Heft 231 Sergey Oladyshkin

Efficient Modeling of Environmental Systems in the Face of Complexity and Uncertainty

Efficient Modeling of Environmental Systems
in the Face of Complexity and Uncertainty

Habilitationsschrift, Fakultät Bau- und Umweltingenieurwissenschaften
und Stuttgart Research Centre for Simulation Technology
der Universität Stuttgart

vorgelegt von
Sergey Oladyshkin
aus Nizhniy Novgorod in Russland

Hauptberichter: Prof. Dr.-Ing. Rainer Helmig

Tag der mündlichen Prüfung: 28.02.2014

Institut für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart
2014

Heft 231   Efficient Modeling of Environmental
Systems in the Face of Complexity
and Uncertainty

von
Dr.-Ing. habil. Sergey Oladyshkin

D93 **Efficient Modeling of Environmental Systems in the Face of Complexity and Uncertainty**

Acknowledgements

Contents

List of Figures

List of Tables

Abbriviations

Abbriviation	Denotation
aPC	arbitrary polynomial chaos
BF	Bootstrap filtering
CCS	carbon capture and storage
CDF	cumulative density function
CO_2	carbon dioxide
CPU	central processing unit
EnKF	ensemble Kalman filter
gPC	generalized polynomial chaos
GSA	global sensitivity analysis
HT	hydro-thermodynamic
IPCM	integrative probabilistic collocation method
ME-gPC	multi-elements generalized polynomial chaos
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
ME	maximum entropy
OTS	open thermodynamic simulator
PCE	polynomial chaos expansion
PDF	probability density function
RSF	random space function
SAIGUP	sensitivity analysis of the impact of geological uncertainty on production

Notation

The following table shows the significant symbols used in this work. Local notations are explained in the text.

Symbol	Definition	Dimension

Greek Letters:

α	multivariate index	
χ	basis concentration	[kg/kg]
δ	binary interaction factor	[-]
ε	system perturbation parameter	[-]
ε	error	
μ	dynamic phase viscosity	[Pa s]
μ	mean value	
μ	raw moment	
ν	chemical potential	[J/mol]
π	fractional mass flow ration	[-]
ρ	density	[kg/m^3]
ϕ	porosity	[-]
ϕ	fugacity factor	
ι	mobility	[1/(Pa s)]
σ	standard deviation	
τ	time of flight	[s]
ω	relative mobility parameter	[-]
ω	model input	
ξ	random input	

ζ	component neutrality function	[-]
$\mathscr{S}$	arbitrary streamline	[-]
Φ	arbitrary function	
Φ	multivariate polynomial basis	
Γ	probability measure	
Λ	space of events	
Υ	functional	
Π	polynomial	
Θ	transformation	
Ω	space of events	
Ω	model output	
ψ	hydraulic phase conductivity	[kg/(Pa m s)]

Latin Letters:

c	mass concentration	[kg/kg]
c	expansion coefficient	
d	order of expansion	
$\mathbf{e}$	unit vector	[-]
f	fractional flow function	[-]
f	frugality	
f	anisotropy ratio	
$\mathbf{g}$	metric tensor	
$\mathbf{g}$	gravity vector	
h	hydraulic head	[m]
l	curvilinear coordinate	[m]
k	relative permeability	[-]
n	mole number	[mol]
n_e	porosity	[-]
m	mass	[kg]
p	pressure	[-]
p	polynomial coefficient	
q	total flow rate	[kg/s]
r	health risk	

s	saturation	[-]
t	time	[s]
t^*	relaxation time	[s]
u	velocity	[m/s^2]
v	velocity	[m/s^2]
w	importance weigth	
x	mole fraction	[-]
$\mathbf{x}$	physical space	
$\mathbf{y}$	prior probabilityy	
z	z-factor	[-]
A	area of cross section	[m]
A	σ-algebra	
C	concentration	[kg/m^3]
C	covariance model	
Ca	capillary number	[-]
D	diffusion coefficient	
D	mass discharge	[kg/s]
F	pseudo-fractional flow function	[-]
G	chemical potential of the pure component	
$\mathbf{H}$	Henkel matrix	
H	Hermite polynomial	
I	integral scale	
J	Bessel function	
K	absolute permeability	[m^2]
L	length	[m]
M	molar mass	[kg/mol]
M	number of expansion therms	
N	number of chemical components	
N	number of input parameters	
P	pressure	[Pa]
P	polynomial basis	
P	number of expansion therms	
R	universal gas constant	[J/(mol K)]
R	risk	
$\mathbf{R}$	covariance matrix of measurement error	

S	sensitivity index	
S	salinity	[-]
T	temperature	[K]
U	Buckley-Leverett velocity	[m/s^2]
V	volume	[m^3]
V	velocity	[m/s^2]
X	particle displacement covariance function	
Y	orthogonal curvilinear coordinate system volume	
Y	stochastic model output	
Z	interface rise	[m]

Subscripts:

abs	absolute
aq	aquifer
c	coefficient
$crit$	critical
h	hydrogen
g	gas phase
i	order
l	liquid phase
in	input
out	output
p	pseudo
rel	relative
w	water
N	normal
Ph	physical
Ω	output

Superscripts:

d	degree of expansion

h	hydrogen
k	k^{th} chemical component
k	degree of polynomial
l	degree of polynomial
w	water
H	Henry
K	Kelvin
Ph	physical
R	Raoult

Operators:

∂	partial differential operator
d	total differential operator
div	divergence
$grad$	gradient
Δ	difference
δ	Dirac delta function
$\int$	integration
$\|\ \|$	norm
^{T}	transposition
$^{-1}$	inversion

Zusammenfassung

Die starke industrielle Entwicklung des letzten Jahrhunderts führte zu einer erheblichen Steigerung der öffentlichen Nachfrage nach verschiedenen Energiearten, was wiederum einen enormen Anstieg der Nachfrage nach natürlichen Ressourcen zur Folge hatte. Natürliche Ressourcen bilden einen bedeutenden Bestandteil unserer Umwelt. Die Entwicklung einer Vielzahl von Technologien zur Gewinnung von natürlichen Ressourcen verursacht eine fortwährende Zunahme von Umwelteingriffen. Gleichzeitig sind Umweltsysteme eine der bedeutendsten Erscheinungsformen von komplexen dynamischen Systemen. Aus diesem Grund brauchen wir ein besseres Verständnis unserer natürlichen Umgebung, das uns eine effiziente und sichere Interaktion mit der Umwelt und ein nachhaltiges Ressourcenmanagement zum Wohle der Menschheit ermöglichen kann. Insbesondere die Fähigkeit vorherzusagen, wie sich unsere Umwelt im Laufe der Zeit verändern oder auf geplante Eingriffe reagieren wird, ist unverzichtbar. Unsere Umwelt verhält sich jedoch in verschiedenen zeitlichen und räumlichen Dimensionen sehr komplex. Darüber hinaus sind viele Umweltsysteme nichtlinear und dominiert von Echtzeiteinflüssen externer treibender Kräfte. Ein vollständiges Bild der Umweltsysteme ist oft nicht verfügbar, denn viele dieser Systeme lassen sich nicht direkt beobachten und eine Beschreibung kann nur anhand von spärlich vorhandenen Messwerten hergeleitet werden. Hinzu kommt, dass Umweltdaten kaum erhältlich und teuer zu beschaffen sind. Mehrere Jahrzehnte der Forschung haben gezeigt, dass Modellierung eine sehr wichtige Rolle bei der Rekonstruktion eines vollständigen Bildes von Umweltsystemen spielt und eine einzigartige Möglichkeit bietet, das Verhalten solcher vielschichtiger Prozesse vorherzusagen. Aufgrund der Komplexität unserer Umwelt, besteht die größte Herausforderung in der Konstruktion zuverlässiger und praktikabler Modelle, die eine adäquate Beschreibung physikalischer Konzepte unter gleichzeitiger Berücksichtigung von Unsicherheiten liefern. Die vorliegende Arbeit befasst sich mit Forschung auf dem Gebiet der Umweltmodellierung unter Beachtung von Komplexität und Unsicherheit. Die Arbeit besteht aus drei Teilen und behandelt Anwendungsfälle wie Erdöllagerstätten, Grundwasserströme, die Lagerung radioaktiver Abfälle sowie die unterirdische Speicherung von energierelevanten Gasen. Jeder Teil enthält mehrere Kapitel, die jeweils auf ausgewählten

Publikationen basieren und unabhängig voneinander gelesen werden können. Teil II konzentriert sich auf physikalische Konzepte und stellt mehrere Möglichkeiten vor um die Modellierung zu beschleunigen. Teil III beschäftigt sich mit effizienten Methoden der Modellreduktion zur Quantifizierung von Unsicherheiten. Teil IV zeigt Anwendung als die Speicherung von energierelevanten Gasen in geologischen Formationen und diskutiert die damit verbundenen Herausforderungen. Der aktuelle Abschnitt bietet einen allgemeinen Überblick in deutscher Sprache.

