chaos . . . 35
4.8 Conclusions . . . 40

5 Analysis of a repository design in a rock salt formation **41**
5.1 Uncertainties in model parameters . . . 41
5.2 Standard Techniques used for Uncertainty Quantification . . . 42
5.3 Polynomial Chaos Applied to the EMOS code for UQ . . . 44

5.3.1 Choice of Polynomials, Truncation Order and Integration Grid 45
5.3.2 UQ Results obtained by applying PC 46
5.4 Conclusion 50

6 Discussion and Recommendations 51

A Monte Carlo 53
A.1 Implementation of the Monte Carlo method 53
A.2 Results 53

B Non Intrusive Spectral Projection Applied to the Compartment model 57
B.1 Implementation 57
B.2 Results 58

C Pamina Results 70
C.1 cdf's of the maximum dose rate 70

1 Introduction

Radioactive or nuclear waste can be described as a material containing concentrations of radioactive isotopes which are higher than the minimum measurable concentration and for which there are no known applications available. A well know producer of nuclear waste are nuclear reactors, but also medical applications such as radiotherapy and nuclear medicine produce large quantities of radioactive waste. An other producer of nuclear waste is the oil and gas industry where the waste is produced during the mining process. This is due to the fact that radioisotopes are (naturally) mixed with the product to be mined, this is also known as a Natural Occurring Radioactive Material (NORM).
Nuclear waste can be categorized into;

- Low and intermediate level waste, of which the measurable activity is higher than the background activity and produces no more than 2 kW/m^3 of radiative heat. This group can be divided into two subgroups;
  - short lived waste with an averaged maximum activity of 400 Bq/g.
  - long lived waste with an activity exceeding the averaged maximum of 400 Bg/g for the short lived waste.
- High level waste, which mostly consists of long lived isotopes and produce more than 2 kW/m^3 of radiative heat.

In the Netherlands, COVRA (central organisation for radioactive waste) is responsible for the management of radioactive waste. The owners or producers of nuclear waste do not only pay COVRA for processing and storage of nuclear waste, but also for research into long-term solutions for handling, processing and storing of the nuclear waste.
Nowadays, nuclear waste is stored in bunkers owned by COVRA. These bunkers are situated at ground level and are designed to prevent leakage of radionuclides into the geosphere and biosphere under various threat conditions such as floodings, storms and air-plane crashes. COVRA expects that is has a sufficient storage capacity to store all nuclear waste produced within the coming one hundred years, before a long-term solution has been found.
One of the possible solutions is a underground waste repository situated in a salt rock or clay formation. Those types of rock or soil are know to be plastic which means that cracks (due to for example construction of the repository) in these formations are self-healed. Furthermore there is no, to almost no transport of fluids in those formations due to the fact that there are no cracks in the formations due to the self-healing effect. This way a very stable and robust natural barrier to encapsulate nuclear waste is obtained. So if the engineered barriers have failed, there is a large natural barrier left in which there is almost no transport of radionuclides possible.

To define which locations are suitable for such a repository, a performance assessment of all possible repositories including site-specific natural barriers needs to be carried out. In this way a quantification and a demonstration of the safety of the total repository (engineered and natural barriers) performance can be specified. Such an assessment can be carried out by modelling the properties of the repository with a mathe-

matical model which describes the transport of radionuclides through the barriers into the biosphere.

In order to perform a site-specific calculation, site-specific model parameter values are necessary. These parameter values are usually obtained through measurements, and limitations during these measurements can cause difference between the measured value and the actual value. The properties of the quantity are said to be uncertain and in general this is the case for every measurable quantity. Uncertain quantifies are usually expressed by a best estimate and an associated range. The influence of the uncertain properties on the final performance assessment can be estimated with the use of Uncertainty Quantification (UQ).
There are various methods available to perform UQ such as *Monte Carlo* and *Perturbation techniques*. A relatively new method used for UQ is *Polynomial Chaos expansion* and it has already been applied to applications including stochastic fluid dynamics, stochastic finite elements, solid mechanics, nonlinear estimation, and probabilistic robust control. In the presented work the implementation of this method on performance assessment codes of repository designs is the main topic of this thesis.

In Chapter 2 we introduce the notion of Uncertainty Quantification and briefly discuss the techniques which can be used for this. We continue in Chapter 3 with the introduction of Polynomial Chaos and how the technique can be used to find a solution for stochastic models. A stochastic model for radionuclide release into the biosphere is introduced in Chapter 4 and a Polynomial Chaos expansion is compared with Monte Carlo approach to quantify the uncertainty related to this model. In Chapter 5 we apply a Polynomial Chaos expansion for the analysis of a repository design in a salt rock formation and compare the results with the results of a Monte Carlo approach.

2 Uncertainty Quantification

2.1 What is UQ

Throughout the research and design process numerical modelling and simulations are essential tools for scientists and engineers, especially when dealing with complex systems.
The reliability of a simulation is a main concern when interpreting and analysing its results. In order to yield reliable information regarding the system under investigation, these numerical simulations have to be carefully designed, performed and verified. Nevertheless, differences between simulations and the actual real world system can occur. It is therefore critical to understand what the origin of these differences are, if one wants to quantify the reliability of the simulation. In general the errors in the simulations, which lead to discrepancies between model predictions and the actual system can be grouped into; 1. Model errors, 2. Numerical errors and 3. Data errors.

1. Model errors
   Simulations rely on the fact that the process or system under investigation can be cast in a mathematical model. Simplifications are often made in such a way that the mathematical model only consists of the essential physical principles and characteristics of the system. This means that physical phenomena which may occur in the real process or system are left out of the mathematical model because of their negligible contribution to the process or system. Therefore, it is impossible that the behaviour of the real system will be exactly reproduced by the mathematical model.
   In this case, the validity of the assumptions made have to be verified beforehand to make sure that the predictions will remain sufficiently accurate for interpretation and analysis.

2. Numerical errors
   As discussed before, a simulation consists of a mathematical model, which has to be solved by using a numerical method. The numerical method to be chosen usually depends on the type (differential, integral, algebraic,...) of equations to be solved. Furthermore, one needs to keep in mind in what kind of phenomena one is interested and the length-scales associated to these phenomena. The resolution with which the problem is solved should be such that the smallest length-scales of interest can be captured. In order the meet this resolution criteria an appropriate discretization method should be chosen.
   Theoretically, the numerical error can be reduced to any arbitrary (low) level by using finer discretizations and more computational resources. However, in general the numerical error will always be nonzero because of the finite representation of numbers in computers.

3. Data errors
   Usually a large set of data is necessary to run a simulation. This data is used for example to define the system's geometry, boundary and initial conditions, where parameters are used for physical or model constants. Most of the data is specified through experiments. And since a quantity can not be measured with infinite precision, an error is induced on the measured quantity. Also limitation in experiments can affect the precision so that the quantity is only known in a (broad) range. And so using data, which not fully covers the physics of the real system, also will induces simulation errors.

The present work considers only errors associated with variabilities in the data and methods to analyse the impact of these variabilities on the model predictions are considered. Such an analysis can be done for the purpose of, e.g.; (1) Validation, (2) Variance Analysis, (3) Risk Analysis and (4) Uncertainty Management.

1. Validation
   A simulation must be validated, this can be done by comparing simulation results with results obtained by measurements performed on the real system. One has to keep in mind that there are also uncertainties affecting the measurement. So during validation, simulation errors as well as measurement errors have to be considered.
2. Variance Analysis
   Usually the best-estimate of a certain quantity will be calculated during a simulation run. With the use of variance analysis it is possible to specify a confidence measure in this best-estimate and the controllability of the system.
3. Risk Analysis
   It is often desired to conduct reliability or risk assessment analysis. For such an analysis one needs to define probabilities that the system will exceed certain critical values or thresholds.
4. Uncertainty Management
   In order to effectively observe, manage and ultimately reduce sources of uncertainty, it is necessary to quantify the relative impact of these sources on the response of the system.

2.2 Techniques for UQ

Suppose a mass is attached to a spring which exerts an attractive force on the mass proportional to the extension or compression of the spring. We are interested in the extension of the spring over time and using Newton's second law we can write an equations which models our problem and ignore friction and gravity.

$$m\frac{\partial^2 x(t)}{\partial t^2} + kx(t) = 0 \tag{2.1}$$

In Eq. 2.1 the extension over time is expressed with $x(t)$, the mass attached to the spring by m and the spring constant by k.

The model displayed in Eq.2.1 is a second order linear differential equation. For this particular type of differential equation a analytical solution can be found, however this is not the case for every differential equation.
In Eq. 2.1 is is assumed that the extension of the spring is only depending on the mass attached to it and the spring constant and that these quantities are exactly known. In real-life physics this is never the case, take for example the spring constant. During manufacturing of this spring there could be a number of things

which have a influence on the spring constant, or perhaps the temperature of the spring has an effect on the spring constant. We can thus assume that the spring constant is uncertain and usually is expressed with a probability law or a range.
Since the the spring constant is dependent on for example manufacturing errors we can say that the spring constant is depending on some random variables (RV's) $\bar{\xi} = \{\xi_1, \xi_2, \xi_3, \cdots\}$ and thus $k \equiv k\left(\bar{\xi}\right)$. Now we would like to know how the extension of the spring is influenced by these dependencies, in other words what is the influence of $\bar{\xi}$ on $x(t)$ (or $x(t, \bar{\xi})$). For this we can now right the following,

$$m\frac{\partial^2 x(t,\bar{\xi})}{\partial t^2} + k(\bar{\xi})x(t,\bar{\xi}) = 0. \tag{2.2}$$

Eq. 2.2 is a stochastic differential equation and in general a analytical solution of such an equation is hard to find, and instead of analytical solutions numerical approximations are used.
If the probability laws of $k(\bar{\xi})$ are known, then it is possible to find the probability law of $x\left(\bar{\xi}\right)$. This can be done for example by using a Monte-Carlo approach or by using a Spectral Method.

2.2.1 Monte-Carlo

The Monte-Carlo method relies on a random sampling of the RV's in order to construct a sample of independent input data, this can be written as: $k\left(\bar{\xi}\right) = k\left(\xi_1, \xi_2, \xi_3, \cdots\right)$. Consequently, multiple of these input data samples can be constructed such that a set of input samples can be defined: $\{k_1, k_2, k_3, \cdots\}$. For each individual data sample the system will be solved, giving a set of unique solutions $\{x_1, x_2, x_3, \cdots\}$. This collection is called the sample solution set. The statistics of this sample solution set can be used to estimate for example the mathematical expectation, the variance or a probability law of a particular observable.
The mathematical expectation for example, can be calculated according to:

$$\langle x \rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{2.3}$$

Usually it is impossible to sample an infinite amount of times and instead a sufficiently large number of samples (N) is used. Eq. 2.3 now becomes,

$$\langle x \rangle \approx \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{2.4}$$

The variance in x can also be calculated,

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} \left(x_i - \langle x \rangle\right)^2. \tag{2.5}$$

Furthermore is is also possible to estimate the accuracy of the approximation of the mathematical expectation by calculating the standard deviation of the mean,

$$\sigma_{\langle x \rangle} = \frac{\sigma}{\sqrt{N}}. \tag{2.6}$$

Based on the standard deviation of the mean we can determine if the number of samples is chosen sufficiently large enough.
In Eq.2.6 we see that relative accuracy in $\langle x \rangle$ scales with $\frac{1}{\sqrt{N}}$. So if one want to determine $\langle x \rangle$ with a relative accuracy of 1% than N should be 10.000 or larger. This low convergence rate is the main drawback of Monte Carlo, since the computation time scales with N.
The main advantages of the method is that it is quit straightforward to use and it can be used for both stochastic problems, which involve probability, and deterministic problems which are without probability.

2.2.2 Spectral Methods

Spectral Methods are based on the construction of the functional dependence of the solution on the RV's. Typically this functional is expressed in terms of a series:

$$x\left(\bar{\xi}\right) = \sum_{i=0}^{\infty} c_i \Psi_i \left(\bar{\xi}\right). \tag{2.7}$$

Here the functionals depending on the RV's are represented by Ψ_i and the deterministic coefficients by c_i. The functionals should be selected beforehand according to the problem, whereas the deterministic coefficients need to be found by solving a mathematical problem.
When the coefficients are available, it is possible to determine the statistics of x analytically or by sampling the RV's.
A Polynomial Chaos Expansion is an example of a spectral expansion where an expansion is used to represent a process in a stochastic-space. It is similar to a Fourier Expansion where a expansion is usually used to represent a process in a frequency-space.
In the following section the polynomial chaos expansion will be introduced and a method is given how to find the deterministic coefficients.

3 Polynomial Chaos

3.1 Introduction

The polynomial chaos technique originates from the *homogeneous chaos*, which has been defined by Wiener [2]. This *homogeneous chaos* was introduced to describe a Gaussian process by Hermite polynomial functionals. This functional describing a Gaussian process can be written in a Fourier-Hermite series. For any process with finite second order moments, the expansion will converges to this process. And so, a stochastic process can be represented with an expansion of orthogonal Hermite polynomials.
The *generalized Polynomial Chaos* (gPC) was proposed in [3] to supply an optimal basis, in terms of convergence for stochastic processes represented by random variables of commonly used distributions. For this the Askey scheme was used [4], and within this light the *homogeneous chaos* can be seen as a subset of the gPC.
Providing these properties, PC gives the possibility to reduce a stochastic differential equation to a high-dimensional deterministic problem.

3.2 Orthogonal Polynomials of the Askey scheme

The orthogonal polynomials belonging to the Askey scheme are so-called hypergeometric orthogonal polynomials. They satisfy a specific type of differential or difference equation. Each of these orthogonal polynomials is a subset of the Askey-scheme and has different weight functions in their orthogonal relationship.
For some of these orthogonal polynomials their weight function correspond or is identical to the probability function of a certain random distribution [3]. These polynomials are used to describe a stochastic process and for this an expansion of polynomials (usually of the same kind) is used. It has been shown that the rate of convergence is optimal, if the orthogonal polynomial corresponding to the stochastic processes has been chosen [3]. Furthermore, if the optimal orthogonal polynomial has not been chosen than the solution will still converges, but the rate of convergence is much lower. In other words, the expansion consists of is less terms when the correct polynomials have been chosen than when polynomials are chosen of which the weighting function does not correspond with the distribution of the stochastic process.
Further on in this section an example is given why only two terms in an expansion are needed to represent a Gaussian process when Hermite polynomials are used.
In table 3.1 the Askey-scheme is depicted, we see that for the commonly used continuous and discrete distributions associated families of orthogonal polynomials exist. In the remainder of this monograph we will only focus on the PC based on continuous distributions, and therefore discrete distributions and polynomials are not discussed in the following.

Table 3.1: Families of probability laws and corresponding families of orthogonal polynomials

	Distribution ξ	Polynomials $\psi_p(\xi)$	Support
Continuous RV	Gaussian	Hermite	$(-\infty,\infty)$
	Gamma	Laguerre	$[0,\infty)$
	Beta	Jacobi	$[a,b]$
	Uniform	Legendre	$[a,b]$
Discrete RV	Poisson	Charlier	$\{0,1,2,...\}$
	Binomial	Krawtchouk	$\{0,1,2,...n\}$
	Negative binomial	Meixner	$\{0,1,2,...\}$
	Hypergeometric	Hahn	$\{0,1,2,...n\}$

3.3 Properties of orthognal Polynomials

A system of polynomials can be defined as $\{P_n(x), n \in \mathcal{N}\}$. Where $P_n(x)$ is a polynomial of degree n, and the dimension of the system can be infinite, $\mathcal{N} = \{0,1,2...\}$ or finite, $\mathcal{N} = \{0,1,..N\}$. The inner product is defined as,

$$< P_n(x), P_m(x) >= \int_S P_n(x)P_m(x)\mathrm{d}\phi(x) = h_n^2\delta_{n,m}. \tag{3.1}$$

Where S is the support of the measure ϕ, h_n^2 is a nonzero constant and $\delta_{n,m}$ is the Kronecker delta. The set of polynomials is called orthonormal if $h_n^2 = 1$.
The measure ϕ often has a density $w(x)$, so the inner product becomes,

$$< P_n(x), P_m(x) >= \int_S P_n(x)P_m(x)w(x)\mathrm{d}x = h_n^2\delta_{n,m} \tag{3.2}$$

The density $w(x)$ needs to be a positive and continuous function and the integral $\int_S w(x)\mathrm{d}x$ must exist.

3.4 Definitions and Notations

Before we can discuss the PC expansion of a random process, we first need to introduce some definitions and notation. Only $\mathbb{R}$-valued random processes U, defined on a complete probability space (Θ, Σ, P) are considered. The distribution of U is the probability measure induced by the mapping as in (3.3).

$$U : \Theta \mapsto \mathbb{R} \tag{3.3}$$

These $\mathbb{R}$-valued random processes, also denoted as $U(\theta)$, live in a Hilbert space, denoted by $L_2(\Theta,\Sigma,P)$. This vector space is equipped with an inner product $< .,. >$ and a mathematical expectation E[.]. The mathematical expectation of E[UY] is expressed as,

$$E[UY] =< U,Y >= \int_{\Theta} U(\theta)Y(\theta)dP(\theta), \tag{3.4}$$

the mathematical expectation of E[U^2] is expressed as,

$$E[U^2] =< U^2 >=< U,U >= \int_{\Theta} U(\theta)U(\theta)dP(\theta), \tag{3.5}$$

and the mathematical expectation of E[U] is expressed as,

$$E[U] =< U >=< 1,U >= \int_{\Theta} U(\theta)dP(\theta). \tag{3.6}$$

As said earlier, it has been shown in [2] and [5] that any random process in can be represented by a series expansion of orthogonal polynomials in the random space Θ. This can be written as:

$$U(\theta) = \sum_{i=0}^{\infty} c_i \Psi_i(\xi(\theta)) \tag{3.7}$$

The deterministic expansion coefficients c_i are the PC coefficients and the orthogonal polynomials are expressed as Ψ.

The following example shows that there are only two terms needed if Legendre polynomials are used to represent a uniform process $U(\xi)$.
We leave from the point that the process to be represented is uniform distributed and that we would like to represent this process with $\psi_0 = 1, \psi_1 = \xi\}$, so the expansion will look like,

$$\begin{aligned} U(\xi) &= c_0 \cdot \psi_0 + c_1 \cdot \psi_1 \\ &= c_0 \cdot 1 + c_1 \cdot \xi, \end{aligned} \tag{3.8}$$

where ξ is uniform distributed.

It is known that any uniform distribution, or process, can be represented by an expectation value μ and a standard deviation σ. And thus we may also write,

$$U(\xi) = \mu + \sigma\xi, \tag{3.9}$$

From Eq.3.8 and Eq.3.9 we may conclude that c_0 is the expectation value of $U(\xi)$ and c_1 the associated standard deviation. This shows that any uniform distribution can be expressed using only 2 polynomials. This example also hold for normal, beta and gamma distributed process.

3.4.1 One Dimensional PC Basis

When a stochastic process depends on a single random variable ξ we speak of one dimensional chaos. Recall that a weight function is involved in the orthogonality of the polynomials. We define the functional $p_\xi(\theta)$ as the probability density of the random variable ξ. This functional then can be used as a weight function under the condition that this functional corresponds to a weight function belonging to an orthogonality relationship listed in the Askey-scheme. In the one dimensional PC case, $\xi(\theta)$ as well as the weight function are scalar functions and the orthogonality relationship can be expressed as:

$$< \psi_i, \psi_j >= \int_{\mathbb{R}} \psi_i(\theta)\psi_j(\theta)p_\xi(\theta)\mathrm{d}\theta = \delta_{ij} < \psi_i^2 > . \tag{3.10}$$

Where ψ_p is the one dimensional chaos of order p.