Von Komplexen Systemen zu machbaren Modellen   Viele Umweltprobleme werden von Echtzeiteinflüssen externer treibender Kräfte beherrscht. Leider ist ein vollständiges Bild von Umweltsystemen nicht verfügbar - insbesondere wenn man versucht, das physikalische Verhalten von unterirdischen Systemen wie Grundwasserströmen, Erdöllagerstätten, Lagerstätten von radioaktiven Abfällen oder von Gasspeichern zu untersuchen. Erschwerend hinzu kommt, dass Umweltdaten kaum erhältlich und teuer zu beschaffen sind. Für die Rekonstruktion eines vollständigen Bildes solcher Umweltsysteme benötigen wir Zusammenhang adäquater und effizienter physikalischer Modelle. In diesem Kontext spielen Modelle eine sehr wichtige Rolle für das Verständnis von Umweltprozessen und bieten eine einzigartige Möglichkeit, das Verhalten komplexer Systeme vorherzusagen.

Die beteiligten Mehrphasenströmungs- und Transportprozesse in unterirdischen porösen Medien werden von komplexen und nicht-linearen Multi-Physik- und Multi-Skalen-Gesetzen beherrscht, die unter anderem instabile Mehrphasenströmungen, Phasenübergänge, Gesteinsdeformationen, geochemische Prozesse und Temperatureffekte beinhalten. Darüber hinaus ist zu berücksichtigen, dass Umweltsysteme Auswirkungen in sehr großen räumlichen und zeitlichen Dimensionen haben können. Aufgrund der Komplexität von Umweltsystemen sind numerische Simulationsmodelle von Mehrphasenströmungs- und Transportprozessen oft zu rechen-aufwändig für realistische großflächige Anwendungen. Selbst die Einbeziehung von parallelem Hochleistungsrechnen kann das Problem nicht vollständig lösen. Die große Schwierigkeit besteht folglich in der Entwicklung von konsistenten und machbaren Umweltmodellen, die sowohl angemessene konzeptuelle Beschreibungen liefern, als auch einen praktikablen Zeitrahmen für die Simulationen wahren.

Teil II konzentriert sich auf die konzeptuellen Besonderheiten von Umweltsystemmodellierungen und eröffnet vielfältige Möglichkeiten um den Modellierungsprozess zu beschleunigen. Kapitel 5 beinhaltet eine kurze Diskussion möglicher Fehlerquellen innerhalb des gesamten Modellierungsprozesses und zeigt ihre möglichen Auswirkungen am Beispiel der

unterirdischen Speicherung von Kohlendioxid. Kapitel 5 unterstreicht die möglicherweise größte, übergeordnete Herausforderung, welche darin besteht, einen gesunden und vernünftigen Kompromiss zwischen numerischen Verfahren, konzeptionellen physikalischen Formulierungen, der Schätzung von Modellparametern und ihren Unsicherheiten sowie dem Rechenaufwand zu finden. Kapitel 6 stellt eine effiziente Methode zur Handhabung thermodynamischer Eigenschaften einer kompressiblen kompositorischen zweiphasigen Strömung in porösen Medien vor, die eine extrem kurze Rechnungzeit hat. Kapitel 7 präsentiert einen effektiven Ansatz basierend auf der Stromlinientechnik für mehrphasige Mehrkomponentenströmungen in großflächigen heterogenen Erdöllagerstätten.Kapitel 8 befasst sich mit Wasserstoffmigration rund um Lagerstätten von radioaktiven Abfällen. Dieses Kapitel bietet eine kalibrierte analytische Beschreibung der Löslichkeit von Wasserstoff in Wasser (Phasendichten, Viskositäten und Phasenkonzentrationen), die die weitere Modellierung wesentlich beschleunigt.

Effiziente Methoden der Modellreduktion zur Quantifizierung von Unsicherheiten Umweltsysteme wie unterirdische Mehrphasenströmungen, Grundwasserströmungen oder Kohlendioxidspeicher sind Erscheinungsformen komplexer dynamischer Systeme. Mangelnde Informationen über ihre Systemeigenschaften führen zu Modellunsicherheiten bis zu einem Ausmaß, bei dem die Quantifizierung von Unsicherheiten zur dominierenden Frage der Modellierungs-, Simulations- und Anwendungsaufgaben werden kann. Die Vorhersage des Verhaltens von weiträumigen Umweltsystemen hängt stark von unserer Fähigkeit ab, ihre Unsicherheiten zu quantifizieren. Aktuelle numerische Simulationsmodelle sind oft zu rechenintensiv für fortgeschrittene Anwendungsaufgaben, welche die genaue Quantifizierung von Unsicherheiten, Risikobewertung, robustes Design sowie Modellkalibrierung umfassen. Selbst einzelne deterministische Simulationen erfordern parallele Hochleistungsrechnungen. Weil die beteiligten komplexen Strömungsprozesse in porösen Medien einen äußerst nichtlinearen Charakter haben, ist das Problem zu nichtlinear, um quasilineare und andere vereinfachte stochastische Werkzeuge anzuwenden.

Teil III befasst sich mit der Entwicklung von Lösungsansätzen für diese erweiterten Herausforderungen basierend auf stochastischen Verfahren der Modellreduzierung. Die Reduktion wird durch Projektionen auf orthonormale Polynombasen erreicht, die eine sogenannte "Antwortfläche" bilden. Auf diese Weise wird die Reaktion des Modells auf Änderungen unsicherer Parameter durch multivariate Polynome repräsentiert und erlaubt so eine nichtlineare Fortpflanzung der Modellunsicherheiten auf das vorherzusagende Risiko. Dieses Verfahren ist bekannt als polynomielle Chaos-Expansion auf dem Gebiet der stochastis-

chen partiellen Differentialgleichungen. Das von der Antwortfläche repräsentierte reduzierte Modell ist erheblich schneller als das ursprüngliche, komplexe Modell und bietet so einen vielversprechenden Ausgangspunkt für nachfolgende Aufgaben, zu denen die globale Sensitivitätsanalyse, die Quantifizierung von Unsicherheiten, robustes Design, probabilistische Risikobewertung sowie Modellkalibrierung zählen. In vielen Fällen wird die Tatsache, dass die Antwortoberfläche bekannte polynomische Eigenschaften hat, helfen, diese Aufgaben weiter zu vereinfachen. Auf diese Weise bietet stochastische Modellreduzierung eine Möglichkeit, Unsicherheiten in nichtlinearen Systemen mit einem vertretbarem Aufwand zu quantifizieren.

Kapitel 9 konzentriert sich auf das arbitrary polynomielle Chaos-Expansion, das sich an beliebige Formen der Wahrscheinlichkeitsverteilung von Eingabeparametern anpassen kann und eine verbesserte Konvergenz im Vergleich zu klassischen Chaosexpansionsverfahren bietet. Kapitel 10 bietet einen flexiblen und effizienten Ansatz für die globale Sensitivitätsanalyse, um die Auswirkungen von Parametervariationen oder Parameterunsicherheiten auf die Gesamtunsicherheit des Modells zu quantifizieren. Dieser Ansatz kann Informationen einer globalen Sensitivitätsanalyse mit einem Rechenaufwand ermitteln, der kaum größer ist als der Aufwand für eine lokale Analyse. Kapitel 11 präsentiert einen neuen Ansatz zur Modellkalibrierung und zum "History Matching", basierend auf arbitrary polynomielle Chaos-Expansion und strengen Bayesschen Grundsätzen. Durch diese Kombination erhalten wir eine statistische Methode des "History Matching", die ein hohe Genauigkeit besitz, und dennoch eine Rechengeschwindigkeit aufweist, die mehr als ausreichend für die Echtzeit-Anwendung ist.

Anwendung auf die Speicherung von energierelevanten Gasen in geologischen Formationen Die Modellierung von Umweltsystemen erfordert oft die Erfüllung von Aufgaben wie des Vorhersage möglicher Szenarien, der Ausgestaltung geplanter Engineering-Maßnahmen, die Abschätzung von Risiken, des Kalibrierung von Modellen anhand verfügbarer Daten etc. Der gemeinsame Kern all dieser Aufgaben ist die Häufige Auswertung des zugrunde liegenden Modells mit variierenden Modellparametern. Unabhängig davon sollte die gesamte Modellierungs innerhalb eines akzeptablen Zeitrahmens durchgeführt werden. Darüber hinaus ist der Modellierungsprozess im Bereich der Umwelttechnik eine Kette von vielen Aufgaben. Wie jede Kette, ist sie nur so stark wie ihr schwächstes Glied. Die Abschätzung von Unsicherheiten ist ein Glied der Kette und die Unterschätzung ihres Einflusses kann dieses Glied immens schwächen. Kapitel 5 zeigt, dass Modellierungsfehler verschiedene Ursachen haben können und Auswirkungen auf den gesamten

Modellierungsprozess möglich sind. Somit bietet eine adäquate Kombination aus einem zuverlässigen physikalischen Modell und aktuellen stochastischen Werkzeugen eine leistungsfähige Basis für die Vorhersage des Verhaltens von Umweltsystemen. Teil IV enthält Anwendungsbeispiele von Modellreduktionsverfahren auf Probleme der Speicherung von energierelevanten Gasen in geologischen Formationen.

Viele Umweltprojekte müssen so gestaltet werden, dass selbst unwahrscheinliche Gefahren unterhalb eines akzeptablen Risikoniveaus bleiben. Kapitel 12 diskutiert einen integrativen Ansatz für ein robustes Design bei bestehenden Unsicherheiten und eine probabilistische Risikobewertung für CO_2 Speicherung in geologischen Formationen. Es zeigt, dass sich Versagenswahrscheinlichkeiten aus einem Wechselspiel zwischen steuerbaren technischen Aspekten und unsicheren Aspekten des Systems ergeben. Kapitel 13 befasst sich mit der Bewertung von steigendem Druck in Speicherstätten, der unter Umständen zu Deckgesteinsbrüchen oder Soleverschiebungen führen kann. Hierzu präsentiert Kapitel 13 eine Studie über den Einfluss der Dimension hochpermeabler Kanäle und ihrer Permeabilität auf die Druckentwicklung in einem Kanalsystem während der CO_2.

Es ist bekannt, dass die Quantifizierung von Unsicherheiten in unterirdischen Systemen Annahmen über die Wahrscheinlichkeitsverteilungen aller Modellparameter erfordert, was entweder eine große Menge an verfügbaren Daten voraussetzt oder eine sehr subjektive Wahl des Verteilungsformen erfordert. Kapitel 14 stellt einen minimal subjektiven Ansatz zur Quantifizierung von Unsicherheiten vor, der auf einer neuen und rein datengesteuerten Formulierung der polynomielle Chaos-Expansion basiert. Formulieng die Analyse von Unsicherheiten die Berücksichtigung verschiedener Arten von Unsicherheiten je nach Anwendung des Modells. Kapitel 15 unterscheidet zwischen Szenario- und statistischen Unsicherheiten. Insbesondere befasst sich Kapitel 15 mit der Schätzung des Risikos von Soleeinleitung in Süßwasseraquifere durch CO_2-Injektion in geologische Formationen. Kapitel 16 nutzt die globale Sensitivitätsanalyse und die Risikobewertung auf Grundlage des Chaosexpansionsansatzes zur Untersuchung eines realistischen, großflächigen Problems der CO_2-Speicherung. In dieser Studie wurde der Einfluss von unsicheren Parametern untersucht, die die strukturellen Heterogenitäten steuern. Hierzu zählen Barrieren, Aggradationswinkel, Durchleitungsvermögen von Gesteinsspalten sowie die externe Einflüsse durch Aquifere.

Part I.

Introduction

This thesis contains research in the field of environmental modeling in the face of complexity and uncertainty. Research over several decades has shown that modeling plays a very important role in reconstructing (as far as possible) the complete complex picture of environment systems and offers a unique way to predict behaviors of the multifaceted processes at play in such complex systems. Due to the complexity of the surrounding environment, the greatest challenge here is the construction of reliable and feasible models that can adequately describe physical concepts and, at the same time, account for uncertainty. The presented thesis is divided into three Parts and refers to diverse applications such as underground petroleum reservoirs, groundwater flow, radioactive waste deposits and storage of energy relevant gases. Each Part contains Chapters that are based on selected publications and can be read independently. Part I is the overall introduction, Part II focuses on complex physical concepts and their mathematical description. Also, it offers several possibilities to accelerate the modeling process. Part III deals with efficient model reduction methodologies for uncertainty quantification. Part IV demonstrates application to the storage of energy relevant gases in geological formations and discusses related challenges.

1. The challenge of modeling environmental systems

Strong industrial development of the last century has led to a significant increase in public demand for different types of energy and, as a consequence, to an enormous increase in demand for natural resources. Naturally, all types of nature resources form a part of our surrounding environment. In order to extract natural resources a wide variety of technologies has been developed. This has led to a strong rise in interventions in the environment continuing up to the present days. At the same time, environmental systems form one of the largest and most important classes of complex dynamic systems. For this reason, society needs a better understanding of the environment in order to have an efficient and safe interaction for the sake of maximized welfare and sustainability in resources management. In particular, the ability to predict how the environment changes over time or how it will react to planned interventions is indispensable. However our surroundings behave non-trivially in various time and spatial scales. Moreover, many environmental systems are heterogeneous, non-linear and dominated by real-time influences of external driving forces. Unfortunately, a complete picture of environmental systems is not available, because many of these systems cannot be observed directly and only can be derived using sparse measurements. Moreover, environmental data is hardly available and expensive to acquire. Overall, this leads to limited observability, and an inherent uncertainty in all modeling endeavors. Still, research over several decades has showed that modeling plays a very important role in reconstructing (as far as possible) the complete and complex picture of the environment systems and offers a unique way to predict behavior of such multifaceted systems.

The current thesis mainly deals with environmental systems that are mostly situated below the surface (see Figure 1.1), such as underground petroleum reservoirs, groundwater flow, flow around radioactive waste subsurface storage sites and deep geological formations for storage of energy relevant gases. The mentioned systems have already been exploited for a long time in diverse locations. However, they still stay very challenging and still attract a lot
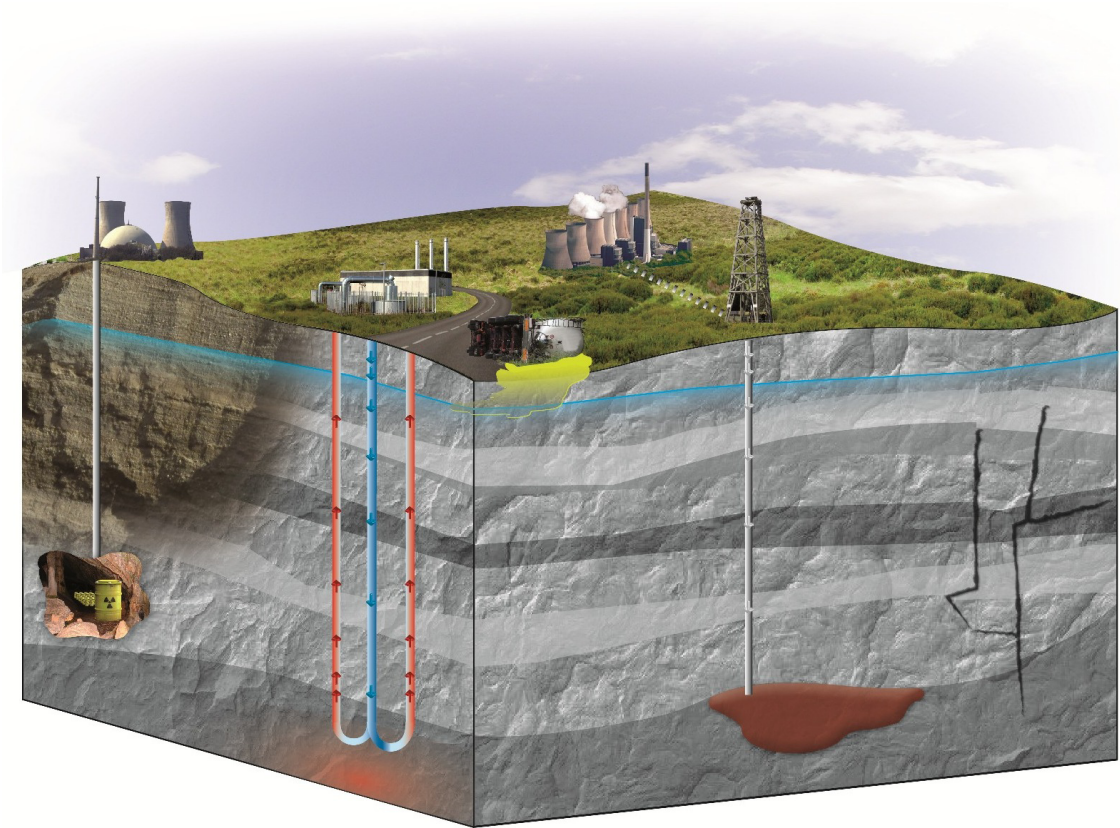
Figure 1.1.: The subsurface and its various competing use types as an example for a complex environment system

of attention. Let us briefly describe the discussed systems.