3.4.2 Multidimensional PC Basis

We can speak of a multi dimensional PC if the stochastic process is depending on multiple RV's. These RV's can be defined as $\bar{\xi} = \{\xi_1, \ldots, \xi_N\}$. Due to the fact that the RV's are independent of each other, the probability of $\bar{\xi}$ is given by,

$$p_{\bar{\xi}}(\bar{\theta}) = \prod_{i=1}^{N} p_\xi(\theta_i). \tag{3.11}$$

Now we can also define multivariate polynomials, for this we first need to define a multi-index, $\gamma = \{\gamma_1, \cdots, \gamma_N\}$,

$$\lambda(p) = \left\{ \lambda : \sum_{i=1}^{N} \gamma_i = p \right\}, \tag{3.12}$$

$\lambda(p)$ denotes the set of multi-indices. When the multi-index γ is known we can construct the kth multidimensional polynomial Ψ_k using the following expression,

$$\Psi_k(\xi_1, \xi_1, \cdots, \xi_N) = \left\{ \bigcup_{\gamma \in \lambda(p)} \prod_{i=1}^{N} \psi_{\gamma_i}(\xi_i) \right\}, \tag{3.13}$$

where ψ_{γ_i} is the univarite polynomial of order γ_i depending on ξ_i. In the case of 2D $(N = 2)$ polynomials based on one dimensional polynomials up to $p = 2$ the multi-index will look like:

$$
\begin{array}{lcccccccl}
\lambda(0) & : & \gamma & = & (0,0) \rightarrow & \Psi_0 & = & \psi_0(\xi_1)\psi_0(\xi_2) \\
\lambda(1) & : & \gamma & = & (1,0) \rightarrow & \Psi_1 & = & \psi_1(\xi_1)\psi_0(\xi_2) \\
 & : & \gamma & = & (0,1) \rightarrow & \Psi_2 & = & \psi_0(\xi_1)\psi_1(\xi_2) \\
\lambda(2) & : & \gamma & = & (2,0) \rightarrow & \Psi_3 & = & \psi_2(\xi_1)\psi_0(\xi_2) \\
 & : & \gamma & = & (1,1) \rightarrow & \Psi_4 & = & \psi_1(\xi_1)\psi_1(\xi_2) \\
 & : & \gamma & = & (0,2) \rightarrow & \Psi_5 & = & \psi_0(\xi_1)\psi_2(\xi_2)
\end{array}
$$

3.5 Truncation of the PC Expansion

The dimension of the set of RV's and of the PC expansion is usually finite, as can be seen in (3.11). For computational purposes is it also necessary to describe the process by a finite number of polynomials, this leads to a PC expansion of finite dimension. This means that $\{\psi_k\}_{k=0}^{\infty}$ has to be truncated down to $\{\psi_k\}_{k=0}^{p}$ where p is the maximum order. If on chooses truncation order p and a N dimensional stochastic space, the number of terms in the multidimensional expansion is equal to:

$$P+1 = \frac{(N+p)!}{N!p!}. \tag{3.14}$$

The truncated PC expansion can now be expressed as:

$$U = \sum_{i=0}^{P} c_i \Psi_i(\bar{\xi}) + \varepsilon(N,p). \tag{3.15}$$

Where $\varepsilon(N,p)$ is the truncation error. It has been shown in [6] that the truncated expansion converges in the mean square sense as N and p go to infinity.

3.6 First- and Second-Order Estimates

Prevously it has been shown that a second order random variable can be represented by an expansion of orthogonal polynomials as:

$$U \approx \sum_{i=0}^{P} c_i \Psi_i(\bar{\xi}). \tag{3.16}$$

We assume that the orthogonal polynomial basis has the indexing such that $\Psi_0(\bar{\xi}) = 1$. Now the orthogonal basis helps us by calculating the expectation of the random variable U:

$$E[U] =< U(\bar{\xi}) >=< \Psi_0(\bar{\xi}), U(\bar{\xi}) >= \sum_{k=0}^{P} c_i < \Psi_0(\bar{\xi}), \Psi_k(\bar{\xi}) >= c_0. \tag{3.17}$$

Furthermore it is possible to obtain the variance of U from the definition of random variable variance:

$$\sigma_U^2 = E[(U - E[U])^2] = E[(\sum_{k=1}^{P} c_k \Psi_k)^2] = \sum_{k,l=1}^{P} c_k c_l < \Psi_k, \Psi_l >= \sum_{k=1}^{P} c_k^2 < \Psi_k^2 > . \tag{3.18}$$

So the variance of U can be expressed as a weighted sum of the squared PC coefficients. The expressions for higher order moments of U can be derived similarly, but those expressions are not as simple as for the first and second order ones. It is therefore much easier to obtain the higher order moments via a sampling strategy of the realization of U. Once the PC coefficients are known, it is possible to build the stochastic solution by using Eq. (3.16) by sampling the stochastic quantity $\bar{\xi}$. Consequently the statistics of U, such as densities, cumulative density functions and probabilities can be estimated when necessary. This Monte Carlo procedure is basically the same procedure as discussed in 2.2.1, the only difference here is that not mathematical model describing U but the right side of Eq.(3.16) is sampled. Generally sampling the right side of Eq.(3.16) requires much less computational resources compared to sampling the mathematical model. Sampling the right side of Eq.(3.16) is relatively cheap because it is in fact only the evaluation of of the term $\Psi(\bar{\xi})$.

3.7 Spectral Projection

In general it is possible to simulate a process by a mathematical model for which there usually exists a solution. There are many methods available to find a solution of a particular mathematical problem, but in general most methods are not suitable or comprehensive enough when dealing with models of stochastic processes.
These stochastic models can be seen as a mathematical problem concerning the process and a mathematical problem covering the uncertainies on the process. As it has been shown in Eq. (3.16), the solution of such a model can be cast in the form of an expansion. The solution can now be found by solving for the PC coefficients. There are mainly two approaches for solving stochastic models using spectral methods, namely intrusive and non-intrusive.
Intrusive methods basically solve the solution of a model by directly searching the PC coefficients, a Galerkin Projection Method is usually used for this method.

The main advantage of this method is that it can be fast, depending on the number of unknown coefficients and the type of system to be solved. Disadvantages are that it can not be applied to a model of which the numerical source code is not accessible or implementation is difficult because of the complexity of the model. Furthermore when the number of unknown coefficients is too large $(O(10^6))$, an intrusive method is usually slow. A description and examples of how Intrusive PC can be used can be found here [3, 7, 8] and will further on in this chapter be no more discussed.

3.8 Non-Intrusive Spectral Projection

A Non-Intrusive Spectral Projecction (NISP) method works in two or more stages, first the model is solved in specific points of the stochastic domain and subsequently the PC coefficients are determined. For finding the solution in these specific points a deterministic solver can be used, so no particular adaptation of existing codes is necessary. This is in contrast with a Galerkin Projection Method, where a reformulation of the problem is required.
Advantages of NISP methods are that the model can be used as a black-box, so it can be applied to complex models, industrial applications and situations where only deterministic codes are available [8]. And since the model needs to be solved in a deterministic set of sample points, it is possible to perform this process in parallel.

3.8.1 Orthogonal Projection

As mentioned in **??**, NISP aims at determining the PC coefficients. Again, the Polynomial Chaos expansion of the output variable U is given by:

$$U \approx \sum_{i=0}^{P} c_i \Psi_i(\bar{\xi}). \tag{3.19}$$

Now the orthogonal relationship of the polynomials can be used to compute the expansion coefficients. If we multiply both sides of Eq. (3.19) by the qth polynomial of the set $\{\Psi_i\}_{i=0}^{P}$ and we integrate over the stochastic domain $\bar{\xi}$, we get the following expression of the qth coefficient.

$$\begin{aligned} \langle \Psi_q, U \rangle &= \left\langle \Psi_q, \sum_{i=0}^{P} c_i \Psi_i \right\rangle \\ &= \langle \Psi_q c_q, \Psi_q \rangle \\ &= c_q \langle \Psi_q, \Psi_q \rangle, \end{aligned} \tag{3.20}$$

or, rewriting,

$$c_q = \frac{\langle \Psi_q, U \rangle}{\langle \Psi_q, \Psi_q \rangle}, \tag{3.21}$$

where $q = 0, 1, 2 \cdots p$. It is important to note that the normalisation factor $\langle \Psi_q, \Psi_q \rangle$ is depending only on the chosen basis and that this factor is analytically known. The numerator of the right hand side of Eq. (3.21) is the only term that needs to be computed. Furthermore it is clear to see that the coefficients are independent of each other. While the same coefficients would be coupled in case of a Galerkin method.

There are a couple of existing techniques to compute the numerator of the right hand side of Eq. (3.21) In the next paragraphs some of these methods will be discussed.

3.8.2 Monte Carlo Integration

One of the simplest techniques to perform numerical integration is the Monte Carlo Method. It is based on a sample set of independent realizations of the RV, in this case $\bar{\xi}$, which is normally generated by using a random number generator. Consequently the solution U_i depending on a certain $\bar{\xi}_i$ is sought, this is done by solving the model for $U\left(\bar{\xi}_i\right)$. Once all the solutions corresponding to the sample set $\left\{\bar{\xi}_i\right\}_{i=1}^{M}$ are known, the integral can be computed by;

$$\left\langle \Psi_q, U \right\rangle = \frac{1}{M}\sum_{i=1}^{M} U_i \Psi_q\left(\xi_i\right) + \varepsilon_M, \tag{3.22}$$

where M is the dimension of the sample set and ε_M the sampling error. According to the law of large numbers it can be shown that the expectation and variance of the sampling error go to zero if M is sufficiently large. The rate of convergence of the sampling error is $\frac{1}{\sqrt{M}}$ and this rather slow convergence is the main limitation of using Monte Carlo Integration for NISP techniques. On the other side the convergence rate is not depending on the dimensionality of the problem and the smoothness of the term $U\left(\xi\right)\Psi_q\left(\xi\right)$. This suggest a robust behaviour in case that the solution lives in a high dimensional space and or is non-smooth. Nevertheless, because of the slow rate of convergence it is worth looking in to alternative methods such as deterministic integration techniques.

3.8.3 Deterministic Integration

A deterministic integration method is a method which can be used to numerically compute a definite integral, such as

$$\int_a^b f\left(x\right) dx < \infty. \tag{3.23}$$

If the function which has to be integrated, (in this case $f\left(x\right)$) is smooth and lies in a small dimensional space, then a deterministic method can be used.
Generally all deterministic integration methods are based on a finite sum of evaluations of the function to be integrated. These evaluations are defined in a set of deterministic points of the integration domain. Usually a numerical integration over a one dimensional space (e.g. univariate integration) is called numerical quadrature, whereas a integration over more than one dimension (e.g. multivariate integration) is sometimes described as cubature.
An univariate integration can be approximated by the following quadrature,

$$\int_a^b f(x)\,dx \approx \sum_{i=1}^{N_Q} f(x_i)\,w_i, \tag{3.24}$$

where the quadrature approximation can be written in operator-form by Q,

$$\int_a^b f(x)\,dx \approx Qf(x) = \sum_{i=1}^{N_Q} f(x_i)\,w_i. \tag{3.25}$$

3.8.3.1 Gauss Quadrature Over the time, various quadrature methods have been invented such as the rectangle rule, trapezoidal rule and the Gaussian quadrature. The latter has a high order of accuracy, and is very efficient, especially when integrating polynomials.
Every polynomial belonging to the Askey scheme has an associated Gaussian quadrature with which the integral can be exactly computed up to $p = 2n_l - 1$, where n_l is the number of points used and p the order of the polynomial. For a comprehensive introduction to Gaussian quadratures, see [8, p. 499].

3.8.3.2 Tensor Product Formulas When performing an multivariate integration of an N-dimensional function the associated cubature belonging to this integral can be built using a one-dimensional quadrature formalism. If the same type of quadrature can be used for each integration variable, then the cubature can be written as:

$$Q^{(N)} f(\bar{x}) = \left(Q_l^{(1)} \otimes \cdots \otimes Q_l^{(1)}\right) f(\bar{x}). \tag{3.26}$$

If each integration variable needs a different type of quadrature, then the previous definition can be extended by using different quadrature rules in each integration direction, being l_1 up to l_N. This gives,

$$Q^{(N)} f(\bar{x}) = \left(Q_{l_1}^{(1)} \otimes \cdots \otimes Q_{l_N}^{(1)}\right) f(\bar{x}), \tag{3.27}$$

As discussed, to numerically integrate a function $f(x)$, it is necessary to know the outcome of the function in a finite set of points. This set of points can be seen as a grid, especially when dealing with multivariate integrals. In Fig. 3.1 three two-dimensional grids constructed using Eq. (3.26) are displayed. These grids are based on a one-dimensional Gauss-Legendre quadrature rule, where for (a) this rule gives 5 points, (b) 10 points and (c) 15 points. The figure shows clearly that the number of grid point increase exponentially with increasing number of quadrature points for the one-dimensional rule. And in fact the number of points in the grid can be computed by ; $N_Q = (n_l)^N$, where N_Q is the number of cubature points, n_l the number of points in the underlying 1D quadrature rule and N the dimension of the integration space.

The number of grid points can be slightly optimized by using Eq. (3.27), where $N_q = \prod_i n_{l_i}$, nevertheless in both cases N_Q increases exponentially with N. This exponential increase is also known as the curse of dimensionality and shows that a Gaussian quadrature is not optimal for high-dimensional models and models which need a moderate to high order of stochastic basis functions.

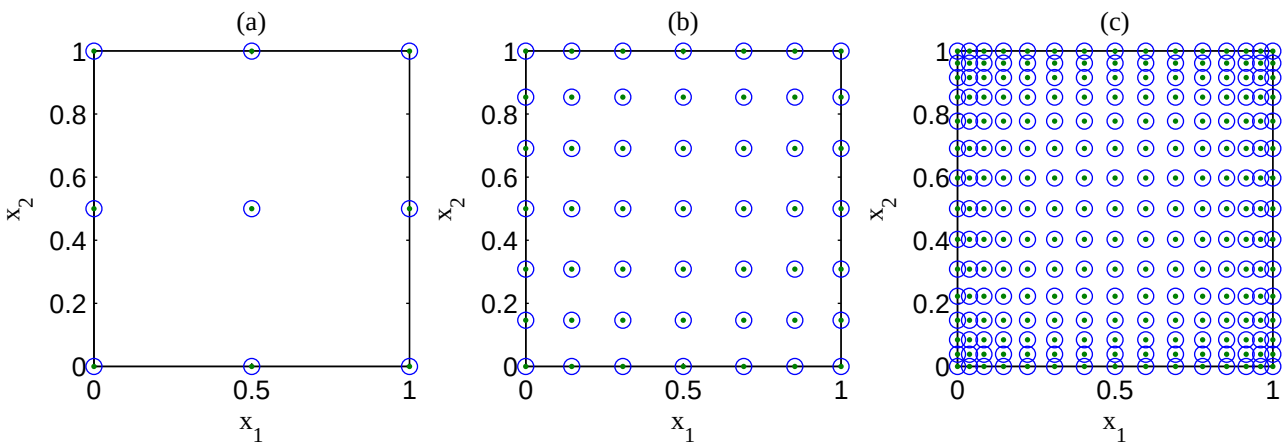


Figure 3.1: Two dimensional cubature grids for (a)the underlying 1D rule consists of 3 points whereas for (b) consists of 7 points and (c) of 15 points. For (a) $N_Q = 9$, (b) $N_Q = 49$ and (c) $N_Q = 225$.

Clearly the selection of N_Q is an important issue for NISP. The Gaussian quadrature is exact for polynomials up to degree $2n_l - 1$, unfortunately this is only valid when the integrand is a polynomial and this is not necessarily the case for every integrand. The difficulty lies in $\langle \Psi_q(\xi), U(.,\xi) \rangle$, since $U(.,\xi)$ is not necessarily polynomial in ξ. And thus there is no clear rule for the selection of N_Q, except that in general $n_l > p_{max}$, where p_{max} is the maximum order of the 1D-polynomials. [8, p. 149].

3.8.3.3 Sparse Grids As discussed in the previous paragraph the large number of grid points generated when using tensorized quadrature formulas of high dimensional spaces can cause serious computational issues, known as the *curse of dimensionality*.
An efficient way to reduce or temper this issue can be achieved by using a sparse grid instead of a grid based on the tensorized formula. The sparse grid method is a numerical discretization technique which was first introduce by Smolyak [9].
Cubature formulas involve $O\left((n_l)^N\right)$, where sparse grids only involve $O\left(n_l \log(n_l)^{N-1}\right)$ number of nodes to achieve the same degree of exactness. So to a certain extent the exponential dependence on the dimension of the problem can be overcome by using a sparse grid.
Nevertheless there is still no general method to chose the optimal number of nodes, which is also the case when using a cubature formula.
A lot of research has been done on sparse grids, mainly to further reduce the number of nodes providing the same degree of exactness especially for high dimensional grids. An introduction and additional information

regarding sparse grids and the underlying theory can be found here [8, p. 56] and [10, 11].

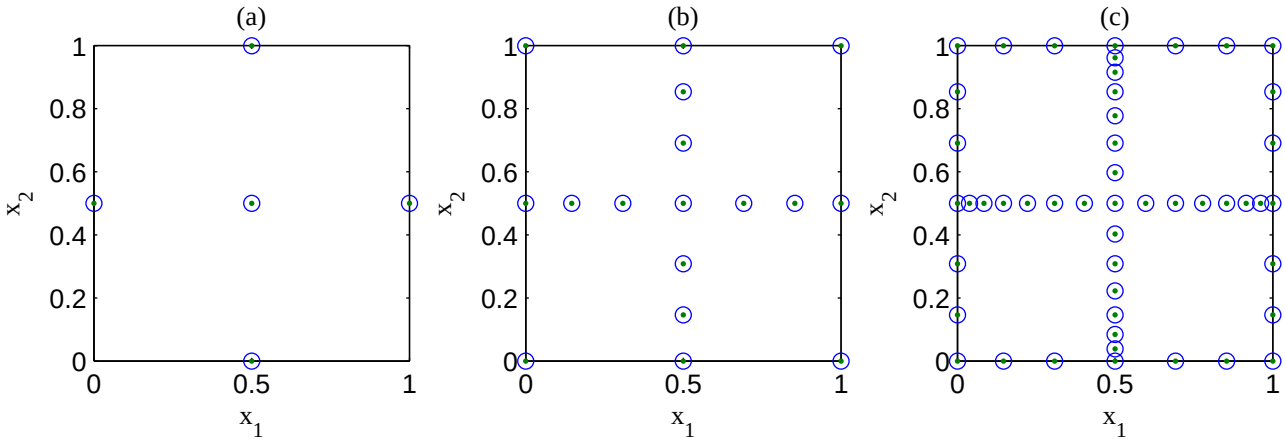


Figure 3.2: Two dimensional sparse grids for (a)the underlying 1D rule consist out of 3 points whereas for (b) consist of 7 points and (c) of 15 points. For (a) $N_Q = 4$, (b) $N_Q = 17$ and (c) $N_Q = 49$.

Table 3.2: 2 Dimensional shifted-Legendre polynomials of maxmum order 2 (constructed from the 1 dimensional polynomials of order $p \leq 2$)

Ψ_n	$\psi_i\psi_j$		
Ψ_0	$\psi_0\psi_0$	=	1
Ψ_1	$\psi_1\psi_0$	=	$2x-1$
Ψ_2	$\psi_0\psi_1$	=	$2y-1$
Ψ_3	$\psi_2\psi_0$	=	$6x^2-6x+1$
Ψ_4	$\psi_1\psi_1$	=	$4xy-2(x+y)+1$
Ψ_5	$\psi_0\psi_2$	=	$6y^2-6y+1$
Ψ_6	$\psi_1\psi_2$	=	$12xy^2-12xy+2x-6y^2+6y-1$
Ψ_7	$\psi_2\psi_1$	=	$12yx^2-12xy+2y-6x^2+6x-1$
Ψ_8	$\psi_2\psi_2$	=	$36x^2y^2-36yx^2+6x^2-36xy^2+36xy-6x+6y^2-6y+1$

In Fig. 3.2 three 2D sparse grids are shown. The underlying rule with which these sparse grids where constructed consists of (a) 3 nodes, (b) 7 nodes and (c) 15 nodes. For the construction of the sparse grids shown in fig. 3.1, the same rules where used.
In Table 3.2 the 2 dimensional polynomials, constructed from the 1 dimensional shifted Legendre polynomials up to order $p = 2$, are listed. The first six polynomials from Table 3.2 form a set which can be used for a polynomial chaos expansion of order 2 (in the other polynomials 3th and 4th order terms are involved and can therefore not be used). For the integration of $< \Psi p, \Psi q >$ a tensorized grid as in Fig. 3.1a can be used. With the use of this grid we can integrate all possible inner products including inner products which consists of the 6th, 7th and 8th order polynomials, but we are only interested in the polynomials up to order five. In other words, the grid in Fig. 3.1a contains more points, or information than needed. So some of the point

in the grid are not necessarily needed to integrate the polynomials used in the expansion. If this is the case, we can then remove those points from the grid and only use the computational power there where needed. When we compare Fig. 3.1a with Fig. 3.2a we can see that the four outer points of the tensorized grid have been removed in the sparse grid but have the same degree of exactness for the inner products associated to the first six polynomials of Table 3.2.