During the last century our society hugely exploited the hydrocarbons that occur naturally in subsurface. The hydrocarbons still play an extremely important role in satisfying the energetic demands of our society. It can be gas or oil that typically form extremely big fields in the underground. Multiphase multi-component flows in large-scale heterogeneous petroleum reservoirs are represented by very complex models. One of the most challenging problems in this field is to derive models for flow in complex reservoirs accounting for their geology and petrophysical properties. The modelling should take into account the simulation of the reservoir parameters that matches historical dynamic production data in order to predict pressure and saturation changes in the reservoir during exploitation.

Over the last decades storage of carbon dioxide (CO_2) in geological formations has been intensively discussed as an interim technology with a high potential for mitigating CO_2 emissions. This technology comprises capturing CO_2 at industrial facilities, compressing it into

a fluid or supercritical state and disposing of it in deep underground formations. In recent years research efforts have been directed towards understanding the physical processes in CO_2 storage (e.g. (IPCC, 2005)). Like in the case of petroleum reservoirs, the multiphase flow and transport processes involved are strongly non-linear. They include phase changes in the region of the critical point and effects such as gravity-induced fingering and convective mixing as well as geo-chemical and geo-mechanical processes etc. The principal hazards that can occur in CO_2 storage need to be addressed. Such hazards can be, for example, leakage of CO_2 from the reservoir back into the atmosphere, pollution of drinking water resources or structural failures caused by very high pressure peaks.

Another concern of our society is the processing and storage of radioactive waste resulting from nuclear energy. The danger here is that radioactive elements can leak from the subsurface storage site. Such elements can be transported by other molecules that migrate from the storage of radioactive waste. For example, substantial quantities of hydrogen (H_2) are produced by physical process within the nuclear waste. Other gases, such as CH_4, C_2H_6, C_2H_2, etc., can be produced as well, but in quantities that are very small in comparison to H_2. The hydrogen is formed as a result of physical and chemical processes in storage, which are caused by an intensive corrosion of metal. The storage frame has several hundred thousand tons of metal, and the surface of corrosion reaches several hundred thousand m^2. Hydrogen production by corrosion essentially in presence of water, but it still can be formed as a result of other processes, such as radiolysis waters and decomposition of organic substances. A particular property of hydrogen is its small molecular size. This why H_2 has a strong ability of transport in porous medium even with almost impermeable properties. The big danger consists in transfer of radionuclides from the waste storage site through molecules of hydrogen. That is why the analysis of gas transfer (mainly hydrogen transfer) in porous media saturated by water is a non-trivial problem and demands better understanding of transport processes.

Another irreplaceable and maybe even the most essential natural resource for our society is drinking water consumed in everyday life. Huge amounts of freshwater and drinking water are located in the subsurface and represent an important part of the hydrological cycle. Such subsurface water systems can be easily polluted and need to be monitored efficiently. In particular, the groundwater can be contaminated by agricultural, urban and industrial processes or by other ecological sources. This can directly affect the quality of drinking water and also can have impact on plants and animals. Because of an extremely large scale of groundwater resources, their heterogeneity and their connectivity it is very challenging to predict the behavior of such nontrivial systems

2. Complexity of environmental systems and models

Flow and transport processes in subsurface porous media often involve several phases, such as water, oil, gas, and so forth. These process are governed by complex and non-linear multi-physics and multi-scale laws which could take into consideration multiphase flow, phase transitions, rock deformation, geochemistry, temperature effects etc. Although these laws are traditionally derived for small scales, many modeling tasks require considering impacts on very large spatial and temporal scales as well. Such multifaceted environmental systems demand adequate physical model concepts for the reconstruction of a full and complex picture of the environment. Due to the complexity of the surrounding environment, numerical simulation models for multiphase flow and transport processes are often too cumbersome for realistic large-scale applications. Even the increasing power of parallel high performance computing offers only limited help and can never completely solve that problem. The big difficulty here is the definition of consistent and feasible environmental models that can provide adequate conceptual descriptions and that can simultaneously maintain a reliable time frame of simulations. This Chapter discusses several possibilities to accelerate the modeling process for a selected class of environmental problems, keeping at the same time the underlying physical aspects. These possibility will be explained in more detail in the Chapter 5 to Chapter 7 in Part II of this thesis, and the corresponding cross-references will be provided here.

2.1. On possible sources of modeling errors

In spite of all possible conceptual and numerical finesse, the possibly greatest overall challenge consists in finding a healthy and reasonable compromise between robust and efficient numerical techniques, adequately complex and accurate conceptual physical formulations,

reliable estimation of model parameters and their uncertainties, overall computational effort, etc. This compromise is very relevant for diverse environmental systems that offer only limited observability, and where multifaceted and imperfectly determined processes are involved, owing to heterogeneity, complexity, dynamics, scarcity of data and uniqueness of place (Pappenberger and Beven, 2006). In Chapter 5, we use the example of carbon dioxide (CO_2) storage in geological formations to demonstrate the impact of several modeling error sources, such as uncertain choice of physical concepts, error in numerical schemes, parameter uncertainty and human subjectivity in data interpretation.

Unfortunately, as for many environmental problems, modeling underground CO_2 storage involves many conceptual and quantitative uncertainties (Hansson and Bryngelsson, 2009a). The lack of information on subsurface properties may lead, depending on the specific question at hand, to parameter uncertainties up to a level where the uncertainties dominate or even override the influence of secondary physical processes (see (Oladyshkin et al., 2010)). However, a significant part of the scientific community still refrains from considering uncertainty in modeling, although the corresponding arguments are discussed and rejected one by one in (Pappenberger and Beven, 2006). Obviously, handling of uncertainty is necessary, but unfortunately it is a very challenging task, which will be discussed in detail in the upcoming Part III. Model-based uncertainty analysis can help to better judge the potentials and hazards in many engineering applications. This requires to specify the probability distributions of
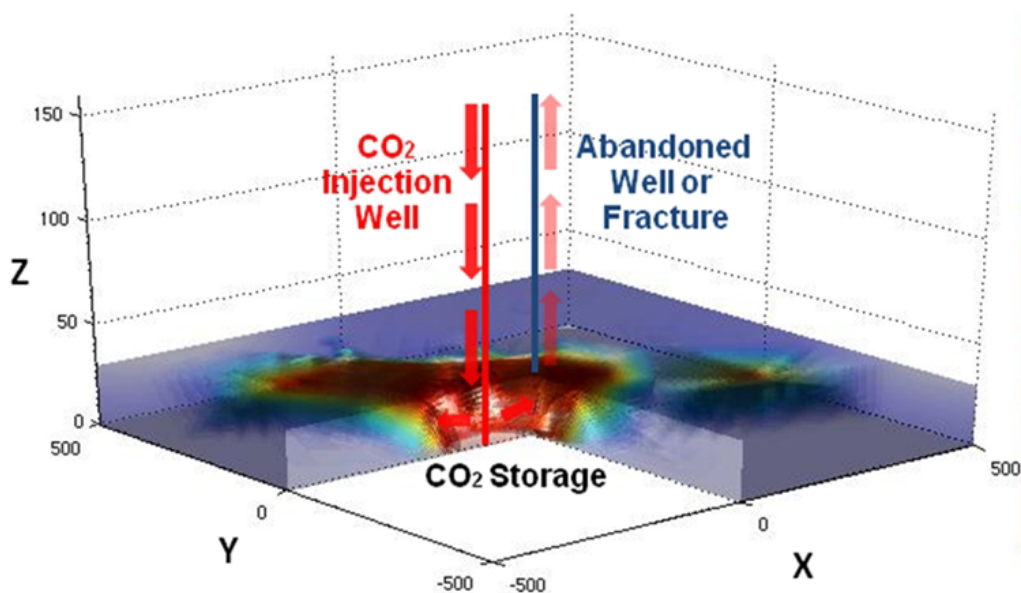
Figure 2.1.: Illustration of CO_2 leakage well problem

all model parameters, posing a huge demand on data availability or requiring highly subjective assumptions on distribution shapes to compensate for missing data. Even the outcome of well-established simulation models (see illustration in Figure 2.1) can lead to surprisingly different results betweeen different numerical schemes, different subjective conceptual model choices, or different subjective choices in raw data interpretation. Chapter 5 attracts attention to the often ignored or forgotten fact that efforts invested in improved physical conceptualization, numerical codes and stochastic modeling can easily be overwhelmed by error through human subjectivity in data interpretation and in histogram analysis at a very early stage of modeling.

2.2. Thermodynamic properties of multiphase compositional flow

A specific challenge in petroleum reservoir modeling and simulation is handling of thermodynamic properties of compositional flow. This section introduces research on how to efficiently include compositional effects and inter-phase mass transfer into multi-phase flow simulations, using the example of gas-oil reservoirs.

Mass exchange between phases strongly affects the flow process. Typically, the compositional fluid consists of a large number of chemical components. The components are able to form different thermodynamic phases separated from one another by an interface. There are three types of components in the system: light (N_2, CH_4, C_2H_6, ...), heavy (C_5H_{12}, C_6H_{14}, ...) and neutral (C_3H_8, C_4H_{10}, ...). The compositional fluid in porous media takes the form of a finely dispersed system with a characteristic scale of liquid drops or gas bubbles of the order of pore size. The interface between the phases is then highly dispersed and irregular. Under such conditions the examined system may be effectively described in terms of a phenomenological approach when both phases are considered as two interpenetrating continua, so that each of them occupies the whole space. Moreover, the system may be assumed to be in local equilibrium, due to a highly developed interface area ensuring fast mass transfers. To describe this system, a model based on high order nonlinear transcendent equations is used, which can only be solved numerically (Aziz and Settari, 1979), (Coats, 1980). Some simplified versions of the compositional model may be obtained by reducing the number of components up to two (Koldoba and Koldoba, 2003), or by assuming a flow stationarity (Whitson et al., 2003), (Dinariev, 1996). However, the simplified models have

a limited range of applicability and therefore accounting for the compositional effects at a higher complexity and accuracy level is very relevant for practical applications.

Generally, one can distinguish two typical physical scenarios of thermodynamic behavior in the compositional system. The first one corresponds to the closed thermodynamic system behavior. This type of behavior means that the phase transition between a gas sub-volume and a liquid sub-volume happens in a closed volume. Physically this type of behavior corresponds to the thermodynamic experience in a so-called closed PVT bomb, which describes the most fundamental thermodynamic data such as pressure, volume and temperature (PVT). The thermodynamic theory of these systems is well developed in fundamental works: (Sedov, 1976), (Nigmatulin, 1987), as well as in reservoir engineering: (Coats, 1980), (Nikolaevski et al., 1968), (Firoozabadi, 1999), (Danesh, 1998), (Batalin et al., 1992). The second type of behavior corresponds to the open thermodynamic system behavior. In this case, there is not only phase exchange between the two individual sub-volumes of gas and liquid, but also a motion of the phases relative to each other.

Usually, for this kind of analysis, a full compositional model consisting of thermodynamic relations as well as hydrodynamic equations is used. Actually, the mathematical description consists of mass balance equations for each component, equations of phase state and equations of phase equilibria relating the chemical potentials of each component in both phases, phase pressures and phase temperatures. Unfortunately, to understand the thermodynamic behavior in an open system, it is not sufficient to consider the classical thermodynamic relations for closed systems. Due to hydrodynamic coupling, the analysis of behavior in an open system becomes very complicated.

Chapter 6 presents a new type of thermodynamic model and a corresponding simulator that describes phase equilibrium in a thermodynamically open system. The new open thermodynamic simulations are validated by full compositional flow simulations using the commercial software Eclipse E300. However, the new open thermodynamic simulator is independent of space and time and hence is much faster than the fully coupled compositional simulator. Moreover, this new approach does not limit the number of chemical components in the system. Chapter 6 also offers the possibility to quantify the difference between closed and open thermodynamic systems. This judgment is based on mass conservation in an individual volume and is achieved by simple analytical relations. The closed description can be reasonable if the incoming gas in an individual volume is not different from the leaving gas; otherwise the use of the open approach is inevitable.

2.3. Efficient modeling of multiphase flow using streamline technique

This section motivates the research on streamline-based flow and transport simulation that will be discussed in more detail in Chapter 7. Multiphase-multicomponent flow in petroleum reservoir is computationally expensive to simulate. This holds in particular for large-scale reservoirs. An additional challenge for computational efficiency is posed by the need for inverse modeling, when the task is to find permeability fields such that past dynamic production data are reproduced. In this context, streamlines are often used to find an approximation of the flow and saturation equations in order to accelerate dynamic simulations. The main idea underlying the streamline-based concept is to reduce a multidimensional flow problem to a set of $1D$ flow problems along streamlines. Streamline-based flow simulation is often presented as an alternative method to the classical traditional flow modelling approaches for multi-phase flow problems based on finite differences or finite element techniques. Moreover, the streamline simulation approach is an effective technology to simulate fluid flows in large heterogeneous geological models of petroleum reservoirs. The basis of any streamline method is the simulation of the pressure field. A typical technique is presented in (Aarnes, 2004) which uses a multi-scale mixed finite-element method to compute the pressure on a coarse grid. The velocity field is used to define streamlines and then gives an estimated on an underlying finer grid using base functions fitted on the coarse grid. This numerical methodology together with streamlines gives an efficient and robust method to solve detailed flow patterns on the underlying fine grid. A significant element of any streamline technique is the introduction of the time-of-flight, i.e. a travel time coordinate, used instead of the longitudinal curvilinear coordinate along each streamline. The parameterization of streamlines with the time-of-flight was suggested in (Thiele and Batycky, 2001), (Wang and Kovscek, 2000), (Lialin and Silniov, 2005), (Batycky, 1997), (Batycky et al., 1996).

The streamline technology is also promising alternative to solve the special case of compositional flow (Christie and Clifford, 1997). In several field-scale compositional studies, serious numerical difficulties occur in conventional methods. A common solution often used is to reduce the number of cells and/or to simplify the geological complexity which affects directly the numerical performances. In opposition, the streamline approach remains unrivaled in the ability to efficiently model the transport of species along the flow paths, even in the presence of extreme permeability/porosity contrasts. In the very specific case of multi-component and multi-phase flow characterized by an extreme degree of complexity, the reduction of a three-
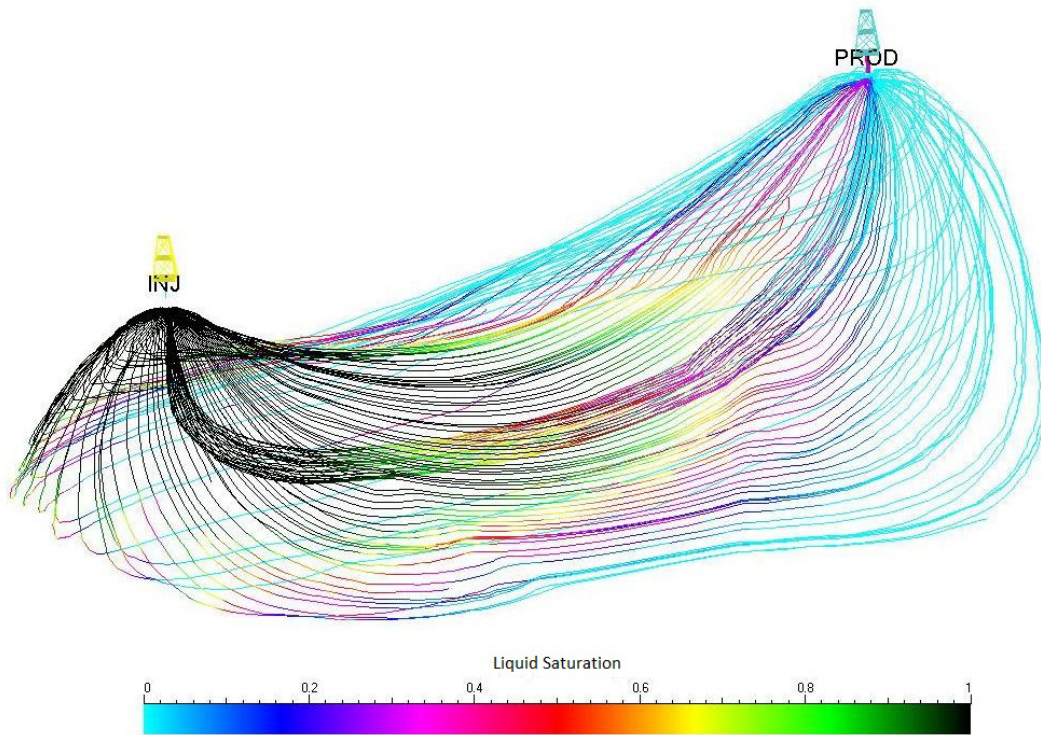
Figure 2.2.: Multiphase flows in a large-scale heterogeneous petroleum reservoir

dimensional problem to the series of one-dimensional solutions is so attractive in controlling the numerical dispersion that it cannot be overlooked, as shown in (Thiele and Edwards, 2001). For large oil fields exploited by gas injection, streamlines unlike traditional technologies offer some advantages to model full-field scenarios with a reasonable inter-well spatial resolution in acceptable run times on affordable computers.