4 Compartment Model for radionuclide release into the Biosphere

4.1 Introduction

A number of countries are currently considering deep geological storage of nuclear waste as a disposal option. For the selection of suitable sites, a performance assessment needs to be carried out in order to demonstrate the safety of the sites and usually mathematical models are used for this.
In general the behaviour of radionuclide concentrations in specific areas over long time scales is of main concern. A qualitative indication of the behaviour of these radionuclide concentrations can be obtained through the construction of mathematical models for radionuclide transport through the biosphere.
In the following paragraphs we will discuss the physical processes which will affect the behaviour of the radionuclide concentrations in the biosphere and introduce the radionuclide transport equations and show how a compartment model is derived from this transport equation.

4.2 The Transport Equation

Given the assumption that the motion of a fluid through porous media can be described by advection-dispersion model [12], the equations for radionuclide migration in porous media can be written as:

$$\frac{\partial}{\partial t}\varepsilon C^k(\mathbf{r},t) = -\nabla\cdot\mathbf{J}^k(\mathbf{r},t) - \lambda^k\varepsilon C^k(\mathbf{r},t) + \lambda^{k-1}\varepsilon C^{k-1}(\mathbf{r},t) - f(F^k,C^k) + S^k(\mathbf{r},t) \tag{4.1}$$

$$\frac{\partial}{\partial t}(1-\varepsilon)F^k(\mathbf{r},t) = -\lambda^k(1-\varepsilon)F^k(\mathbf{r},t) + \lambda^{k-1}(1-\varepsilon)F^{k-1}(\mathbf{r},t) - f(F^k,C^k) \tag{4.2}$$

with:

$$\mathbf{J}^k(\mathbf{r},t) = -D(\mathbf{r},t)\nabla C^k(\mathbf{r},t) + \mathbf{U}(\mathbf{r},t)C^k(\mathbf{r},t) \tag{4.3}$$

Where

Where :

ε	Porosity of the rock.
$C^k(\mathbf{r},t)$	Molar concentration in water of radionuclide k at position $\mathbf{r}$ and time t.
$\mathbf{J}^k(\mathbf{r},t)$	Current of the fluid at position $\mathbf{r}$ and time t.
λ^k	Decay constant of radionuclide k.
$F^k(\mathbf{r},t)$	Molar concentration of adsorbed radionuclide k on solids at position $\mathbf{r}$ and time t.
$D(\mathbf{r},t)$	Dispersion coefficient of water in the media at position $\mathbf{r}$ and time t.
$\mathbf{U}(\mathbf{r},t)$	Advective flow velocity of water in the media at position $\mathbf{r}$ and time t.
$f(F,C)$	Adsorption isotherm.
$S^k(\mathbf{r},t)$	External source of radionuclide k.

The physics behind Eq. (4.1) and Eq. (4.2) is as follows. The left hand side of Eq. (4.1) describes the rate of change over time of the concentration of radionuclide species k dissolved in water. The first term on the right hand side describes the movement due to dispersion and advection. The second term is a sink due to radioactive decay and the third term is a in-growth due to decay of the preceding radionuclide in the decay chain. If k is the first radionuclide in the chain than the third term is zero. Radionuclides can also be removed from the solution by adsorption on the surface of the host rock. The inverse process, desorption can also take place and so there is a relationship between the concentration in solution and that being adsorbed on the rock. This is called the *adsorption isotherm* relation. The concentration of radionuclides that is being absorbed by the host rock is described by Eq. (4.2). The left hand side of this equation describes the rate of change over time of the concentration of radionuclide species k adsorbed on the rock. There has not been accounted for an advection-dispersion process in this equation because the rock, on which the radionuclides are adsorbed, do not move or hardly move. The other terms in Eq. (4.2) are analogous to those in Eq. (4.1).
It has been shown in [13] that, under the condition that C^k and F^k are proportional, (4.2) can be subtracted from (4.1) and we will obtain,

$$\frac{\partial}{\partial t}(R^k \varepsilon C^k) = -\nabla \cdot \mathbf{J}^k - \lambda^k R^k \varepsilon C^k + \lambda^{k-1} R^{k-1} \varepsilon C^{k-1} + S^k. \tag{4.4}$$

Where R^k is the *retardation factor* of radionuclide species k, which is defined as:

$$R^k = 1 + \frac{1-\varepsilon}{\varepsilon} K_d^k \rho_s. \tag{4.5}$$

And K_d^k the *distribution coefficient* of radionuclide species k. The retardation coefficient is equal to the fraction of radionuclides which is in the mobile phase, e.g. diluted in water and is a dimensionless number. The distribution coefficient is equal to the ratio of concentration of radionuclides in solution phase over concentration of radionuclides in solid phase.

4.3 Derivation of the Compartment Model

Equation (4.4), which is a second order partial differential equation, can in principal be solved with the use of a number of numerical methods. Unfortunately there is not enough detailed knowledge about how dispersion coefficient (D) and the advective flow ($\mathbf{U}$) of water behave in the domain of interest and so those numerical mehods are difficult to implement.
What can be done is splitting up the domain in a convenient number of volumes or compartments and consequently average equation (4.4) over these volumes. And so it is possible to define the concentration in a certain compartment V_i by integrating over the volume of V_i.

$$\int_{V_i} \mathrm{d}\,\mathbf{r} R^k \varepsilon C^k(\mathbf{r},t) = R_i^k C_i^k(t) \tag{4.6}$$

Furthermore it is possible to express the transport from compartment i to j, therefore the average transport in a volume V_i can be written as:

$$\int_{V_i} \mathrm{d}\mathbf{r}\nabla \cdot \mathbf{J}^k(\mathbf{r},t) = \int_{S_i} \mathrm{d}\mathbf{S} \cdot \mathbf{J}^k(\mathbf{r}_s,t) \tag{4.7}$$

The Gauss' Theorem is used to rewrite the volume integral into a surface integral and consequently this surface integral can be written as:

$$-\int_{S_i} \mathrm{d}\mathbf{S} \cdot \mathbf{J}^k(\mathbf{r}_s,t) \approx \sum_j r_{ji} C_j^k(t) - C_i^k(t) \sum_j r_{ij} \tag{4.8}$$

The rate transfer coefficients $r_{i,j}$ in Eq. (4.8) can be obtained from separate measurements and calculation of earth processes. They describe the rate at which a radionuclide is transported into and out compartment i over a boundary which compartment i shares with compartment j. It is convenient to redefine the rate transport coefficient as:

$$a_{i,j} = r_{j,i} \qquad i \neq j$$
$$a_{i,j} = -\sum_{j=1}^{N_c} r_{j,i} = -\sum_{j=1}^{N_c} a_{j,i} \qquad i \neq j$$

Where N_c is the number of compartments. To cover conservation of mass, it is necessary to set:

$$\sum_{i=1}^{N_c} a_{i,j} = 0 \tag{4.9}$$

Equation (4.7) can be written as:

$$\int_{S_i} \mathrm{d}\mathbf{S} \cdot \mathbf{J}^k(\mathbf{r}_s,t) \approx -\sum_j a_{i,j} C_j^k(t) \tag{4.10}$$

Substituting the relations of Eq. (4.6) and Eq. (4.10) into Eq. (4.4) gives use the compartment model:

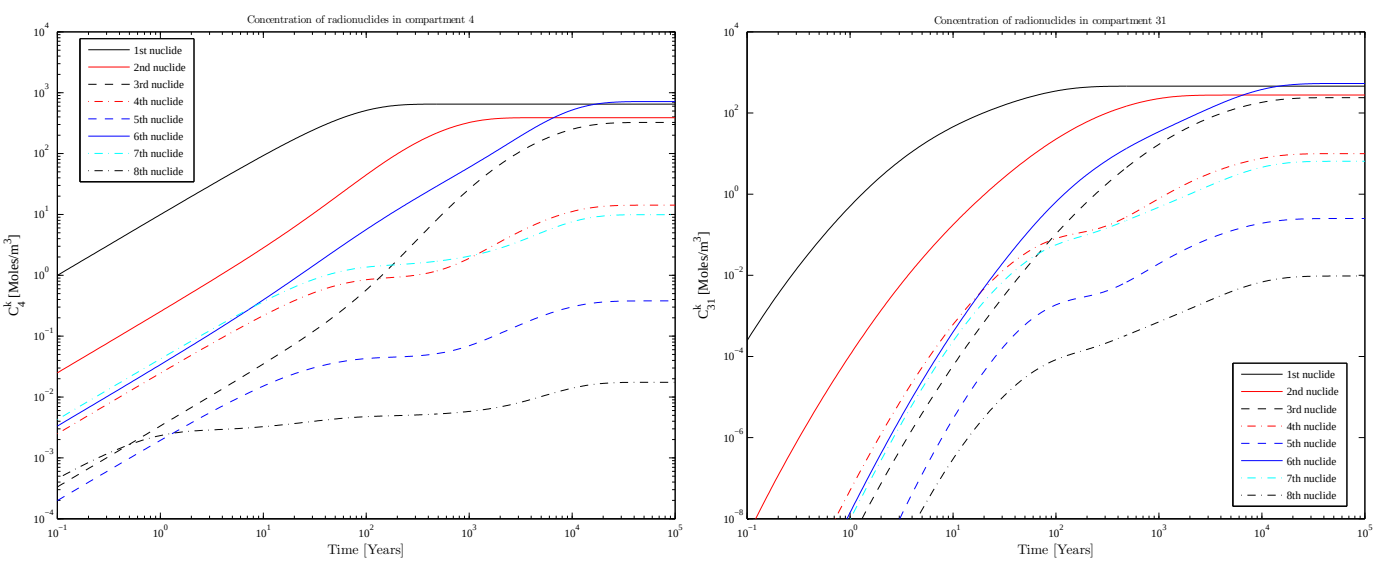
$$\frac{\partial}{\partial t} R_i^k C_i^k = \sum_{j=1}^{N_c} a_{i,j} C_j^k - \lambda^k R_i^k C_i^k + \lambda^{k-1} R_i^{k-1} C_i^{k-1} + S_i^k \tag{4.11}$$

The left-hand side of Eq. (4.11) describes the rate of change of the concentration of radionuclide k in compartment i over time. the first term in the right-hand side describes the flux of concentration of radionuclide k from compartment i into compartment k. The second term describes the decay of radionuclide k, whereas the third term describes the in-grow of concentration of radionuclides from the preceding radionuclide in the decay chain. The last term in Eq. (4.11) describes the source of radionuclides.

4.4 Solving the Compartment Model

Equation (4.11), which has been derived from (4.4) is a set of $N_c \cdot N_r$ coupled first order, Ordinary Differential Equations. This system has been solved in [13] by using parameters which do not correspond to a physical element or a real case scenario. According to [13] the presented case was used for illustrative purposes. In the presented work we decided to use the model presented in [13] as a benchmark.

The *'illustrative case'* in [13] uses a decay chain containing 8 radionuclides and 31 compartments which gives a system of 248 coupled first order differential equations. Furthermore, due to large variations in rate transfer coefficients the set of equations is stiff. This requires a solver that can handle a moderately large set of couple and stiff equations. The time-span over which the concentrations should be calculated reaches from 0 to 100,000 years. In figures 4.1a and 4.1b the results of the compartment model are plotted for compartments 4 and 31. The plots clearly indicate that the concentration will saturate after about 10.000 years. This because, in this *illustrative case*, the source terms has been assumed to be constant with time, e.g. there is a constant influx of radionuclides into the total domain.



(a) The radionuclide concentration in compartment 4 as a function of time.

(b) The radionuclide concentration in compartment 31 as a function of time.

Figure 4.1: The radionuclide concentration in compartment 4 and 31 as a function of time.

4.5 Uncertainties in the Compartment Model

In Chapter 4.3 it has been shown how to tackle the problem of having limited data available for the dispersion coefficient and the velocity field of the advective flow of the underground water. This problem, the limited

amount of available data, is unfortunately not only bounded to the dispersion and advection terms but also counts for the distribution coefficient. This is due to the fact that the physical and chemical processes behind the absorption-desorption processes are not fully understood yet. The distribution coefficient is furthermore depending on a variety of parameters such as, but not limited to; temperature, soil texture, pH and the solute. Due to these dependencies it can be assumed that the distribution coefficient is a function of a random variable, ξ. The concentration of radionuclides, which is depending on the distribution coefficient through the retardation coefficient, becomes a functional depending on ξ. Equation (4.11) will now become:

$$\frac{\partial}{\partial t} R_i^k(\xi) C_i^k(t,\xi) = \sum_{j=1}^{N_c} a_{i,j} C_j^k(t,\xi) - \lambda^k R_i^k(\xi) C_i^k(t,\xi) + \lambda^{k-1} R^{k-1}(\xi)_i C_i^{k-1}(t,\xi) + S_i^k \tag{4.12}$$

Where the retardation coefficient is now be expressed as:

$$R_i^k(\xi) = 1 + \frac{1-\varepsilon}{\varepsilon} K_d^k(\xi) \rho_s. \tag{4.13}$$

We assume that $\xi = (\xi_1, \xi_2, \cdots, \xi_M)$ are independent identically distributed random variables and M indicates the number of dimensions in random space. The uncertainties in the distribution coefficient have been already investigated in [7], where it has been chosen to vary K_d^k according to a log-uniform distribution.
We decided to use the same log-uniform distribution for K_d^k, so that [7] can be used as a reference. Moreover, a log-uniform distribution of the distribution coefficient also mimics the physical reality since the values of K_d^k can span over several orders of magnitude [14]. This allows us to write a probability density function (pdf) for the values of K_d^k:

$$p\left(K_d^k(\xi)\right) = \frac{1}{K_d^k(\xi) \ln\left(\frac{K_{d,max}^k}{K_{d,min}^k}\right)}, \tag{4.14}$$

where $K_{d,min}^k$ the minimum value and $K_{d,min}^k$ the maximum value is of the distribution coefficient for radionuclide species k. With the use of the cumulative distribution function, it is possible to define an analytical expression for the distribution coefficient (K_d^k) as function of the RV ξ:

$$\xi_n = \int_{K_{d,min}^k}^{K_d^k(\xi)} p\left(K_d^{k'}(\xi)\right) \mathrm{d}K_d^{k'}(\xi) = \frac{\ln \frac{K_d^k(\xi)}{K_d^{k,min}}}{\ln \frac{K_{d,max}^k}{K_d^{k,min}}}, \tag{4.15}$$

isolating $K_d^k(\xi)$ out of Eq.(4.15) gives:

$$K_d^k(\xi) = K_{d,min}^k \left(\frac{K_{d,max}^k}{K_{d,min}^k}\right)^{\xi_n}. \tag{4.16}$$

Since Eq. (4.12) is depending on a random variable ξ, this equation can be viewed as a stochastic differential equation. As said earlier, there are a couple of methods available to deal with stochastic differential equations with which the stochastic behaviour of the solution of such equations can be quantified.

4.6 UQ of the Compartment Model with Monte Carlo

As said before, the Monte Carlo method is a computational algorithm which can be used to quantify the uncertainty of stochastic differential equations. A description of how MC is implemented can be found in Appendix A.1.

4.6.1 Results of UQ with Monte Carlo

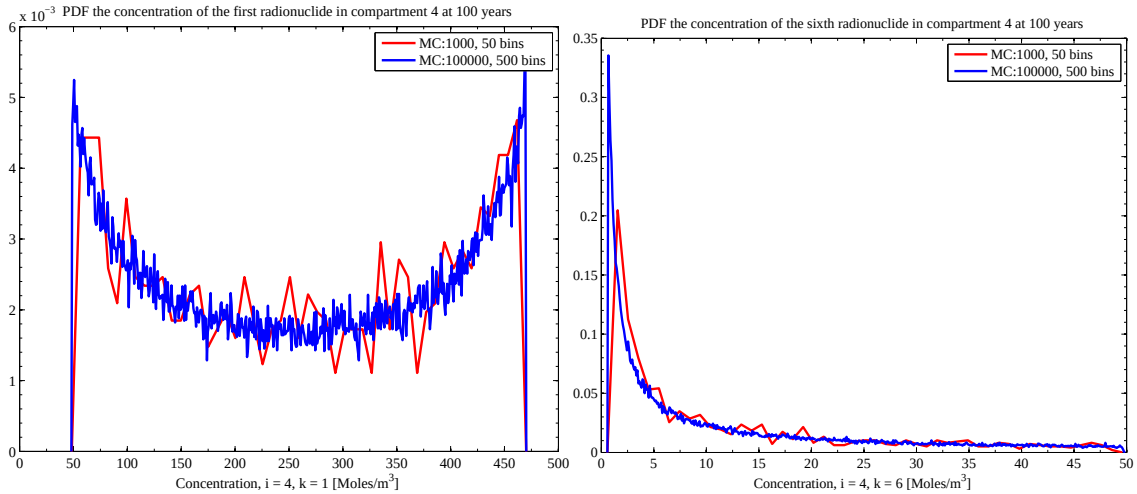
In Table 4.1 the calculated mean value, standard deviation and the standard deviation of the mean of the concentration of radionuclides $k = 1$ and $k = 6$ in compartement $i = 4$ at time 100 years are depicted. Both ($k = 1$ and $k = 6$) have been calculated with 10^3 and 10^5 samples (N). The calculation for 10^3 samples took about 12 minutes, where as the 10^5 samples took about 19 hours. The results of all the radionuclide concentrations in compartment 4 and 31 at time 100 years and 100.000 years and calculated with 10^3 and 10^5 number of samples can be found in tables A.1 until A.4.

Table 4.1: Mean and standard deviation of the concentrations of radionuclides k = 1 and k = 6 in compartment 4 at 100 years.

$k = 1$						
i	N	time	$E[C_4^k]$	$\sigma_{C_4^k}$	$\sigma_{E[C_4^k]}$	$\frac{100\%}{\sqrt{N}}$
4	10^3	100	265.2288	138.4375	4.3800	3.16 %
4	10^5	100	260.3267	138.5777	0.4382	0.32 %
$k = 6$						
4	10^3	100	11.4379	12.5557	0.3972	3.16 %
4	10^5	100	11.4202	12.7023	0.0402	0.32 %

In figure 4.2 the pdf's of the concentrations in compartment 4 for the first radionuclide (4.2a) and for the sixth radionuclide (4.2b) at 100 years are given. The MC:1000 pdf has been constructed by dividing the range of the concentrations in 50 equidistant bins and counting the number of realizations per bin. The number of realizations per bin have been normalised by dividing it by the total number of realizations. We see that the pdf is not a smooth function, if less bins are used the smoothness of the function will increase. The disadvantage of this is that information will be lost due to the averaging over larger bins. This already can be seen if we compare MC:1000 with MC:100000. The peaks in MC:100000 are more pronounced which can be described the the fact that more bins are used, this besides the fact that more realizations are used.

Although the pdf's still contain noise, they give a good description of the actual pdf's and it can be shown that, if the number of bins and samples go to infinity the pdf's will converge the the actual pdf's.



(a) (b)

Figure 4.2: The pdf's of the first radionuclide in compartment 4 at 100 years (a) and the pdf of the fourth radionuclide in compartment 4 at 100 year (b), both computed with Monte Carlo.

4.7 UQ of the Compartment Model with Polynomial Chaos

As discussed in Chapter 3, using a Polynomials Chaos expansion, a stochastic process can be written as a weighted sum of orthogonal polynomials which span the stochastic space. In the case of the compartment model, the concentration can be seen as a process which evolves over time in a stochastic space. The distribution coefficient is log-uniform distributed and in Eq.(4.16) we have shown that the distribution coefficient is depending on RV ξ_n.

From Table 3.1 it can be seen that there does not exist an associated family of orthogonal polynomials to a log-uniform distribution. Since the distribution coefficient is a function of ξ, and ξ is uniform distributed it is convenient to choose Legendre polynomials. Note that the RV $\xi \sim U(0,1)$ and thus the used Legendre polynomials are not the standard ones, but the 'shifted' Legendre polynomials which are orthogonal on [0,1] are used. Furthermore, the maximum order of the polynomials have been set to $p = 5$, so only the first 6 shifted Legendre polynomials are used. From these shifted Legendre polynomials the multidimensional polynomials will be constructed, and since there are 8 independent random variables these polynomials are 8 dimensional functions.