In the present chapter, we examine one large class of a compositional flow that corresponds to gas-condensate systems which contain N chemical components like hydrocarbons from CH_4 to $C_{10}H_{22}$ and even heavier ones. The flow of such a mixture is described by the full compositional model that consists of a coupled system involving the momentum balance equations for each phase, the mass balance equations for each chemical component present in the mixture, and the thermodynamic closure relations that describe the phase equilibria and the phase states. Usually such a model represents a high order nonlinear transcendent equation system for which a constructive analysis is very difficult, so that the only method to analyze this system is numerical simulation (Aziz and Settari, 1979), (Coats, 1980). Some simplified versions of the compositional model may be obtained by reducing the number of components down to two (Koldoba and Koldoba, 2003), or by assuming a flow stationar-

ity (Whitson et al., 2003), (Dinariev, 1996). An attempt to provide a justification to such a stationary theory was done in (Chopra and D., 1986) by neglecting dispersion, capillarity and gravity. Another, much less coarse approach was developed in (Oladyshkin and Panfilov, 2007b,a, 2012) where the thermodynamic block was totally split from the hydrodynamic sub-system along streamlines, whatever the number of components. Chapter 7 will exploit this splitting of hydrodynamics and thermodynamics along the streamlines. Namely, a multi-component 1D solver has been implemented using the splitted hydrodynamic equations along the streamlines, see Figure 2.2. The equations that govern the equilibrium between phases are solved separately using a classical non-linear solver. Chapter 7 shows that the splitting of hydrodynamics and thermodynamics coupled to the streamline technology provides an effective tool to solve complex problems involving multi-compositional flow for heterogeneous 3D reservoirs.

2.4. Transport in porous media around subsurface radioactive waste storage sites

Modeling of gas transport around a radioactive waste storage asks for very large time scale simulations, and as consequence, demands high computational power. That is why questions about model complexity describing the flow in porous media and possible assumptions are still open (S. and Panfilov, 2011). In Chapter 8, we consider a multiphase compositional model containing gas and liquid phases, which consist of two main components: hydrogen and water. Chapter 8 also discusses the complexity of the underlying physical model and shows the domain of validity for such simplified relations as Henry's, Raoult's and Kelvin's law for the estimation of phase properties in comparison to full thermodynamic formulations. In additional, we offer a calibrated analytical form of the relation for hydrogen solubility in water (phases densities, viscosities and phase concentrations) which vastly accelerates further modeling. As a result, the large time scale hydrogen transport around a radioactive waste storage sites could be simulated within acceptable time frame using such calibrated relations. An Illustration of a possible scenario of hydrogen transport around the storage of radioactive waste is shown in Figure 2.3. Chapter 8 also demonstrates the influence of compositional effects on hydrogen transport in porous media. To track the appearance of phases, we introduce the novel notion of pseudo saturation.
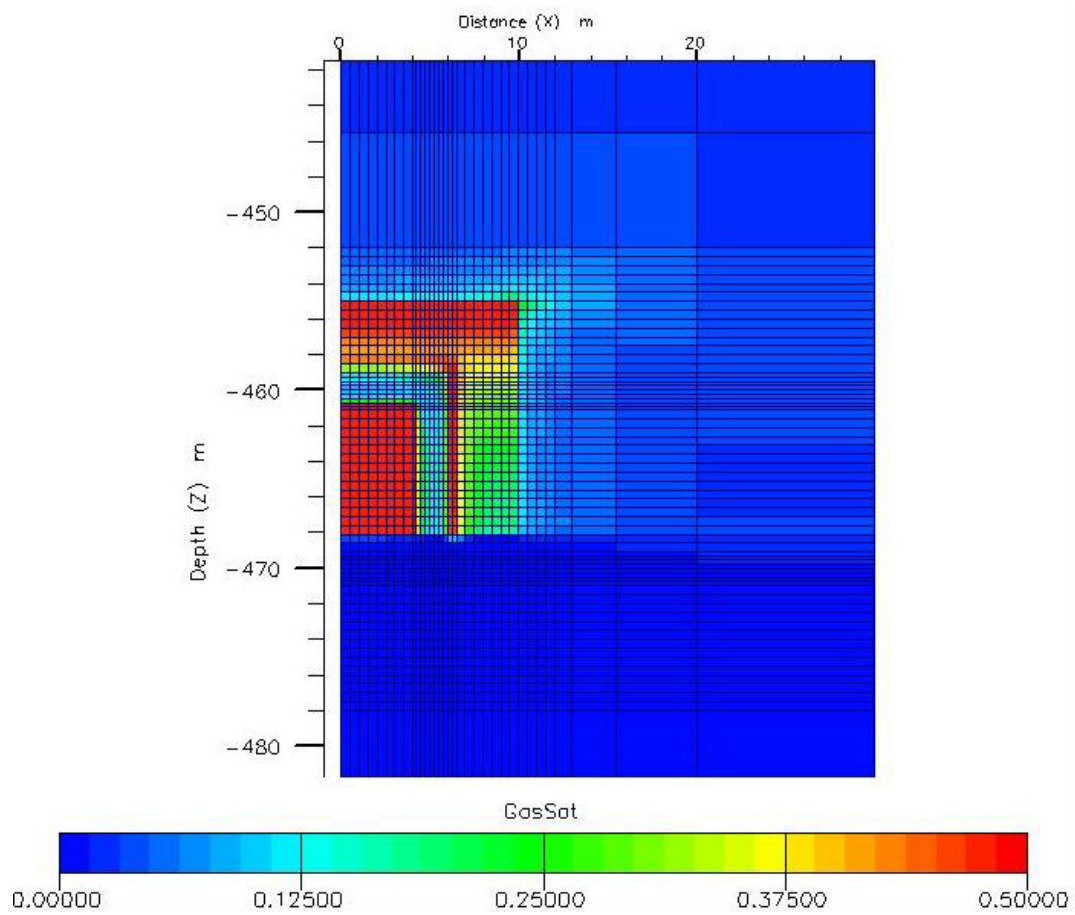
Figure 2.3.: Hydrogen transport around storage site of radioactive waste after 1000 years

3. Uncertainty in environmental modeling

For all systems in general, and for environmental systems in specific, lacking information about their system properties leads to model uncertainties up to a level where quantification of uncertainties may become the dominant question in modeling, simulation and application tasks. Unfortunately, parametric-based uncertainty can change simulation outcomes by factors several order of magnitude in many typical environmental problems. Our ability to quantify such uncertainties and related risks plays a key role and may even be more important than detailed numerical perfection of simulation codes. Current numerical simulation models are often too expensive for advanced application tasks that involve accurate uncertainty quantification, risk assessment, robust design and model calibration. Even single deterministic simulations often require parallel high performance computing. Because the involved complex flow processes in porous media have a significantly non-linear character, the problem is too non-linear for quasi-linear and other simplified stochastic tools.

Due to the complexity of the surrounding environment, the greatest challenge here is the construction of reliable and feasible models that can adequately describe physical concepts and at the same time account for uncertainty. The current Chapter is a preview on Part III of this thesis and discusses advancements in the development of approaches for these advanced challenges, based on massive stochastic model reduction techniques. The reduction is achieved via projections on orthonormal polynomial bases which form a so-called response surface. This way, the model response to changes in uncertain parameters is represented by multivariate polynomials for each output quantity of interest and allows for non-linear model-based propagation of parameter uncertainties onto the predicted quantities of interest. This technique is known as polynomial chaos expansion in the field of stochastic partial differential equations. The reduced model represented by the response surface is vastly faster than the original complex one, and thus provides a promising starting point for follow-up tasks: global sensitivity analysis, uncertainty quantification, robust design, probabilistic risk

assessment as well as model calibration. For many cases, the fact that the response surface has known polynomial properties will help to further simplify these tasks. Thus, stochastic model reduction provides an opportunity to quantify uncertainty in complex non-linear and dynamic systems at acceptable costs. The following three sections provide an overview on the Chapters 9, 10 and 11 in Part III of this thesis.

3.1. Uncertainty quantification via the polynomial chaos expansion

Throughout the following, we will consider model reduction via a response surface in a closed polynomial form. Obviously, a response surface can be constructed in different ways, e.g. it can be constructed directly on a dense Cartesian grid of input parameters at extremely high computational efforts. Likewise, conceptually straightforward numerical Monte Carlo (MC) simulation techniques are computationally demanding since the statistical accuracy of their predictions depends on the number of realizations used. Chapter 9 is exploring an alternative methodology which demands only a very small number of model evaluations to construct a response surface. Our alternative approach is through the polynomial chaos expansion (PCE). Generally, all PCE techniques can be viewed as an efficient approximation to full-blown stochastic modeling (e.g., exhaustive MC). The basic idea is to represent the response of a model to changes in variables through a response surface that is defined with the help of an orthonormal polynomial basis in the parameter space. In simple words, the dependence of model output on all relevant input parameters is approximated by a high-dimensional polynomial. This projection can be interpreted as an advanced approach to statistical regression. The PCE offers an efficient and accurate high-order way of including nonlinear effects in stochastic analysis. One of the attractive features of PCE is the high-order approximation of error propagation (Oladyshkin and Nowak, 2012b) as well as its computational speed when compared to MC (Oladyshkin et al., 2011b), (Oladyshkin et al., 2010).

The original PCE concept can be used only for Gaussian distributed input parameters. Unfortunately, natural phenomena and uncertainty in engineering are often not that simple, and the distribution of physical or model parameters often cannot be considered Gaussian. However, it is possible to put into conformity a physical variable with a normal variable by an adequate transformation called Gaussian anamorphosis or normal score transformation (e.g.
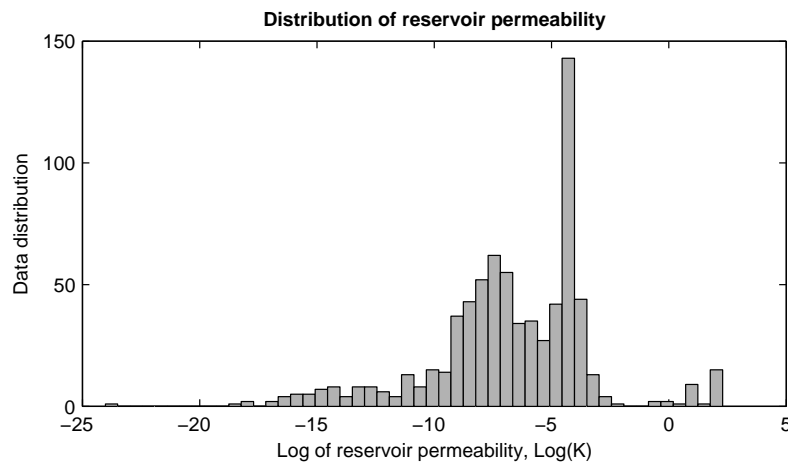
Figure 3.1.: Distribution of reservoir permeability

(Wackernagel, 1998)) or approximate parametric transformations (Ditlevsen and Madsen, 1992). Using transformed variables for expansion cannot be considered an optimal choice because it leads to slow convergence of the expansion (e.g. (Xiu and Karniadakis, 2002a), (Xiu and Karniadakis, 2003)). In recent years, the classical PCE technique was extended to the generalized polynomial chaos (gPC) which accommodates for the use of an increased, yet limited number of statistical distributions. The PCE methods discussed above assume an exact knowledge of the probability density functions of all input parameters and they are optimal only when applied to a finite number of certain parametric probability distributions.

Unfortunately, information about the distribution of data or input parameters is very limited in many realistic applications, especially in environmental engineering and sciences (see also (Red-Horse and Benjamin, 2004)). Data that characterize model parameters often indicate a variety of statistical distribution shapes (e.g., bounded, skewed, multi-modal, discontinuous, etc). Applied research on (partially) natural or complex realistic systems often faces the problem of limited information about the model parameters and even about their probability distributions. For example, material properties of underground reservoirs are insufficiently available to provide a full picture of their distribution. Moreover, the statistical distribution of model parameters can be nontrivial, e.g., bounded, skewed, multi-modal, discontinuous, etc. Figure 3.1 illustrates the distribution of reservoir permeability from a public data base (Comunian and Renard, 2009). Also, empirical parameter distributions derived from raw data sets do in general not follow analytically known distribution shapes. Any attempt to construct probability density functions of any particular shape from samples of limited size or from sparse information introduces additional subjectivity into the analysis, which bears

the severe risk of leading to biased results. For such reasons, application tasks demand further adaptation of the chaos expansion technique to a larger spectrum of distributions.

In the light of thesis challenges, Chapter 9 generalizes the PCE to the arbitrary polynomial chaos (aPC) that accommodates for a wide range of data distributions. The aPC adapts to arbitrary probability distribution shapes of input parameters and, in addition, can even work with unknown distribution shapes when only a few statistical moments (Oladyshkin and Nowak, 2012b),(Oladyshkin et al., 2011a),(Oladyshkin and Nowak, 2012a) can be inferred from limited data or from expert elicitation. In fact, our equations will show explicitly (in closed form) that statistical moments are the only source of information that is propagated in all polynomial expansion-based stochastic approaches. Thus, exact probability density functions do not have to be known and do not even have to exist. The arbitrary distributions for our framework can be either discrete, continuous, or discretized continuous. They can be specified either analytically (as probability density/cumulative distribution functions), numerically as histogram or as a raw data sets. The available information can directly and most purely be used in stochastic analysis, when using our data-driven formulation of aPC. The aPC approach provides improved convergence in comparison to classical PCE techniques, when applied to input distributions that fall outside the range of classical PCE.

3.2. Assessing the impact of model parameters via global sensitivity analysis

Understanding the general role of parameters in models and the impact of varying model parameters on the response of prediction models is a relevant subject in various fields of science and engineering. Characterizing the impact of parameter variations is known as sensitivity analysis and can be subdivided into local and global analysis. In many cases of practical interest, we wish to perform a GSA in order to analyze a model as such or to investigate, quantify and rank the effects of parameter variation or parameter uncertainty on the overall model uncertainty. GSA can also be used to: (1) quantify the relative importance of each individual input parameter in the final prediction; (2) aid engineers to produce more robust designs; and finally (3) help decision makers to allocate financial resources towards better uncertainty reduction. For example, the field of subsurface contaminant hydrology requires uncertainty estimates due to the ubiquitous lack of parameter knowledge caused by spatial heterogeneity of hydraulic properties in combination with incomplete characterization. For

such reasons, we need to rely on probabilistic tools to predict contaminant levels and their overall health effects, and to quantify the corresponding uncertainties (Oladyshkin et al., 2011d). Having efficient computational approaches to estimate uncertainty and to perform GSA in many fields of science and engineering that feature uncertain dynamic or distributed systems is desirable. It can inform modelers about the relevance of processes or parameters in the models they compile, and can inform engineers and decision makers about which parameters require most attention and where characterization efforts should be allocated such that prediction uncertainty can be minimized. Hence, there is an ever-increasing demand for having a GSA method that efficiently quantifies uncertainty and parameter relevance in complex and non-linear systems.

Chapter 10 tackles GSA based on the aPC technique, following the line of work on aPC (Oladyshkin and Nowak, 2012b), (Oladyshkin et al., 2011a). Because the presented framework accounts for arbitrary bounds or weighting functions for input parameters, it provides a weighted global sensitivity (Oladyshkin et al., 2011d). In some sense, the novel sensitivity indexes introduced here can be perceived as a generalization of the Morris method (Morris, 1991) to weighted analysis. Up to presence, the Morris method considers a uniform importance of input parameters within predefined intervals. We also generalize the Sobol indices (Sobol, 1990) for GSA to the aPC context, and provide a novel GSA measure which resembles a weighted square norm of sensitivities. Compared to Sobol indices, the new weighted sensitivity measure is absolute rather than relative. The advantage of an absolute index over a relative one is that it is a quantitative expression for the (averaged) derivative (slope), and hence keeps the original meaning of a sensitivity as known from linear, local analysis. The resulting sensitivity measure shows better convergence properties than Sobol analysis.