The concentration of radionuclides can now be written as a PC expansion:

$$C_i^k\left(t,\bar{\xi}\right) = \sum_{n=0}^{P} C_{i,n}^k\left(t\right)\Psi_n\left(\bar{\xi}\right) \tag{4.17}$$

,

where $C_{i,n}^{k}(t)$ are the PC coefficients, $\Psi_n(\bar{\xi})$ are the 8-dimensional shifted Legendre polynomials and $\bar{\xi}$ an 8-dimensional random variable. Since the polynomials are known, one (only) needs to determine the PC coefficients. This can be done by using a orthogonal projection described in paragraph 3.8.1. Since the dimension of the stochastic space and the amount of base function per dimension are known, eq. 3.14 gives that 1287 PC coefficients need to be determined and thus $P = 1286$.
For the evaluation of the PC coefficients a numerical integration is necessary. In paragraph 3.8 a few methods have been discussed for the evaluation of the coefficients. If a regular grid based on a Gauss quadrature, mentioned in 3.8.3.1, where to be used, than the system of differential equations needs to be solved at least in 6^8 grid points. On average one evaluation of the system takes about 0.69 seconds so the total evaluation time for $6^8(\approx 1.68 \cdot 10^6)$ evaluations would be around 13 days.
If instead of a regular grid a sparse grid is used, than the number of grid points can be reduced while maintaining the same order of accuracy. Here, two types of sparse grids were used. One sparse grid is based on a Gauss quadrature for Legendre polynomials using Smolyak's method to construct the grid, and is being referred to as 'GQU' (Gauss Quadrature with Uniform measure). The second sparse grid is based on a so called 'delayed Kronrod-Patterson sequence Gauss-Legendre quadrature' and is being referred to as 'KPU' (Kronrod-Patterson with Uniform measure). This KPU sparse grid uses an optimization technique such that the number of nodes can be decreased will maintaining the same order of accuracy compared to a normal Smolyak sparse grid [11].
In table 4.2 the number of nodes and cputime are listed, it is clear to see that the KPU sparse grid uses the least computational resources.

Table 4.2: Number of nodes and cputime for given underlying quadrature rules of a normal Cubature, a GQU- and a KPU sparse grid.

	Cubature		GQU		KPU	
n_l	N_Q	cputime [days]	N_Q	cputime [days]	N_Q	cputime [days]
5	390625	3.11	3905	0.03	2177	0.02
6	1679616	13.41	15153	0.12	6657	0.05
7	5764801	46.03	51713	0.41	17921	0.14
8	16777216	133.98	159425	1.27	43137	0.34

Once the PC coefficients are known a functional is obtained describing the stochastic behaviour of the concentration at a specific time. The mean value and variance (or standard deviation) can now be evaluated as described in paragraph 3.6 or by sampling the obtained functional. The first method should be the fastest, but if one is interested in higher order estimators than the sampling method is usually easier [8]. If the random variable $\bar{\xi}$ is in each stochastic dimension N times uniformly sampled over [0,1] and these N samples are substituted into eq. (4.17) we obtain N realizations. From this set of N realizations the statics of the concentration can be measured.

Table 4.3: Mean value of the concentration and the associated standard deviation for the first and sixth radionuclide in compartment 4 at time 100 years, calculated via the deterministic and via the sampling method.

	GQU				KPU			
	deterministic		sampled (N = $1.1 \cdot 10^6$)		deterministic		sampled (N = $1.1 \cdot 10^6$)	
n_l	$E[C_4^k]$	σ_4^k	$E[C_4^k]$	σ_4^k	$E[C_4^k]$	σ_4^k	$E[C_4^k]$	σ_4^k
	$k = 1$, compartment 4							
5	260.5571	139.1826	260.7918	139.0733	260.5074	138.9891	260.7519	138.8833
6	260.5448	138.8723	260.6954	138.8426	260.5074	138.7798	260.6592	138.7495
7	260.4654	138.6981	260.7463	138.6986	260.5244	138.7632	260.8052	138.7633
8	260.5071	138.7487	260.3739	138.7177	260.5321	138.7717	260.3967	138.7395
	$k = 6$, compartment 4							
5	11.5285	12.8007	11.5334	12.8135	11.5225	12.7981	12.5316	2429.4092
6	11.5317	12.8066	11.5376	12.8139	11.5240	12.7985	11.5298	12.8058
7	11.5249	12.8085	11.5220	12.8018	11.5277	12.7961	11.5250	12.7895
8	11.5269	12.7967	11.5128	12.7857	11.5309	12.8062	11.5168	12.7951

Since the multidimensional polynomials where constructed using shifted Legendre polynomials up to $p = 5$, the rule of thumb for choosing a suitable grid tells us that the grid should at least be based on $n_l = 6$ (since $n_l > p$). Calculations have been carried out on sparse grids based on the GQU and KPU method and for both methods n_l has been chosen from $n_l = 5$ up to $n_l = 8$. In Table 4.3 the results of the sampling method and the deterministic method to determine the first- and second-order moments are tabulated for the first and sixth radionuclide in compartment 4. If we compare these results with the results in Table 4.1 it can be seen that the results obtained by the PC method are in agreement within 1% with the Monte Carlo results.
If we look at the results for the KPU $n_l = 5$ case and compare the deterministic and sampled standard deviation of the sixth radionuclide ($k = 6$) with each other, we see that those values do not correspond at all. This can be described to the fact the the higher order coefficients that can not be accurately be calculated using this KPU $n_l = 5$ grid. We see that the evaluated standard deviation for radionuclide 6 ($k = 6$) computed with a KPU $n_l = 5$ grid via the sampling method is not in agreement with the value computed via the deterministic method. In Fig. 4.3 and Fig. 4.4 the pdf's of the first ($k = 1$) (Fig.4.3) and the sixth ($k = 6$) (Fig.4.4) radionuclide are plotted using those sparse grids.

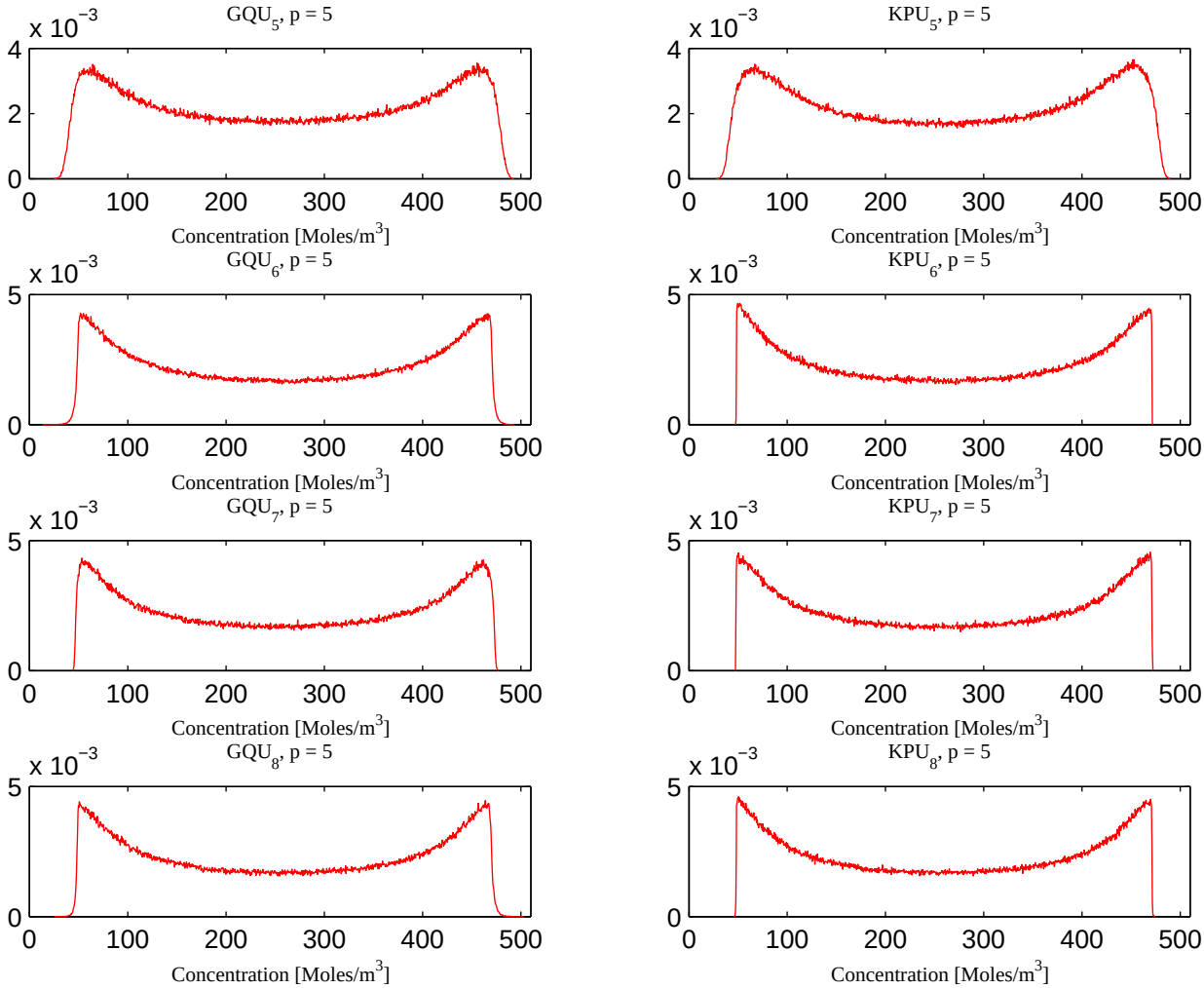


Figure 4.3: pdf's of the first radionuclide in compartment 4 at 100 years, computed with different sparse grids. The four left pdf's were computed by the GQU sparse grid for $n_l = 5$ up to $n_l = 8$. The four pdf's at the right were computed by using the KPU sparse grid for $n_l = 5$ up to $n_l = 8$.

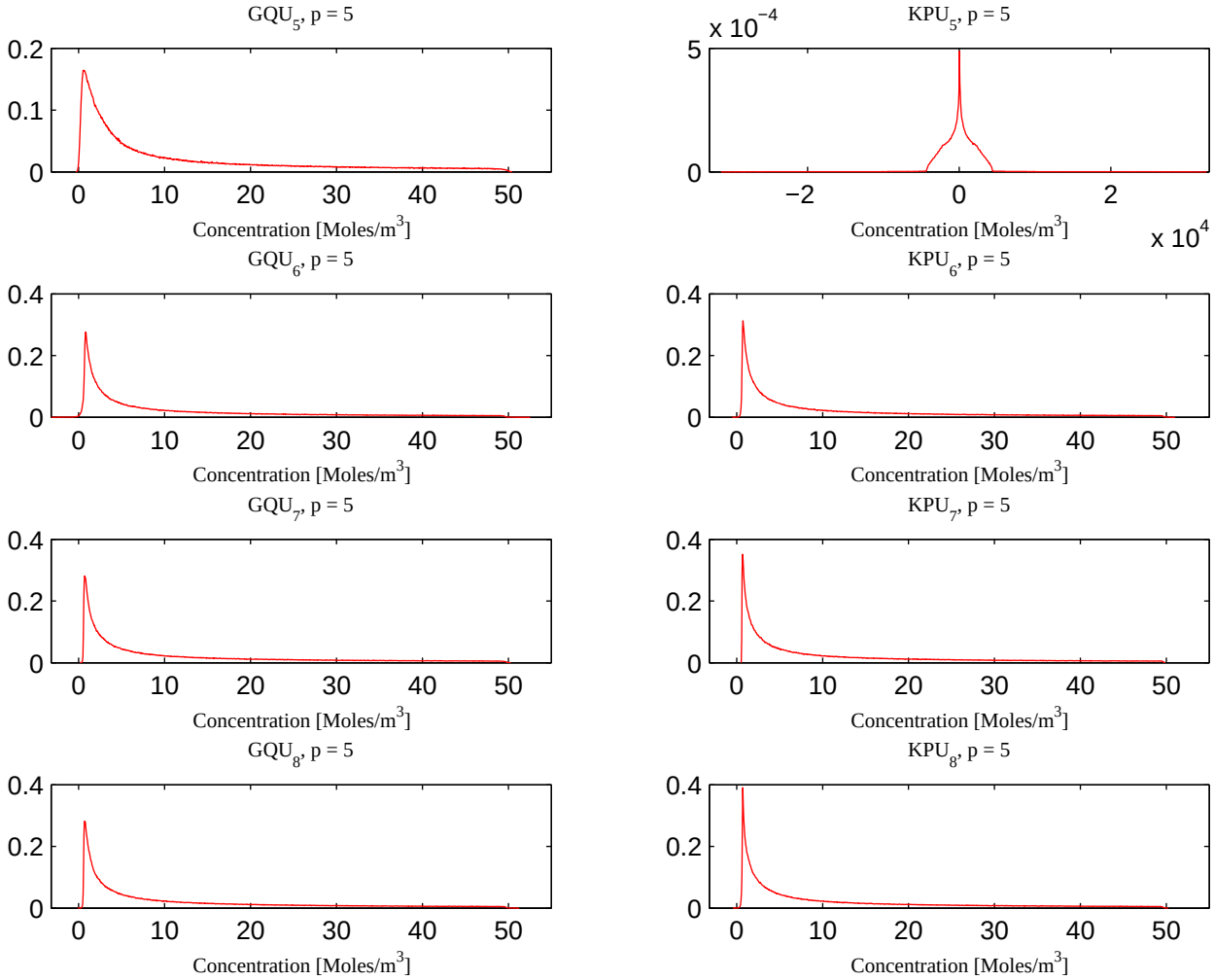


Figure 4.4: pdf's of the sixth radionuclide in compartment 4 at 100 years, computed with different sparse grids. The four left pdf's were computed by the GQU sparse grid for $n_l = 5$ up to $n_l = 8$. The four pdf's at the right were computed by using the KPU sparse grid for $n_l = 5$ up to $n_l = 8$.

From Fig.4.3 and Fig.4.4 we can see that the shapes slightly differ between different GQU calculated pdf's meaning that the coefficients change between the different GQU sparse grids. The pdf based on GQU_5 has a smooth shape, whereas the pdf's calculated with higher GQU grids the pdf's become less smooth, which means that the pdf's are still converging. For the pdf's computed with the use of the KPU grids we see that the first pdf, based on KPU_5 is also smooth but the pdf based on KPU_6 does not has this smoothness around the edges. For higher order KPU grids we see that there is no real differences between pdf's for different grids. We can thus say that the pdf of the first radionuclide is converged at the KPU_7 calculation.

For the pdf's of the concentration of the sixth radionuclide we see similar behaviour for the GQU based results. This indicates that the 'low' GQU grids are not capable of computing the higher order PC coefficients, where as if the number of nodes in the grid is increased we see the shape changing due to a more accurate computation of the higher order PC coefficients.
If we compare the pdf computed with the KPU_5 grid with the other pdf's we see that this pdf does not have the same shape at all. This suggests that the KPU_5 not only fails in determining the higher order PC coefficients, but also is not capable of determining the moderate to low order coefficients. From $n_l = 6$ the KPU grid is capable of accurately determining the PC coefficients, and again this can be seen from the plots in 4.4, since the shapes of the pdf's between KPU_6 to KPU_8 hardly change. In Appendix B the pdf's of all radionuclides in compartment 4 at time 100 years and 100.000 years and also a brief description of how PC is implemented can be found.

4.8 Conclusions

The non-intrusive Polynomial Chaos technique is used to analyse uncertainties corresponding to uncertainties in the distribution coefficients of the compartment model for radionuclide release into the biosphere. It has been shown that by using sparse grids the *curse of dimensionality* can be reduced.
The PC results (mean and standard deviation) obtained with a low level grid ($n_l = 5$) are within 1% agreement with the results obtained with Monte Carlo sampling using 1000 samples. On the other hand, the pdf's constructed with these low level grids are very unreliable, this due to errors in the higher order coefficients.
For all GQU as well as KPU sparse grids we see that the convergence of the mean and the standard deviation are within 1%, compared with the Monte Carlo sampling using 10^5 samples.
The amount of computational resources necessary for PC are, for the low level sparse grids, roughly two to four times the amount necessary for a Monte Carlo (MC:1000) calculation were 1000 samples are used. If the computational resources of seven to eighteen times the MC:1000 resources are accepted we see that the the pdf's, computed with KPU $n_l = 6$ and $n_l = 7$, become much more reliable. If such pdf's where to be build by using Monte Carlo, around 1000 times the MC:1000 resources are necessary.

5 Analysis of a repository design in a rock salt formation

The Netherlands currently considers a deep geological storage for nuclear waste. For the selection of suitable disposal sites, performance assessments need to be carried out in order to qualify and demonstrate the safety of the engineered barriers as well as the natural barrier which are site specific. This can be done by the means of a mathematical model which mimics the transport of the radionuclides through the barriers into the biosphere.
The Nuclear Research and Consultancy Group (NRG) carried out such a performance assessment with the use of EMOS (version 4 ccm2 NRG) , which is a computer program designed by GRS, NRG and ECN. This performance assessment has been carried out within the PAMINA (Performance Assessment Methodologies in Application to Guide the Development of the Safety Case) project.
The EMOS computer program consists of large number of modules, the three most important modules are; The REPOS module, which models the galleries of the repository including engineered barriers and releases from different types of waste canisters and waste forms. The MASCOT/TROUGH module which models the radionuclide transport in the geosphere. The EXPOS module calculates the exposure in the biosphere for which several exposure pathways have been taken into account to compute the maximum dose rate for individuals or for population groups [1, p. 47]. In [1] NRG considered two cases to perform a safety assessment of a nuclear waste repository situated in a salt rock, a so called dry scenario and a scenario where the repository is flooded, a so called brine intrusion scenario.
The brine intrusion scenario can occur when the waste repository has been abandoned and that the surrounding area and repository has been flooded. In the dry scenario it has been assumed that the due to the absence of brine no diffusion takes place.
In the brine intrusion scenario the permeability of the seal plug has not yet been converged and thus it does not fully seal the borehole. Therefore salt water or brine can enter the borehole where this can take up radionuclides. Eventually this contaminated brine can be pressed out the borehole again due to the surrounding pressure of the host rock and the heat produced by the waste. The radionuclides can now be dispersed not only by a diffusion process, but also due to an advection process since they are dissolved in water.
In both scenarios it has also been assumed that not only the waste canister but also the matrix in which the waste is stored has been compromised. This can be the result from the exposure to pressure from the host rock, corrosion or due to heat produced by the waste. The failure of the waste canisters is a conservative assumption since the waste canisters are designed to withstand extreme pressures.

5.1 Uncertainties in model parameters

As mentioned before, a mathematical model is used to mimic the transport of radionuclides within and around the repository. Most of the parameters used in this model have been derived from experiments and usually a *best fit* is used to determine the parameter value. However for some parameters, insufficient, not well distributed or highly varying data is only available which may lead to uncertainties in the parameter due to the fact that this *best fit* procedure is not accurate enough.
For the UQ of a salt rock repository described in [1], 6 parameters related to the porosity-permeability relation of the salt plug and the creep behaviour of the brine are considered to be uncertain not only due to

facts described above, but also because it is assumed that these parameters will have a large influence on the outcome. These parameters are: The recrystallization creep constant, the pressure creep constant, the permeability, the threshold porosity the threshold permeability and a experimentally determined constant of the salt plug. A description of these parameters can be found in [1, p. 32]. In Table 5.1 the parameters and the associated distributions are listed. One parameter is log-normal distributed, one normal and four parameters are log-uniform distributed.

Table 5.1: Distribution and values of the 6 random variables used

parameter		distribution	mean value	standard deviation
permeability	$[\mathrm{m}^2]$	log-normal	$1.39 \cdot 10^{-10}$	$1.65 \cdot 10^{-10}$
experimentally determined constant	$[-]$	normal	4.18	0.258
parameter		**distribution**	**minimum value**	**maximum value**
pressure creep constant	$[\mu\mathrm{m}^2\mathrm{K}(\mathrm{MPa}^{-1}\mathrm{s}^{-1})]$	log-uniform	0.22	3.11
recrystallization creep constant	$[\mu\mathrm{m}^2(\mathrm{MPa}^{-5}\mathrm{s}^{-1})]$	log-uniform	16.68	2501.0
threshold porosity	$[-]$	log-uniform	$7.9 \cdot 10^{-4}$	$3.16 \cdot 10^{-2}$
threshold permeability	$[\mathrm{m}^2]$	log-uniform	$1.0 \cdot 10^{-26}$	$6.31 \cdot 10^{-17}$

5.2 Standard Techniques used for Uncertainty Quantification

NRG applied the Monte Carlo technique to quantify the uncertainty of the modelled closure behaviour of the salt plug by calculating the porosity and permeability of the salt plug. Subsequently the amount of brine pressed out of the disposal cell, fluxes of radiotoxicity into the geosphere and dose rates in the biosphere are calculated. Therefore the 6 parameters have been sampled independently of each other 1000 times.
For the uncertainty analysis of the output variables it has been chosen to study the logarithm of the raw output data. This due to large the dispersion of the results and a lack of symmetry [1, p. 37]. The following measures of the generated outputs have been studied;

- Geometric mean μ_g and median m were used as central tendency.
- Geometric standard deviation to characterise the dispersion in logarithmic scale.
- Quantiles were used to characterise the dispersion of the outputs. Quantiles are points taken at (regular) intervals from the cumulative distribution function (cdf).
- Skewness S to provide a measure of symmetry.
- Kurtosis K to provide a measure of 'peakedness'.

Definition of the geometric mean,

$$\mu_g = \sqrt[n]{\prod_{i=1}^{n} x_i}, \tag{5.1}$$

median,

$$P(X \leq m) \geq \frac{1}{2} \text{ and } P(X \geq m) \leq \frac{1}{2}, \tag{5.2}$$

geometric standard deviation,

$$\sigma_g = e^{\sqrt{\frac{\sum_{i=1}^{n} \ln\left(\frac{x_i}{\mu_g}\right)^2}{n}}}, \tag{5.3}$$

skewness,

$$S = E\left[\left(\frac{X - E[X]}{\sigma}\right)^3\right], \tag{5.4}$$

kurtosis

$$K = \frac{E\left(X - E[X]\right)^4}{\left(E\left(X - E[X]\right)^2\right)^2} \tag{5.5}$$

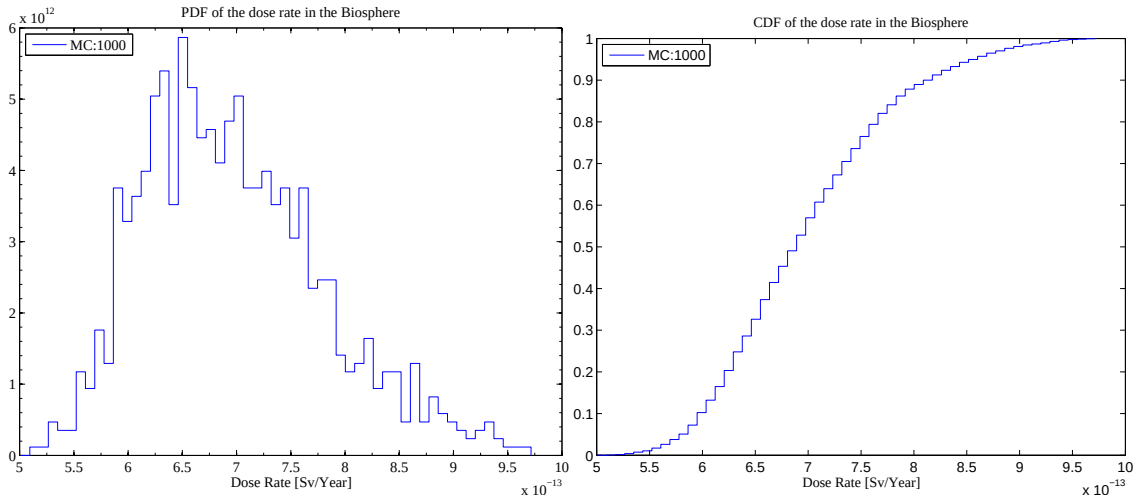
As mentioned before, the EMOS code is capable of computing a large number of different quantities. In [1] a description, and uncertainty analysis of a number of the computed quantities can be found. In the presented work we only focus on the dose rate over time in the biosphere in the brine intrusion scenario.
The code calculates at different times the dose rate in the biosphere, the maximum value of this dose rate over time is analysed. In Table 5.2 estimators of the distribution of the maximum dose rate computed by NRG are listed.

Table 5.2: The different measures of the dose rate in the biosphere calculated by Monte Carlo sampling for 1000 realizations. Taken at the dose rate peak-time. Results have been taken from [1]

N	Geometric mean [Sv/Year]	$\sigma_{geometric}$ [-]	Median [Sv/Year]	Skewness [-]	Kurtosis [-]
1000	6.84e-13	1.13	6.79e-13	0.329	2.758
	Quantiles [Sv/Year]				
N	1%	5%	50%	95%	99%
1000	5.40e-13	5.74e-13	6.79e-13	8.51e-13	9.23e-13

From Table 5.2 it can be seen that the geometric mean is around $6.8 \cdot 10^{-13}$ Sv/year, that the distribution has a skewness of 0.33 and a kurtosis of around 2.76. This means that the distribution is asymmetric with a tail towards the positive direction.
From the obtained results it is also possible to construct a pdf. For this purpose, a range has been defined by finding the minimum and maximum value of the 1000 realisations and dividing it in 55 equi-distant bins. The pdf can then be constructed by counting how many realisations fall within one specific bin and the number of realizations per bin is normalized by dividing it by the total number of realizations. A cumulative distribution function (cdf) can be constructed in a similar way. The pdf and cdf can be found in Fig. 5.1a, as it can be seen the pdf contains a considerable amount of noise, this is due to the relative low number of

realisations (1000). Furthermore it can be seen that the distribution of the realisations is asymmetric, as was predicted by the calculated skewness.



(a) pdf of the maximum dose rate

(b) cdf of the maximum dose rate

Figure 5.1: pdf and cdf of the maximum dose rate based on 1000 realizations.

The performance assessment of a salt-rock repository carried out by NRG is documented in [1]. In this document (p.97) a missing value problem is discussed, this problem is explained as follows:

During the Monte Carlo sampling of the EMOS code a porosity threshold is sampled. If during a run this threshold not has been reached within 10^6 years the run will be terminated and no outputs are available. The may influence the Monte Carlo results when this problem occurs for a relative high number of simulations compared the the total number of simulations.

For the 1000 simulations used in [1] this problem has been observed twice. Due to the robustness of the Monte Carlo technique the results are approximately the same if a low number of simulations is missing.

5.3 Polynomial Chaos Applied to the EMOS code for UQ

The UQ performed by NRG was based on 1000 realisations. Since one realisation is relatively expensive from a computational point of view, 1000 realisations is certainly computational expensive. If one could lower the number of realisations then the computational resources used or the computation time can be reduced.

A reduction of the number of realisations could be achieved by applying the Polynomial Chaos technique.

To this end we assume that a specific outcome or solution of the model, in this case the maximum dose rate in the biosphere, can be cast in the following form;

$$\dot{D}_B\left(t,\bar{\theta}\left(\bar{\xi}\right)\right)=\sum_{i=0}^{N_o}c_i\left(t\right)\Psi_i\left(\bar{\xi}\right), \tag{5.6}$$

where $\dot{D}_B\left(t,\bar{\theta}\left(\bar{\xi}\right)\right)$ is the dose rate at time t which depends on the stochastic processes $\bar{\theta}$ whose distribution is depending on the random variables $\bar{\xi}$. The PC coefficients are expressed by c_i and the multidimensional polynomials are represented by Ψ.
The stochastic processes modelled by $\bar{\theta}$ are the processes listed in Table 5.1 in which the distribution of the stochastic processes is also specified.

5.3.1 Choice of Polynomials, Truncation Order and Integration Grid

The shifted Legendre polynomials, valid on the domain $[0,1]$, have been used to build the multidimensional polynomials. The reason why it has been chosen to use Legendre polynomials is because four of the six stochastic processes are log-uniform distributed and the weight function involved in the orthogonal relationship of Legendre polynomials is the closest to a log-uniform distribution. Therefore the rate of convergence would be the highest or optimal if Legendre polynomials were used [3]. Furthermore the log-uniform distributed processes are only valid on a specific bounded domain, whereas the distributions of the other distributions are valid on a infinite or semi-infinite domain. So if a different polynomial basis associated with another distribution were used, the rate of convergence would possibly be compromised and thus more terms in the expansion would bee needed to achieve the same results.

Usually if the polynomials are correctly chosen the rate of convergence would be optimal, i.e. a low(er) truncation order is necessary. However in general one does not know the optimal order of truncation beforehand. Furthermore it has been shown in [15] that for time dependent problems the convergence may change over time, making it more difficult to define an appropriate truncation order.
It is also expected that the use of only one family of orthogonal polynomials deteriorates the rate of convergence.

It has been decided to investigate different truncation orders ($p=3$, $p=4$ and $p=5$), in combination with the GQU- and KPU-sparse grid introduced in paragraph 3.8.3.3. We have seen in paragraph 4.7 that the KPU grid performs better that the GQU grid and we would expect that the KPU grid would perform better also in this case. With the given truncation order (p is 3,4,and 5) we can define the level (n_l) on which we can build the sparse grids. And since n_l should obey $n_l > p$, we chose for the GQU grid $n_l = 3$ and $n_l = 4$. For the KPU grid we chose $n_l = 3$ up to $n_l = 6$, this way the convergence of the grid could be studied.

5.3.2 UQ Results obtained by applying PC

In paragraph 5.2 we discussed the measures used to study and to quantify the uncertainty of the predicted dose rate and in the presented work we have used the same measures. As mentioned in paragraph 3.6 it is possible to compute statistical moments based on the PC coefficients. We have seen that this was relatively easy for the fist and second order moments (arithmetic mean and standard deviation). Here we also need to compute other measures such as the geometric mean and the geometric standard deviation. These estimators can be computed by sampling $\bar{\xi}$ in the right hand side of Eq. (5.6) uniformly between $[0,1]$ and this sampling procedure requires little computational resources.
In Table 5.3 the results for different orders of truncation and grids are shown.
In the first line of Table 5.3 the results from NRG are shown. The geometric mean calculated via PC depending on different grids and truncation orders correspond with the geometric mean calculated via MC. By the sampling of the right hand side of Eq. (5.6) 10^6 realizations were obtained. We see that for all truncation orders and grid types, except GQU $n_l = 4$, $p = 3$, the geometric mean calculated by means on a PC expansion are within 3% accuracy of the geometric mean obtained with MC. For increasing truncation order it can be seen that the geometric standard deviation increases. Furthermore the skewness and kurtosis increase for increasing order of truncation. From this can be concluded that the distribution still changes for increasing order of truncation. This can also be seen in Table 5.4 in which the quantiles are given.

Table 5.3: Statistics of the maximum dose rate in the Biosphere for the wet scenario, with various numerical integration techniques.

Grid type	n_l	p	N	Geometric mean [Sv/Year]	$\sigma_{geometric}$ [-]	Median [Sv/Year]	Skewness [-]	Kurtosis [-]
NRG:MC	[-]	[-]	1000	6.84e-13	1.13	6.79e-13	0.329	2.758
GQU	3	3	85	6.98e-13	1.16	7.06e-13	-0.297	2.658
GQU	4	3	389	6.61e-13	1.27	6.78e-13	0.462	3.006
KPU	3	3	79	6.97e-13	1.16	7.05e-13	0.484	2.957
KPU	4	3	257	7.04e-13	1.18	7.14e-13	0.909	3.975
KPU	5	3	737	6.94e-13	1.20	6.80e-13	1.731	9.227
KPU	5	4	737	6.91e-13	1.23	6.84e-13	1.848	11.524
KPU	6	3	1889	7.01e-13	1.19	6.88e-13	1.390	6.678
KPU	6	4	1889	7.00e-13	1.20	6.89e-13	1.716	9.355
KPU	6	5	1889	6.97e-13	1.22	6.89e-13	1.989	12.781

Table 5.4: Quantiles for the maximum dose rate in the Biosphere for the wet scenario, with various numerical integration techniques.

				Quantiles [Sv/Year]				
Grid type	n_l	p	N	1%	5%	50%	95%	99%
NRG	[-]	[-]	1000	5.40e-13	5.74e-13	6.79e-13	8.51e-13	9.23e-13
GQU	3	3	85	4.57e-13	5.28e-13	7.06e-13	8.63e-13	9.03e-13
GQU	4	3	389	3.75e-13	4.46e-13	6.78e-13	9.66e-13	1.10e-12
KPU	3	3	79	5.05e-13	5.49e-13	7.05e-13	8.98e-13	9.90e-13
KPU	4	3	257	5.13e-13	5.51e-13	7.14e-13	9.49e-13	1.08e-12
KPU	5	3	737	4.76e-13	5.31e-13	6.80e-13	9.51e-13	1.21e-12
KPU	5	4	737	4.31e-13	5.08e-13	6.84e-13	9.58e-13	1.26e-12
KPU	6	3	1889	4.94e-13	5.45e-13	6.88e-13	9.45e-13	1.62e-12
KPU	6	4	1889	4.79e-13	5.35e-13	6.89e-13	9.55e-13	1.22e-12
KPU	6	5	1889	4.48e-13	5.20e-13	6.89e-13	9.74e-13	1.28e-12

Since the solution is sampled to compute the measures given in Tables 5.3 and 5.4 it is also possible to construct a pdf. The pdf's for the nine different cases are depicted in Fig. 5.2 together with the pdf computed via MC. In Fig. 5.2a the pdf's calculated on a GQU grid are plotted with truncation orders 3 and 4. In Fig. 5.2b, 5.2c and 5.2d the pdf's are plotted for truncation orders 3, 4 and 5 and for each plot different KPU grids were used.

In Fig. 5.2a we see that the computed pdf's do not agree with the MC pdf. This can be due to the fact that the truncation order is not high enough or the grid used is not accurate enough. In Fig. 5.2b a truncation order of 3 was used together with various KPU grids. We see that the pdf's do agree with the MC pdf. In Fig. 5.2c and 5.2d we see that the shape does agree, but the maximum does not, this can be describe due to the fact that these pdf's have longer tails.

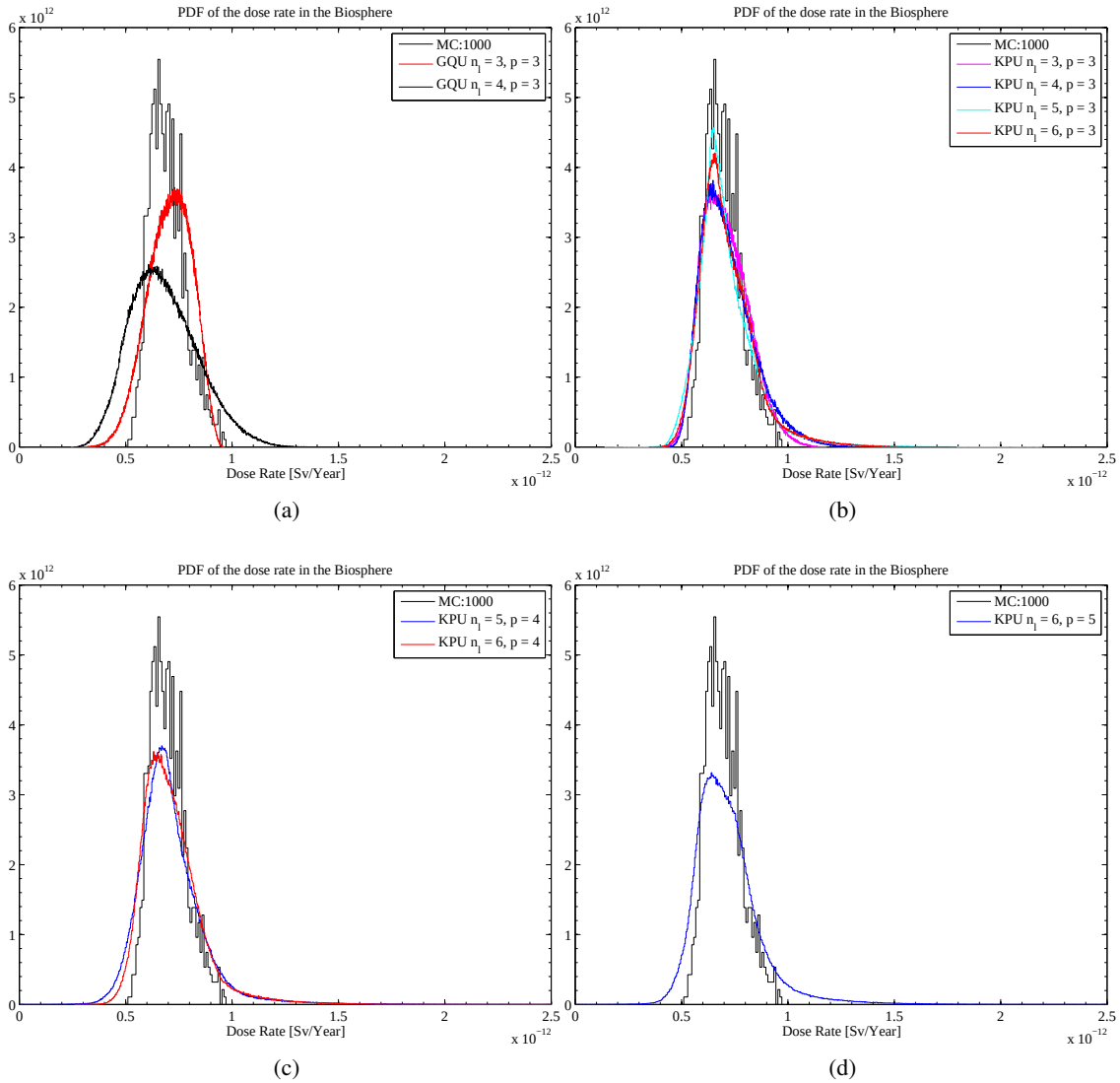
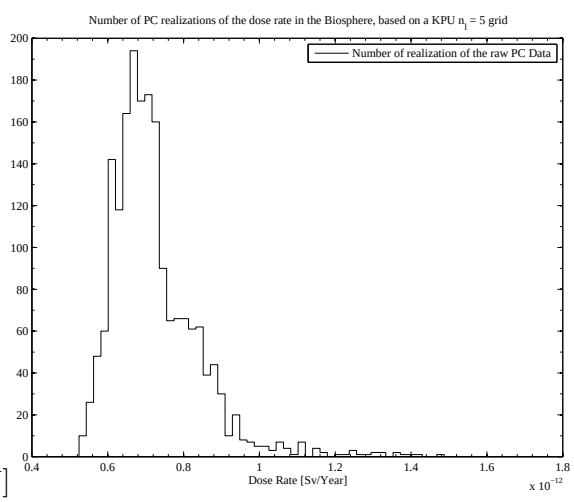


Figure 5.2: (a) pdf of the maximum dose rate computed for $p = 3$ on a GQU $n_l = 3$ and $n_l = 4$ grid. (b) pdf of the maximum dose rate computed for $p = 3$, $p = 4$, $p = 5$ and $p = 6$ on a KPU $n_l = 3$ grid. (c) pdf of the maximum dose rate computed for $p = 5$ and $p = 6$ on a KPU $n_l = 4$ grid. (d) pdf of the maximum dose rate computed for $p = 6$ on a KPU $n_l = 3$ grid.

From the tabulated results in Table. 5.3 and Table. 5.4 as well as the graphs in Fig.5.1a we can conclude that if the GQU grids are used, the method produces unreliable results. The KPU grids in contrast to the GQU grids do produce reliable results as can be seen in 5.1a. In appendix C in C.1 the cdf's can be found.

As mentioned before if $p = 3$ is chosen we see that the obtained solution with PC suits the results obtained with MC. If higher orders of truncation are chosen the solution changes and the distribution gets some what longer tails. Furthermore if we look at Fig. 5.2c we see that the difference between the two PC pdf's is negligible, this suggests that a KPU $n_l = 5$ grid is sufficiently accurate for a fourth order truncation. But yet there is a large discrepancy between the MC pdf and the PC pdf's. This discrepency can also be seen in 5.2d, and at first there is not a clear explanation for this discrepancy.
We see that the MC pdf is bounded between approximately $0.5 \cdot 10^{-12}$ and $1.0 \cdot 10^{-12}$ Sv/Year. However, if we look at the raw PC data, which is the data on which the Legendre polynomials are projected we see that in fact there are realizations which are located outside this MC domain. This can clearly be seen in Fig. 5.7, were we plotted a histogram of the raw PC-data computed with a KPU $n_l = 6$ grid. Also for the other KPU grids, there is seen that there are realizations with a value higher than approximately $1.0 \cdot 10^{-12}$ Sv/Year, and this explains why the PC-pdf's do have longer tails compared to the MC-pdf. And if we look at Table 5.4 we can see from the 95% quantile that the values which are higher than $1.0 \cdot 10^{-12}$ represent approximately 5% of the total data.

Assuming that the EMOS code not has been changed in between MC and PC calculations, the difference in output can only be explained by differences in the type of sampling.
A possible explanation could be that for Monte Carlo 1000 samples are used, which is maybe to low to capture outputs higher the $1.0 \cdot 10^{-12}$. On the other hand, this data represents about 5% of the total data and so for 1000 samples we would expect about 50 realizations to be higher than $1.0 \cdot 10^{-12}$ and so this explanation does not seems very likely.
An other possibility maybe that the input values (for example the distribution, mean values or standard deviations) ,in the MC- or PC-case, are not properly defined, and if this is the case than it would be likely that the distribution of the output will change.
Unfortunately we could not find a definite cause of the problem.



$[H]$ (5.7)

In paragraph 5.2 we discussed that due to the robustness of Monte Carlo that (a low number of) missing values do not influence the the the results. This is however not the case for the Polynomial Chaos technique.

Fortunately this missing value problem did not occur during the PC calculations. But in general it is possible that this problem may occur. If code, in this case the EMOS code, can not be fixed, for instance because it is considered as a black-box, then one should come up with a substitution procedure for the missing value. In general there does not exist a straightforward approach for such a substitution procedure.

5.4 Conclusion

Uncertainties in the dose rate in the biosphere due to uncertainties in closure behaviour of the salt plug have been investigated. A computer code, developed by NRG was used for this end. The ”traditional” Monte Carlo approach to quantify the uncertainties has been used by NRG and for this analysis 1000 calculations were necessary to achieve an accuracy of approximately 3%. For this accuracy, which is related the the number of calculations a relative high amount of computational resources is required.
A Polynomial Chaos expansion to model the stochastic behaviour of the dose rate has been proposed in order to reduce the computational resources necessary.
For this expansion the shifted Legendre polynomials were used and we investigated the convergence of the expansion for different orders of truncation.
We have seen that if polynomials up to 3rd order are used, together with a KPU $n_l = 3$ grid, the results are in agreement with the results of the Monte Carlo approach. For this specific case ($p = 3$, KPU $n_l = 3$) only 79 grid points are necessary, which means that with less than 10% of the computational resources of the Monte Carlo approach the same accuracy can be met.

The Monte Carlo approach failed to capture realizations which are higher than approximately $1.0 \cdot 10^{-12}$ Sv/Year, where the PC-approach did found realizations higher than this value. And according to the PC-approach these values represent approximately 5% of the distribution.

6 Discussion and Recommendations

For the UQ of stochastic systems Polynomial Chaos can be applied non-intrusive. Unfortunately for high dimensional stochastic problems the amount of computational resources increase exponentially. This exponentially increase can be reduced by making use of sparse grids.

If one possesses over an effective and efficient method to find the PC-coefficient then the statistics of the problem can be found analytically or via a sampling method. Both method are from a computational perspective inexpensive. This makes PC an attractive alternative for UQ compared to other techniques like Monte Carlo or Perturbation method, this is especially the case for low dimensional stochastic problems.
Unfortunately the amount of computational resources necessary to find the PC-coefficients scales exponentially with the number of random input variables.
Furthermore the amount of resources scales with the amount of terms in the Polynomial Chaos expansion, and therefore one would like to determine beforehand the amount of terms which would be sufficient with respect to a error margin. Unfortunately, a method to determine an optimal number of terms beforehand does not exist, and this could lead to that one takes more terms in the expansion than needed just to be on the safe side.

If one possesses over an adaptive method to find the optimal truncation order one could reduce the computation time. One of the problems one may find when investigating or applying such methods is the lack of good and efficient convergence estimators.
Furthermore an adaptive method to find the optimal number of points used for the numerical integration can reduce the amount of computational resources.

References

[1] T. Schröder, J. Hart, A. Costescu-Badea, and R. Bolado Lavin, "Pamina task 2.1.ṭechniques for sensitivity and uncertainty analysis; analysis of a repository design in a rock salt formation." 2009.

[2] N. Wiener, "The homogeneous chaos," *American Journal of Mathematics*, vol. 60, no. 4, pp. 897–936, 1938.

[3] D. Xiu and G. E. Karniadakis, "The wiener–askey polynomial chaos for stochastic differential equations," *SIAM J. Sci. Comput.*, vol. 24, pp. 619–644, February 2002.

[4] R. Askey and J. Wilson, "Some basic hypergeometric orthogonal polynomials that generalize jacobi polynomials," *Memoirs Amer. Math. Soc. 319*, 1985.

[5] D. Xiu and G. E. Karniadakis, "The wiener–askey polynomial chaos for stochastic differential equations," *SIAM J. Sci. Comput.*, vol. 24, pp. 619–644, February 2002.

[6] R. H. Cameron and W. T. Martin, "The orthogonal development of non-linear functionals in series of fourier-hermite functionals," *The Annals of Mathematics*, vol. 48, no. 2, pp. 385–392, 1947.

[7] A. W. Hagues, M. M. R. Williams, and M. D. Eaton, "A probabilistic study of the effect of retardation factor uncertainty using a compartment model for radionuclide release into the biosphere," *Annals of Nuclear Energy*, vol. 37, no. 9, pp. 1197 – 1207, 2010.

[8] O. Le Maitre and O. Knio, *Spectral Methods for Uncertainty Quantification With Applications to Computational Fluid Dynamics*. Springer, 2010.

[9] S. Smolyak, "Quadrature and interpolation formulas for tensor products of certain classes of functions," *Dolk. Akad. Nauk SSSR*, vol. 4, pp. 240 – 251, 1963.

[10] T. Gerstner and M. Griebel, *Encyclopedia of Quantitative Finance*. Wiley, 2008.

[11] K. Petras, "Smolyak cubature of given polynomial degree with few nodes for increasing dimension," *Numerische Mathematik*, vol. 93, pp. 729–753, 2003.

[12] G. de Marsily, "Quantative hydrogeology," 1986.

[13] M. Williams, M. Thorne, J. Thomson, and A. Paulley, "Development of a solution method for the differential equations arising in the biosphere module of the bnfl's suite of codes mondrian," *Annals of Nuclear Energy*, vol. 29, pp. 1019 – 1039, 2002.

[14] E. Holzbecher, *Environmental Modeling Using MATLAB*. Springer, 2007.

[15] M. Gerritsma, J.-B. van der Steen, P. Vos, and G. Karniadakis, "Time-dependent generalized polynomial chaos," *Journal of Computational Physics*, vol. 229, no. 22, pp. 8333 – 8363, 2010.

A Monte Carlo

Monte Carlo is computation technique which can be used to perform uncertainty quantification. In this appendix a brief description is given how the technique is implemented. Furthermore the results obtained with the use of this technique are given. Relevant estimators of the concentration of all radionuclides in compartment 4 and 31 for time is 100 and 100.000 year have been tabulated. Next to this pdf's of the concentration in compartment 4 and 31 for time is 100 and 100.000 year have been plotted.

A.1 Implementation of the Monte Carlo method

A short description of the implementation of the Monte Carlo method to solve the stochastic differential equations concerning the compartment model.

- Do for N samples
  - Draw a random number ξ_n.
  - Plug ξ_n in the differential equation.
  - Solve the Differential equation for $C_i^k(t, \xi_n)$.
  - Collect realization $C_i^k(t, \xi_n)_n$.
- end do
- collect the statisics on $\{C_i^k(t, \xi_n)_n\}_{n=1}^N$, such as mean, standard deviation, standard deviation of the mean, and relative standard deviation.

A.2 Results

Monte Carlo Results for the radionuclide concentration in compartment 4.

Table A.1: Monte Carlo results for 1.000 samples at t = 100 years and t = 100,000 years for compartment 4

	100 years				100,000 years			
k	$E[C_4^k]$	σ_C	σ_{E_C}	$\sigma_{REL,C}$	$E[C_4^k]$	σ_C	σ_{E_C}	$\sigma_{REL,C}$
1	265.2288	138.4375	4.3800	0.5220	336.3776	174.5237	5.5217	0.5188
2	75.6467	79.7003	2.5216	1.0536	620.2280	632.9417	20.0254	1.0205
3	1.2359	1.4392	0.0455	1.1645	637.7993	717.0836	22.6875	1.1243
4	1.5803	1.7599	0.0557	1.1136	26.1481	29.1802	0.9232	1.1160
5	0.0911	0.1040	0.0033	1.1418	0.7973	0.9089	0.0288	1.1399
6	11.4379	12.5557	0.3972	1.0977	1294.9996	1345.6173	42.5734	1.0391
7	2.7792	3.0808	0.0975	1.1085	19.5892	21.7724	0.6888	1.1115
8	0.0103	0.0121	0.0004	1.1702	0.0371	0.0434	0.0014	1.1704

Table A.2: Monte Carlo results for 100,000 samples at t = 100 years and t = 100,000 years for compartment 4

	100 years				100,000 years			
k	$E[C_4^k]$	σ_C	σ_{E_C}	$\sigma_{REL,C}$	$E[C_4^k]$	σ_C	σ_{E_C}	$\sigma_{REL,C}$
1	260.3267	138.5777	0.4382	0.5323	330.1944	174.7105	0.5525	0.5291
2	80.0098	82.2543	0.2601	1.0281	654.9797	653.3507	2.0661	0.9975
3	1.2650	1.4550	0.0046	1.1502	651.8185	723.1592	2.2868	1.1094
4	1.6204	1.7716	0.0056	1.0933	26.7618	29.2974	0.0926	1.0947
5	0.0909	0.1053	0.0003	1.1583	0.7939	0.9209	0.0029	1.1600
6	11.4202	12.7023	0.0402	1.1123	1288.2039	1356.4362	4.2894	1.0530
7	2.7572	3.0849	0.0098	1.1189	19.4444	21.8078	0.0690	1.1215
8	0.0102	0.0118	0.0000	1.1564	0.0366	0.0424	0.0001	1.1582

Monte Carlo Results for the radionuclide concentration in compartment 31.

Table A.3: Monte Carlo Results for 1000 samples at t = 100 years and t = 100,000 years for compartment 31

	100 years				100,000 years			
k	$E[C_3 1^k]$	σ_C	σ_{E_C}	$\sigma_{REL,C}$	$E[C_3 1^k]$	σ_C	σ_{E_C}	$\sigma_{REL,C}$
1	170.9172	98.7026	3.1228	0.5775	224.7957	127.1866	4.0240	0.5658
2	45.1812	54.2043	1.7149	1.1997	459.9982	496.5328	15.7096	1.0794
3	0.3163	0.4622	0.0146	1.4614	496.4796	581.6302	18.4020	1.1715
4	0.4967	0.8013	0.0254	1.6132	18.8652	21.3620	0.6759	1.1324
5	0.0073	0.0134	0.0004	1.8354	0.5306	0.6081	0.0192	1.1462
6	4.3592	6.7488	0.2135	1.5482	1060.0823	1172.7089	37.1029	1.1062
7	0.6218	1.0595	0.0335	1.7039	13.6123	15.5095	0.4907	1.1394
8	0.0004	0.0009	0.0000	2.0388	0.0209	0.0245	0.0008	1.1713

Table A.4: Monte Carlo results for 100,000 samples at t = 100 years and t = 100,000 years for compartment 31

	100 years				100,000 years			
k	$E[C_4^k]$	σ_C	σ_{E_C}	$\sigma_{REL,C}$	$E[C_4^k]$	σ_C	σ_{E_C}	$\sigma_{REL,C}$
1	167.4248	98.7745	0.3124	0.5900	220.2939	127.2788	0.4025	0.5778
2	48.1580	55.8911	0.1767	1.1606	486.9753	512.1187	1.6195	1.0516
3	0.3278	0.4741	0.0015	1.4463	508.2403	588.0734	1.8597	1.1571
4	0.5132	0.8062	0.0025	1.5709	19.3553	21.5056	0.0680	1.1111
5	0.0073	0.0135	0.0000	1.8576	0.5300	0.6150	0.0019	1.1605
6	4.3858	6.8387	0.0216	1.5593	1058.4408	1186.2280	3.7512	1.1207
7	0.6180	1.0604	0.0034	1.7159	13.4973	15.5210	0.0491	1.1499
8	0.0004	0.0009	0.0000	2.0613	0.0206	0.0239	0.0001	1.1578

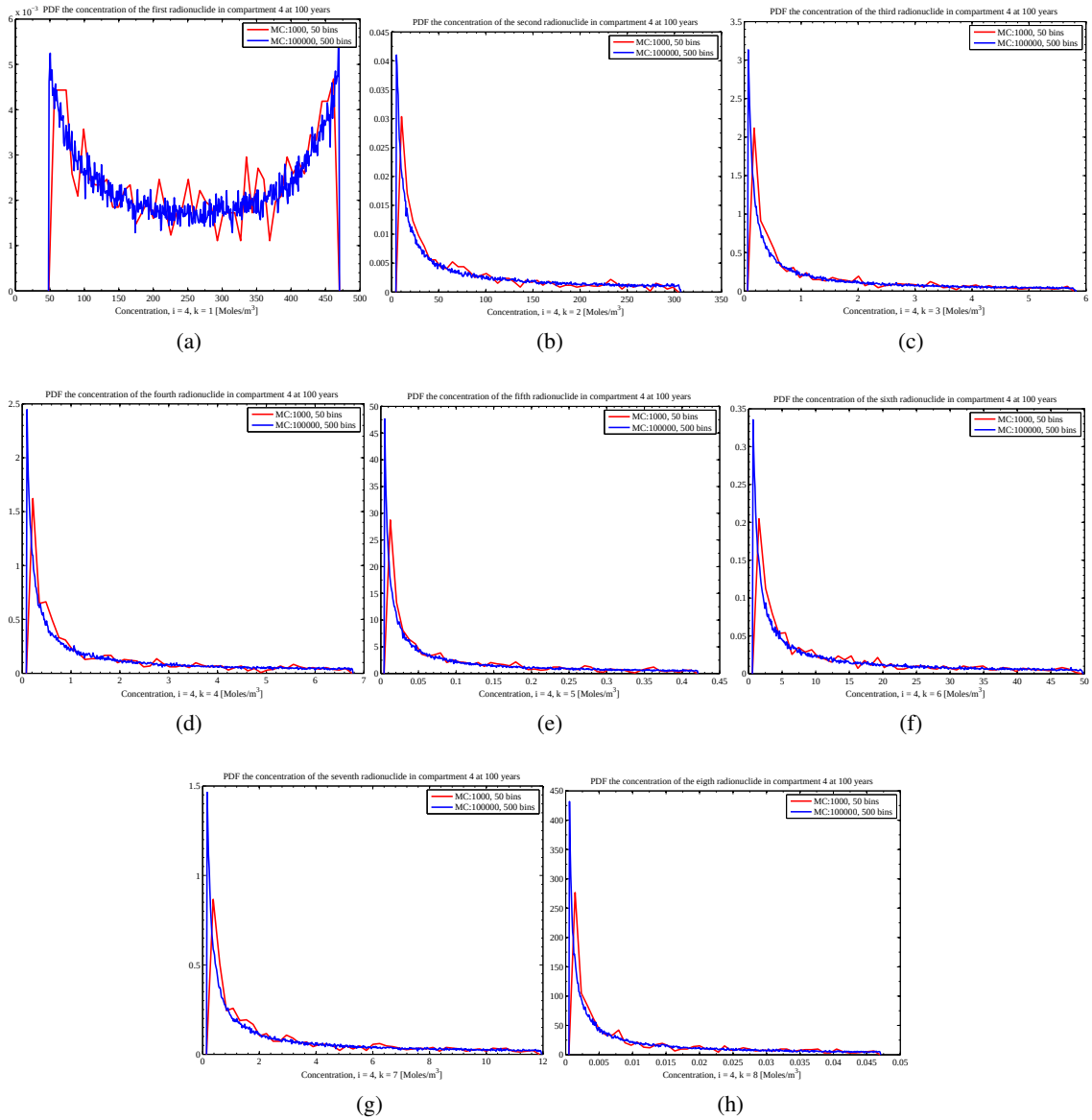


Figure A.1: The pdf's of all radionuclide in compartment 4 at 100 years.

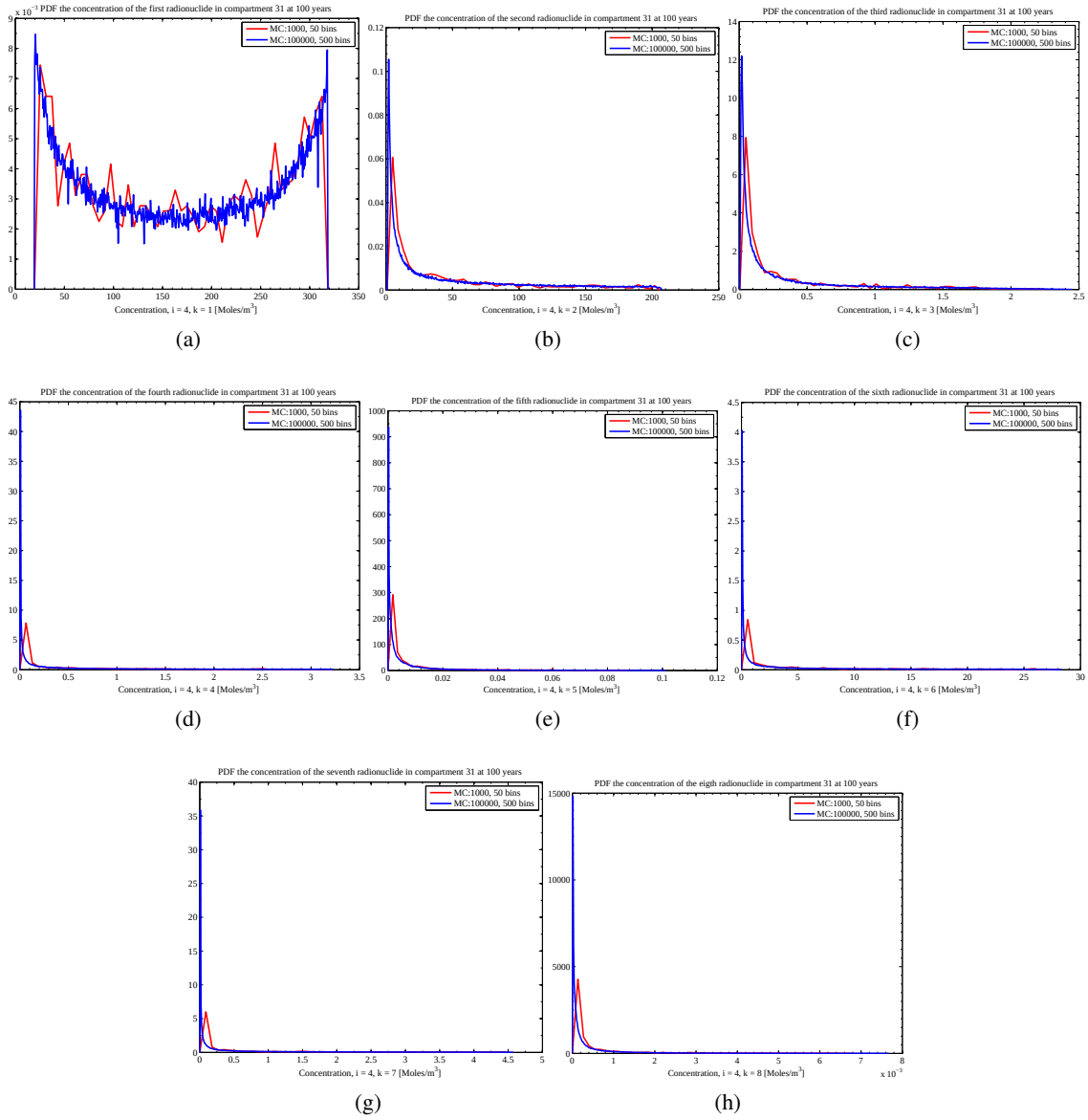


Figure A.2: The pdf's of all the radionuclide in compartment 31 at 100 years.

B Non Intrusive Spectral Projection Applied to the Compartment model

B.1 Implementation

A brief overview of how PC is implemented to perform UQ for the compartment model.

- Define a a set of polynomials which represents the stochastic space: Shifted Legendre polynomials have been chosen with a maximum order of $p = 5$. The multidimensional polynomials where build accordingly.
- Define deterministic sample set $\{\xi_i\}_{i=1}^{N}$. This sample set depends on the chosen numerical integration procedure.
- Solve the Differential equation for $C_i^k(t, \xi_n)$ for al deterministic samples.
- Project the set of deterministic solutions obtained onto the polynomial basis to determine the expansion coefficients.
- Reconstruct the solution by means of a expansion.
- Sample the expansion to obtain the statistics of the solution in the stochastic space.

B.2 Results

Table B.1: Mean and standard deviation of the radionuclide concentrations in compartment 4 at 100 years.

	GQU,N_Q=5		GQU,N_Q=6		GQU,N_Q=7		GQU,N_Q=8	
k	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	260.5571	139.1827	260.5448	138.8723	260.4654	138.6981	260.5071	138.7487
2	80.2344	82.7493	80.1566	82.7633	80.0344	82.5686	80.1986	82.7826
3	1.2654	1.4581	1.2662	1.4563	1.2651	1.4572	1.2653	1.4552
4	1.6258	1.7708	1.6257	1.7708	1.6256	1.7706	1.6256	1.7705
5	0.0914	0.1058	0.0914	0.1056	0.0914	0.1056	0.0914	0.1056
6	11.5285	12.8008	11.5317	12.8066	11.5249	12.8085	11.5269	12.7967
7	2.7449	3.0695	2.7448	3.0683	2.7446	3.0680	2.7447	3.0678
8	0.0102	0.0118	0.0102	0.0118	0.0102	0.0118	0.0102	0.0118

Table B.2: Mean and standard deviation of the radionuclide concentrations in compartment 4 at 100,000 years.

	GQU,N_Q=5		GQU,N_Q=6		GQU,N_Q=7		GQU,N_Q=8	
k	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	330.4380	175.4266	330.4369	174.9352	330.4371	174.9231	330.4370	174.9236
2	656.0483	657.4063	656.0552	656.6898	656.0545	656.6935	656.0545	656.6929
3	652.0699	723.9124	652.0715	723.7166	652.0715	723.7136	652.0715	723.7137
4	26.8511	29.2991	26.8514	29.2841	26.8514	29.2841	26.8514	29.2841
5	0.7984	0.9242	0.7984	0.9233	0.7984	0.9233	0.7984	0.9233
6	1299.2370	1367.4262	1299.2572	1365.9972	1299.2566	1366.0051	1299.2565	1366.0043
7	19.3503	21.6913	19.3505	21.6812	19.3505	21.6812	19.3505	21.6812
8	0.0365	0.0423	0.0365	0.0423	0.0365	0.0423	0.0365	0.0423

Table B.3: Mean and standard deviation of the radionuclide concentrations in compartment 31 at 100 years.

	GQU,N_Q=5		GQU,N_Q=6		GQU,N_Q=7		GQU,N_Q=8	
k	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	167.5932	99.2142	167.5861	98.9889	167.5256	98.8598	167.5566	98.8976
2	48.3361	56.2382	48.2828	56.2575	48.1954	56.1108	48.3118	56.2636
3	0.3279	0.4867	0.3284	0.4743	0.3280	0.4751	0.3279	0.4743
4	0.5149	0.8060	0.5149	0.8055	0.5148	0.8054	0.5148	0.8053
5	0.0074	0.0139	0.0074	0.0136	0.0074	0.0136	0.0074	0.0136
6	4.4482	6.9013	4.4550	6.9005	4.4500	6.9066	4.4484	6.8987
7	0.6123	1.0639	0.6124	1.0535	0.6122	1.0534	0.6123	1.0532
8	0.0004	0.0009	0.0004	0.0009	0.0004	0.0009	0.0004	0.0009

Table B.5: Mean and standard deviation of the radionuclide concentrations in compartment 4 at 100 years.

	KPU,N_Q=5		KPU,N_Q=6		KPU,N_Q=7		KPU,N_Q=8	
k	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	260.5074	138.9891	260.5074	138.7798	260.5244	138.7632	260.5321	138.7717
2	80.1572	82.7257	80.1572	82.6966	80.1815	82.6643	80.1913	82.7183
3	1.2646	1.4574	1.2649	1.4554	1.2655	1.4550	1.2660	1.4567
4	1.6256	1.7705	1.6256	1.7705	1.6256	1.7705	1.6257	1.7706
5	0.0914	0.1058	0.0914	0.1056	0.0914	0.1056	0.0914	0.1056
6	11.5225	12.7982	11.5240	12.7985	11.5277	12.7961	11.5309	12.8062
7	2.7445	3.0686	2.7446	3.0678	2.7447	3.0678	2.7448	3.0680
8	0.0102	0.0118	0.0102	0.0118	0.0102	0.0118	0.0102	0.0118

Table B.4: Mean and standard deviation of the radionuclide concentrations in compartment 31 at 100,000 years.

	GQU,N_Q=5		GQU,N_Q=6		GQU,N_Q=7		GQU,N_Q=8	
k	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	220.4761	127.8113	220.4761	127.4440	220.4761	127.4349	220.4761	127.4353
2	487.9757	515.1125	487.9818	514.8420	487.9816	514.8444	487.9816	514.8443
3	508.5077	589.6217	508.5066	588.5443	508.5066	588.5436	508.5066	588.5435
4	19.4160	21.4957	19.4160	21.4916	19.4160	21.4915	19.4160	21.4915
5	0.5329	0.6179	0.5329	0.6168	0.5329	0.6168	0.5329	0.6168
6	1068.6294	1196.3184	1068.6284	1195.7375	1068.6287	1195.7280	1068.6287	1195.7281
7	13.4332	15.4481	13.4332	15.4345	13.4332	15.4344	13.4332	15.4344
8	0.0206	0.0239	0.0206	0.0239	0.0206	0.0239	0.0206	0.0239

Table B.6: Mean and standard deviation of the radionuclide concentrations in compartment 4 at 100,000 years.

	KPU,N_Q=5		KPU,N_Q=6		KPU,N_Q=7		KPU,N_Q=8	
k	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	330.4371	175.1592	330.4371	174.9211	330.4371	174.9236	330.4371	174.9236
2	656.0542	657.2384	656.0542	656.6948	656.0545	656.6928	656.0545	656.6928
3	652.0716	723.8277	652.0716	723.7095	652.0715	723.7136	652.0715	723.7136
4	26.8514	29.2865	26.8514	29.2840	26.8514	29.2841	26.8514	29.2841
5	0.7984	0.9248	0.7984	0.9233	0.7984	0.9233	0.7984	0.9233
6	1299.2578	1366.4662	1299.2566	1366.0070	1299.2565	1366.0041	1299.2565	1366.0042
7	19.3505	21.6892	19.3505	21.6810	19.3505	21.6812	19.3505	21.6812
8	0.0365	0.0424	0.0365	0.0423	0.0365	0.0423	0.0365	0.0423

Table B.7: Mean and standard deviation of the radionuclide concentrations in compartment 31 at 100 years.

	KPU,N_Q=5		KPU,N_Q=6		KPU,N_Q=7		KPU,N_Q=8	
k	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	167.5567	99.0719	167.5567	98.9188	167.5695	98.9080	167.5750	98.9141
2	48.2842	56.2226	48.2842	56.1999	48.2998	56.1756	48.3075	56.2161
3	0.3277	0.4855	0.3278	0.4743	0.3280	0.4740	0.3284	0.4748
4	0.5148	0.8055	0.5148	0.8053	0.5148	0.8053	0.5149	0.8054
5	0.0074	0.0137	0.0074	0.0136	0.0074	0.0136	0.0074	0.0136
6	4.4472	6.8978	4.4485	6.8982	4.4497	6.8968	4.4527	6.9034
7	0.6122	1.0581	0.6122	1.0532	0.6123	1.0532	0.6123	1.0533
8	0.0004	0.0009	0.0004	0.0009	0.0004	0.0009	0.0004	0.0009

Table B.8: Mean and standard deviation of the radionuclide concentrations in compartment 31 at 100,000 years.

	KPU,N_Q=5		KPU,N_Q=6		KPU,N_Q=7		KPU,N_Q=8	
1	220.4761	127.6118	220.4761	127.4334	220.4761	127.4353	220.4761	127.4353
2	487.9815	515.0703	487.9815	514.8460	487.9816	514.8442	487.9816	514.8442
3	508.5065	589.3277	508.5065	588.5430	508.5066	588.5434	508.5066	588.5434
4	19.4161	21.4920	19.4160	21.4914	19.4160	21.4915	19.4160	21.4915
5	0.5329	0.6177	0.5329	0.6167	0.5329	0.6168	0.5329	0.6168
6	1068.6287	1195.9440	1068.6287	1195.7250	1068.6287	1195.7280	1068.6287	1195.7280
7	13.4332	15.4406	13.4332	15.4343	13.4332	15.4344	13.4332	15.4344
8	0.0206	0.0239	0.0206	0.0239	0.0206	0.0239	0.0206	0.0239

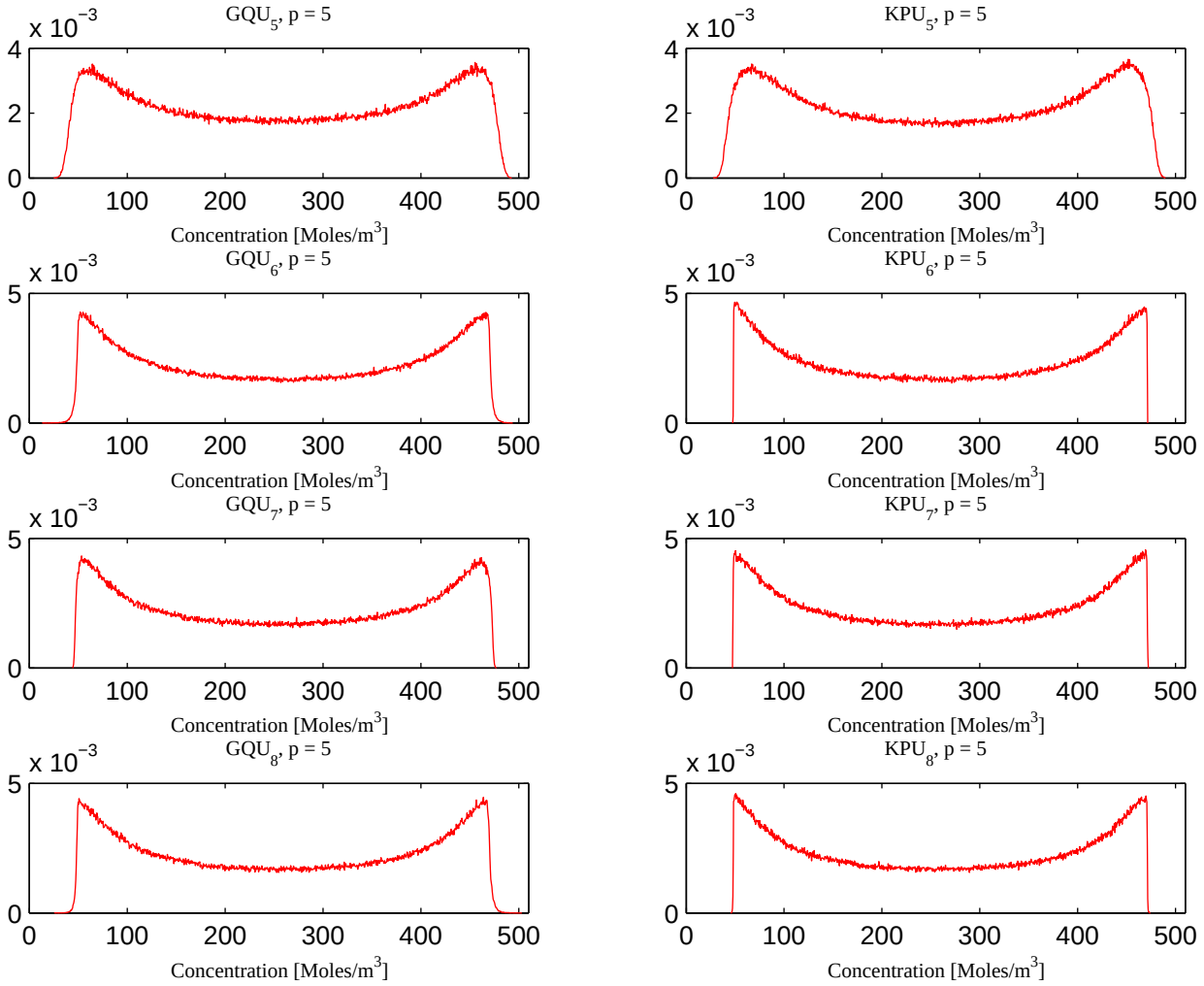


Figure B.1: pdf's of the first radionuclide in compartment 4 at 100 years.

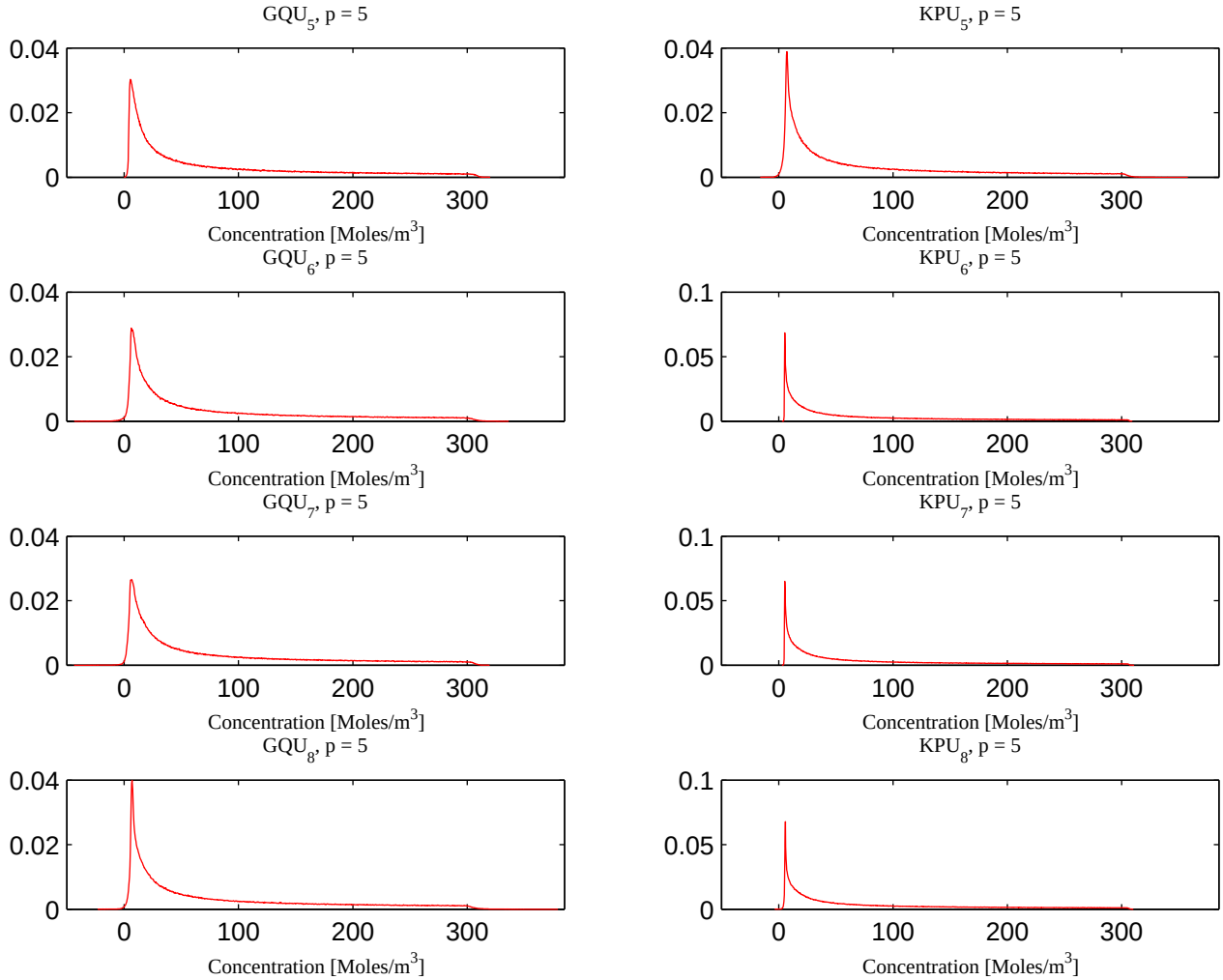


Figure B.2: pdf's of the second radionuclide in compartment 4 at 100 years.

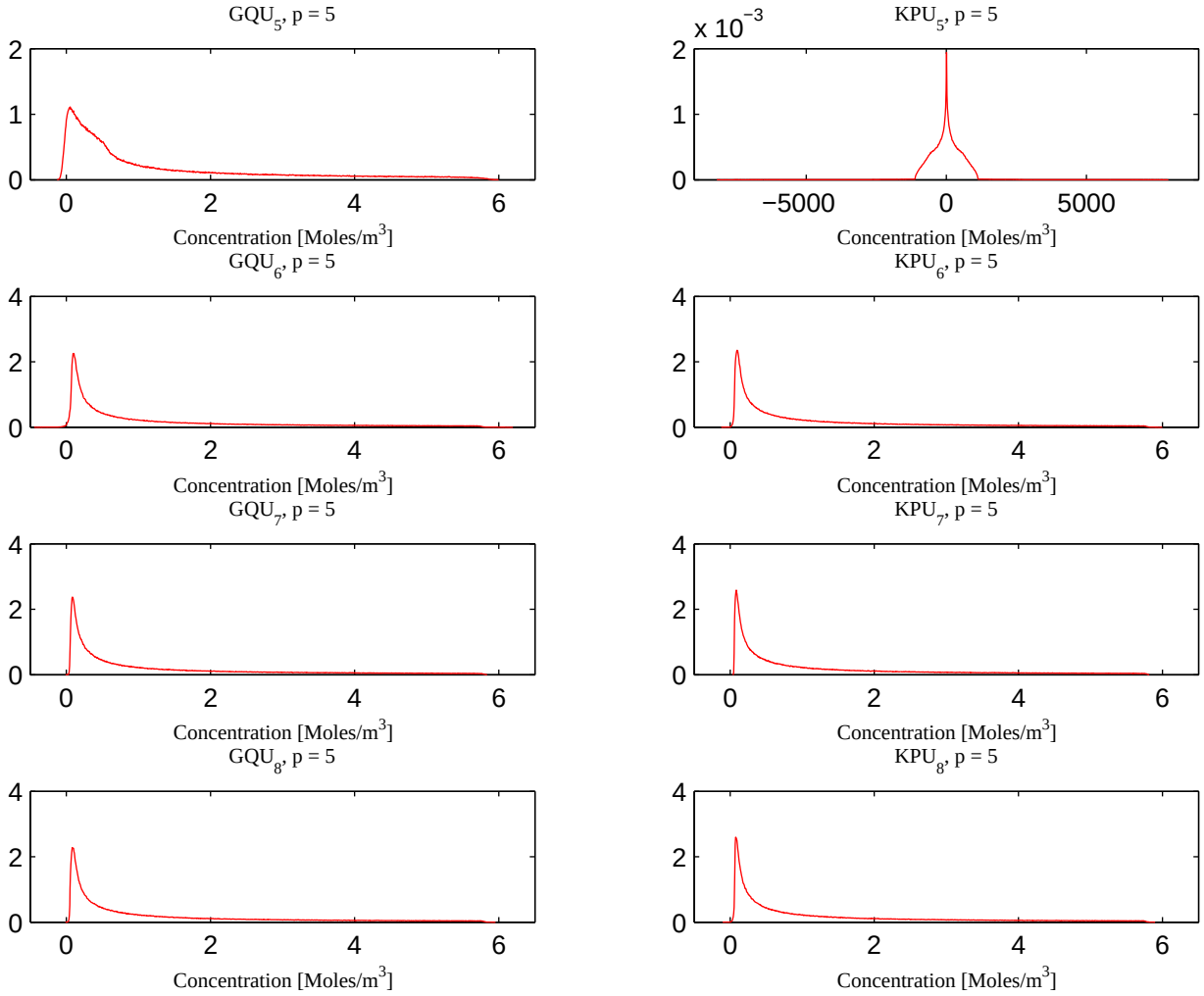


Figure B.3: pdf's of the third radionuclide in compartment 4 at 100 years.

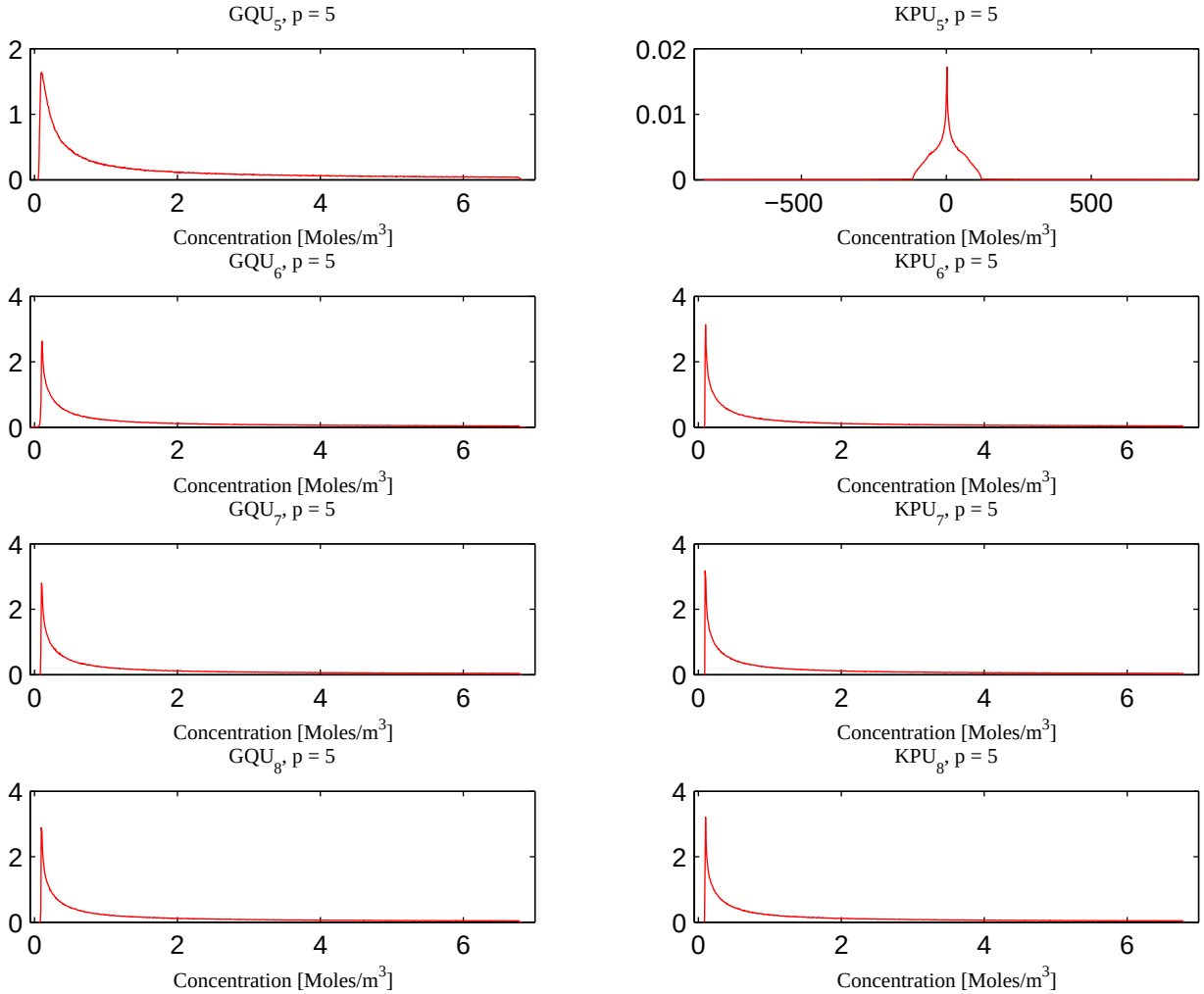


Figure B.4: pdf's of the fourth radionuclide in compartment 4 at 100 years.

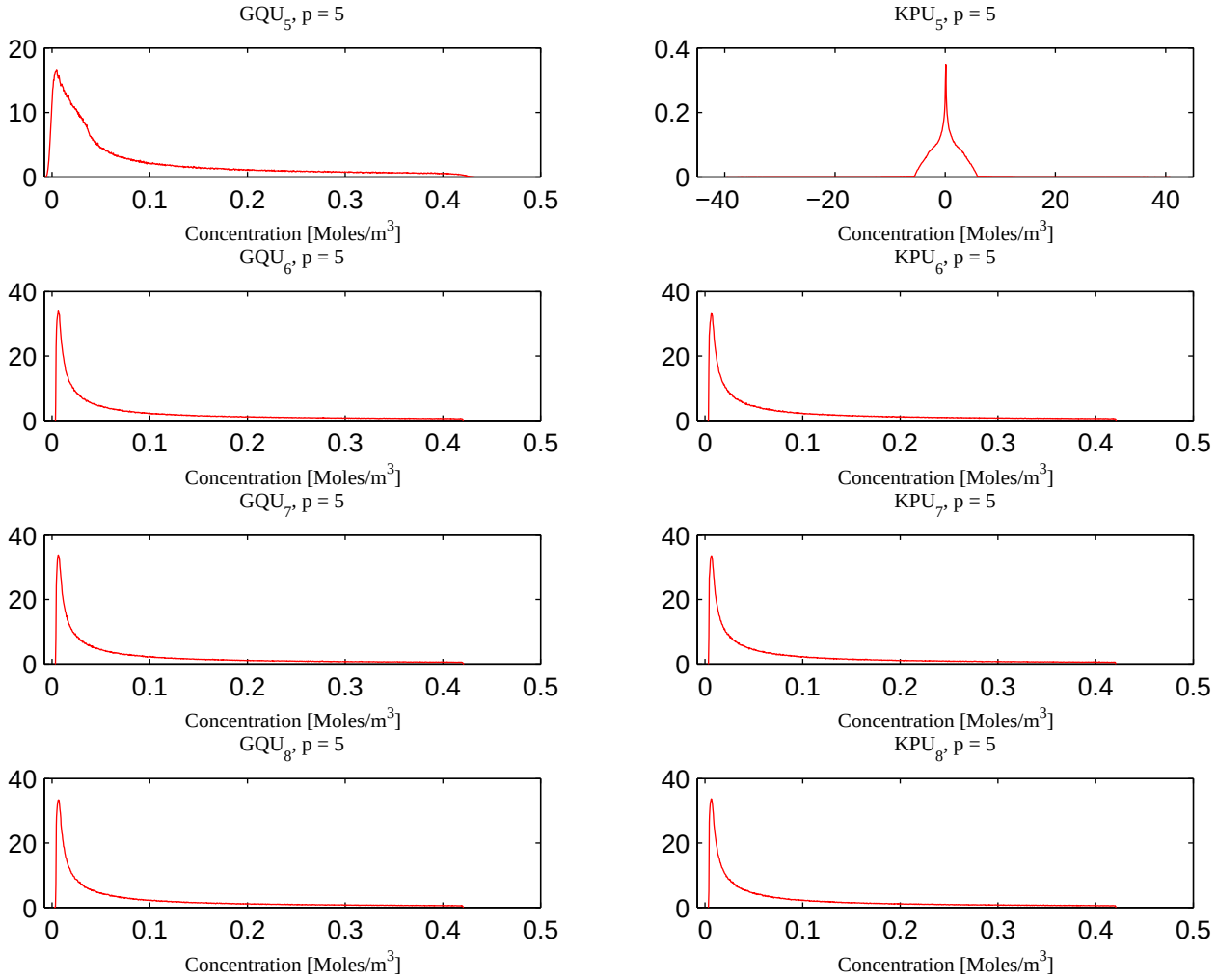


Figure B.5: pdf's of the fifth radionuclide in compartment 4 at 100 years.

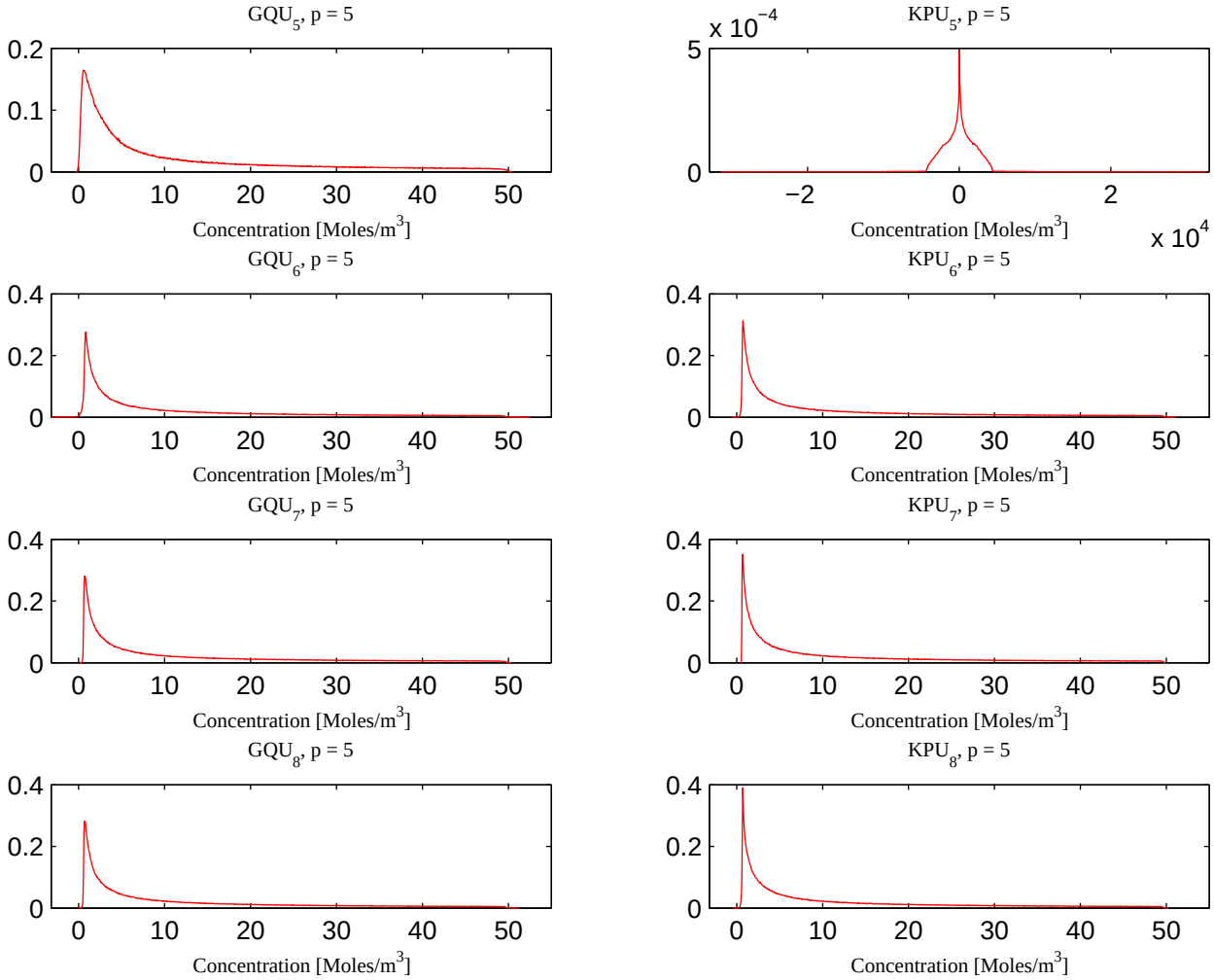


Figure B.6: pdf's of the sixth radionuclide in compartment 4 at 100 years.

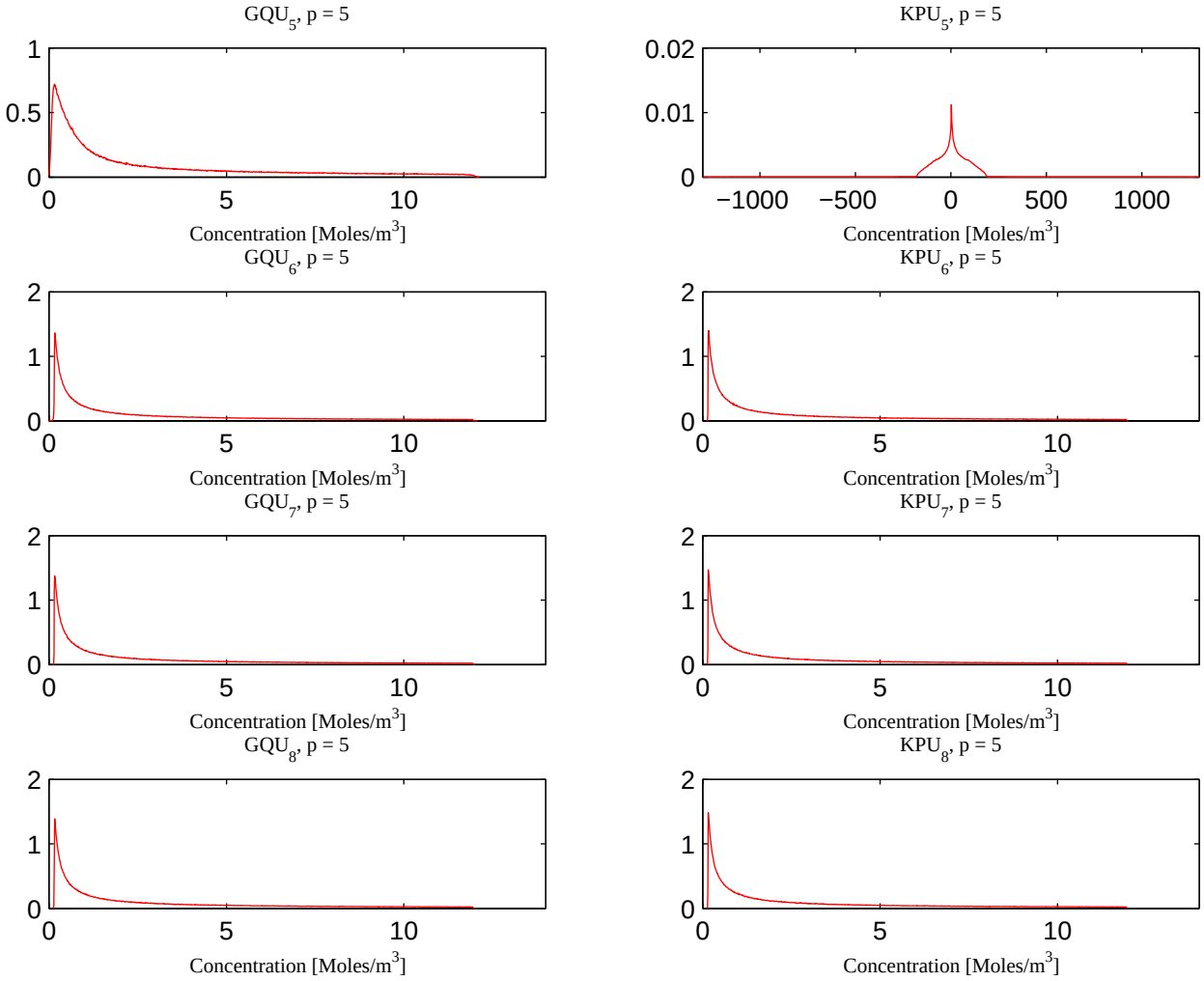


Figure B.7: pdf's of the seventh radionuclide in compartment 4 at 100 years.

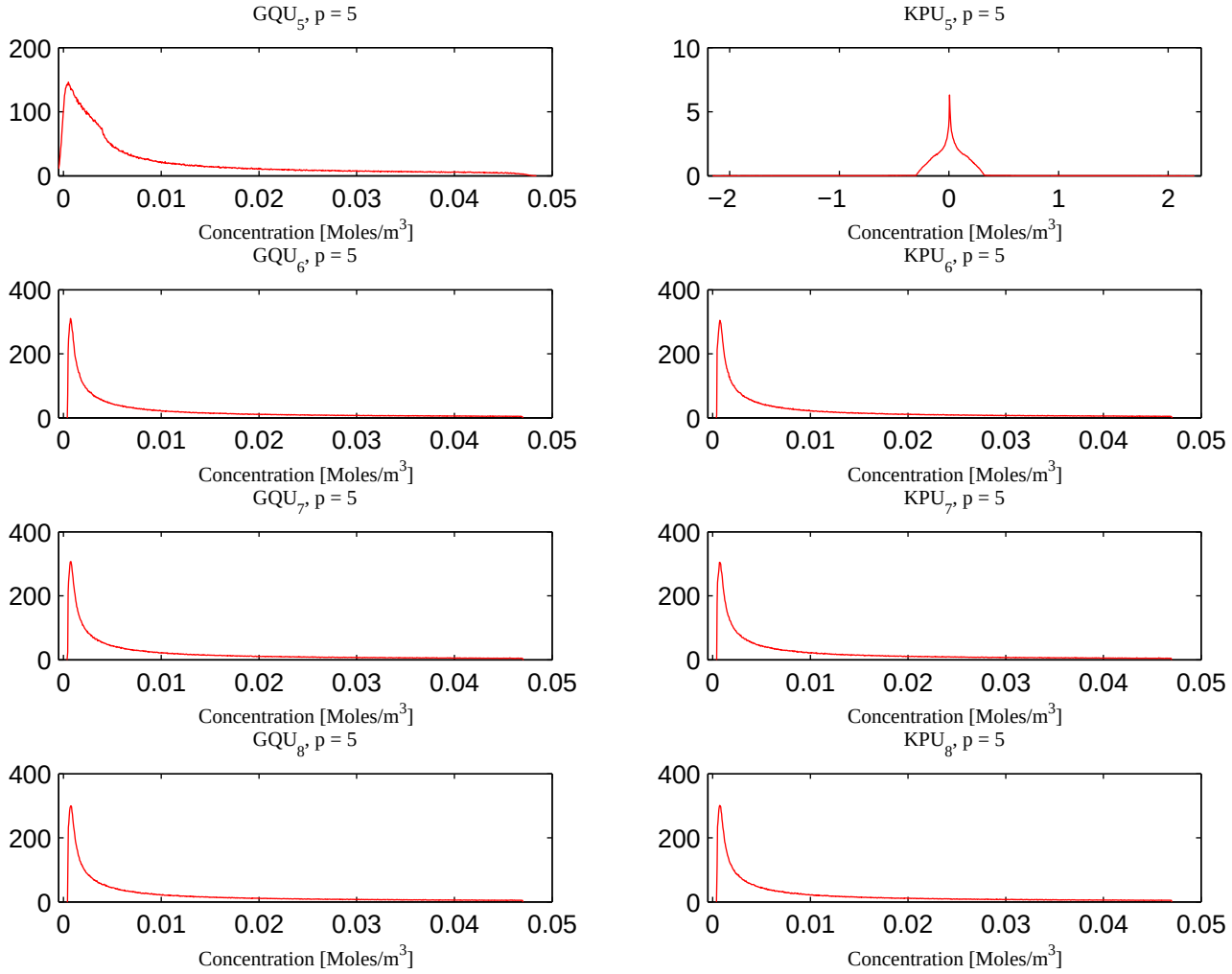


Figure B.8: pdf's of the eighth radionuclide in compartment 4 at 100 years.

C Pamina Results

C.1 cdf's of the maximum dose rate

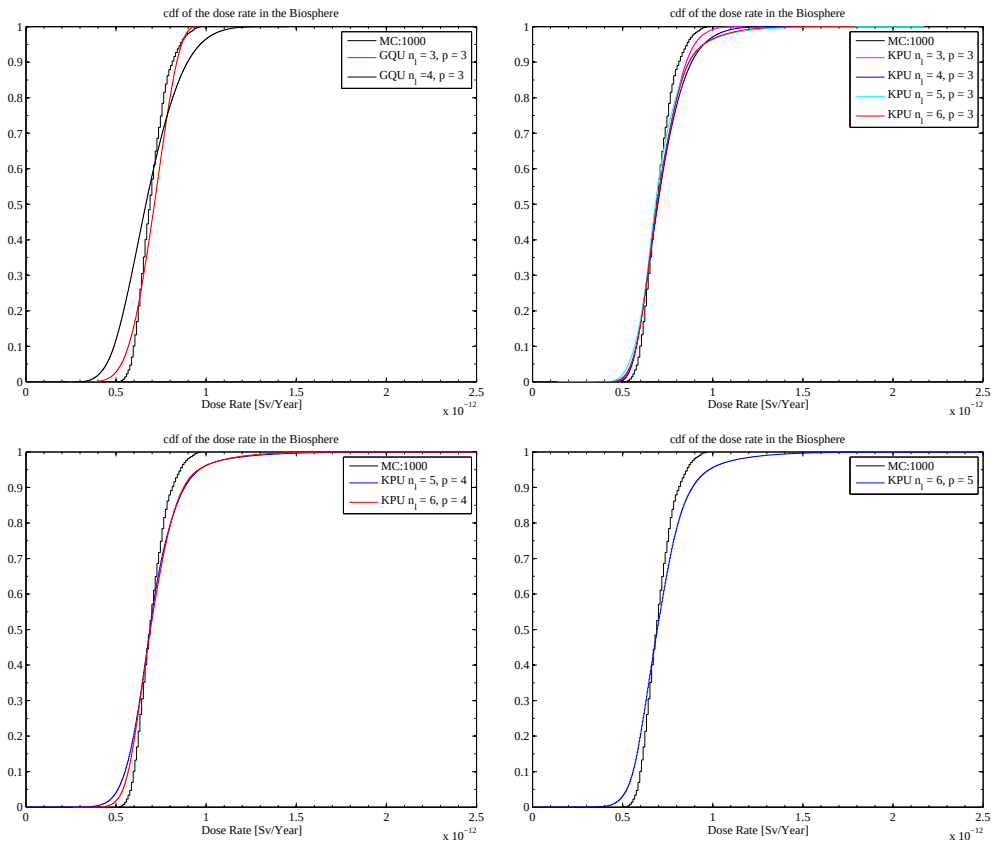


Figure C.1: PDF's calculated for different truncation order and different grids.