Performing GSA requires to evaluate the model at many points in the space of the input parameters. The correct choice of such points within the parameter space is important for adequate and efficient assessment of sensitivity. The aPC approach explicitly offers a method for optimal choice of these points, based on the generalized mathematical theory of Gaussian integration, see (Oladyshkin et al., 2011a). The big advantage of aPC-based GSA (or more generally: GSA based on any PCE technique), is that one can obtain global sensitivity information at computational costs that are hardly larger than those for local analysis (Oladyshkin et al., 2011d), (Oladyshkin et al., 2011c). The reason is the following: Local methods use infinitesimally small spacing between parameter sets for model evaluation to get numerical derivatives evaluated at a single point. The aPC based-method places the parameter sets for model evaluation at an optimized spacing in parameter space. This can be interpreted as fitting secants (or polynomials for non-linear analysis) to the model response. These secants

(polynomials) approximate the model over the entire parameter space in a weighted least-square sense. This is more beneficial to computing a tangent or local second derivatives (compare FORM, SORM methods) that approximate the model well just around one point in the parameter space.

Overall, this provides an alternative procedure to perform GSA that is computationally efficient and highly flexible. In particular, due to aPC, the GSA can be interpreted as exploiting a smart (mathematically optimal) interpolation rule of model output between optimally chosen sets of input parameters, where the number of model evaluations is minimal. Compared to earlier works that related GSA to classical PCE our presented approach: (1) emphasizes a more engineering-like language as compared to otherwise intense mathematical derivations and is based on a clear 3-step procedure to perform sensitivity analysis, (2) provides easy-to-use semi-analytical expressions for frequent use in applications, (3) generalizes PCE-based GSA to arbitrary probability distributions of the investigated parameters, and moreover, (5) allows to align the complexity and order of analysis with the reliability and detail level of statistical information on the input parameters.

3.3. Model calibration via Bayesian updating

Chapter 11 offers a novel combination of methods that improves the tension between accuracy and efficiency in model calibration problem. The construction and calibration of accurate simulation models for complex environmental systems that reproduce the observed behavior of real systems is a very challenging task. This task is a particular challenging in real-world applications, where one features uncertainties and complexities at each stage of modeling. Once a virtual model has been established, inverse modeling and history matching to past production data becomes an extremely important issue in order to improve the quality of prediction. The accuracy of inversion or history matching depends on the quality of the established physical model concepts (including, e.g. seismic, geological and hydrodynamic characteristics, fluid properties, and mathematical laws fro the hypothesized processes, etc.), and on the accuracy of the involved parameter calibration, stochastic inversion or data assimilation techniques. The quality will also depend on the computational efficiency of all involved methods (Oliver and Chen, 2011), because too large computational time would have to be mitigated by compromises in numerical or inversion accuracy. The principal challenge of history matching is that one wishes to obtain a model that reproduces the full behavior of the featured environmental system from a only limited number of local measurements

(Gavalas et al., 1976). The reliability of history matching is increasing with the number of available observations.

History matching is very well known in the field of reservoir engineering (Makhlouf et al., 1993). Traditionally, an iterative manual process of trial and error (see e.g. (Williams et al., 1998)) is applied to adjust the reservoir geological model in order to reproduce past observations of oil or gas production. Such a manual approach is very popular among experts in reservoir engineering and demands a very strong understanding of geology and processes. However, the non-trivial and non-linear interaction of the matched parameters within the model can complicate the history matching procedure a lot (Oliver et al., 2001). Instead of the manual technique, formal optimization methods such as gradient search or the adjoint method (see (Gao and Reynolds, 2006; Li et al., 2003; Rodrigues, 2006)) can be applied. Unfortunately, the mentioned optimization approaches often lead to high computational costs and cannot be easily applied for complex real-world tasks.

Another important point about history matching techniques is that they can produce non-unique solutions (Oliver et al., 2008), which means that several virtual models and parameter sets can match the observation data equally well. In fact, this problem is common to most inverse problems (Sun, 1999; Tarantola, 2005). Stochastic approaches can handle such type of uncertainty occurring during the matching procedure without the need to introduce regularization or to artificially restrict the parameter space. Their result is a probability distribution of possible parameters sets instead of a single best estimation. As an improvement of classical optimization, several successful stochastic approximation methods have been adapted for the history matching problem (Bangerth et al., 2006; Gao and Reynolds, 2007; Li et al., 2011). However, stochastic approaches are more expensive than classical optimization-based (deterministic) calibration techniques, because they need to explore the full range of possible model outcomes with many model runs. In particular, this requires to draw samples from the conditional distribution of the parameters as equally likely calibrated parameter sets. This can be done, e.g., via Markov chain Monte Carlo (Gilks et al., 1996), Bootstrap filtering (Smith and Gefland, 1992), GLUE (Leube et al., 2012a) or rejection sampling (Smith and Gefland, 1992). A particular reason for the high computational costs is that, when exploring the full range of possible model outcomes, many model runs are rejected upon comparison with the data. Therefore, the high computational costs of forward modeling get multiplied by large factors in stochastic calibration methods.

The Ensemble Kalman filter (EnKF) method is one of the simplest yet most successful ways to transfer Bayesian theory (see e.g. (Smith and Gefland, 1992)) to practice for model updat-

ing and forecasting. The EnKF (Evensen, 2006) is derived from a first-order second-moment approximation of error propagation for Bayesian updating. The Special Issue in Computational Geosciences (Naevdal et al., 2011) was fully devoted to (Ensemble) Kalman filtering for model updating. As practically successful attempt to accelerate inverse modeling, the EnKF recently received a lot of attention for history matching (e.g. (Aanonsen et al., 2009; Naevdal et al., 2005)). It is a comparatively cheap method that can generate reasonable history-matched models for real fields. Due to its foundation on first-order second-moment analysis, it is optimal only if all involved model parameters, model states and data follow a joint multi-Gaussian distribution. Thus, the EnKF has a theoretical limitation which does not allow it to deal with strongly non-linear problems. For example, the papers (Wang et al., 2010) and (Zafari and Reynolds, 2007) pointed out that the EnKF suffers from non-linearity and inconsistency in matching.

Therefore, further advancements can be achieved by using more accurate non-linear approaches to Bayesian updating, when combined with sufficiently accurate (and more expansive) model reduction techniques. In fact, techniques based on response surfaces or other surrogate models have lately been combined with more accurate versions of Bayesian updating. In (Jin, 2008) and (Marzouk et al., 2007), the polynomial chaos expansion has been applied to accelerate Bayesian updating via Markov chain Monte Carlo (MCMC) methods. Even the EnKF has been combined with polynomial chaos expansion for accelerated and accurate computation of the required covariance matrices, see (Saad and Ghanem, 2009), (He et al., 2011) and also related work (Pajonk et al., 2012). However, more accurate Bayesian updating approaches incorporate higher order stochastic information on the input parameters (e.g., in form of high-order statistical moments) (Ghanem and Spanos, 1991; Le Maitre and Knio, 2010; Wiener, 1938) which goes beyond the scope of the Gaussian assumption. Also, the Gaussian assumption in stochastic inversion typically goes along with (possibly implicit) linearization of model dependencies. Hence, more accurate Bayesian updating also requires to work with high-order approximations of the involved models than linear ones. Using a more accurate updating rule for higher moments would be inadequate when the involved surrogate model is too inaccurate (e.g. by linearizing strongly non-linear model dependencies). As a consequence, model reduction techniques also need to retain the non-linearity of models at sufficiently high orders.

Chapter 11 offers an advanced methods for statistical (Bayesian) model calibration, which takes into consideration the non-linearity of the forward model and of inversion. The presented approach combines fully accurate Bayesian updating via Bootstrap filtering (Oladyshkin et al., 2013a,b) (or, alternatively, via MCMC) with a model reduction based on

the arbitrary polynomial chaos expansion technique (Oladyshkin and Nowak, 2012a; Oladyshkin et al., 2011a). It provides the necessary flexibility to handle arbitrary distributions of model parameters, including high-order statistical moments. It can also handle the conditional parameter distributions that occur after Bayesian updating. Bootstrap filtering is the most direct yet simple numerical implementation of Bayes theorem, which can account for arbitrary non-linear model equations and for arbitrary distribution shapes in comparison to (Ensemble) Kalman Filters. Hence, Bootstrap filtering is a perfect match for combination with the aPC technique. This combination is superior to the previous combinations of the classical PCE with (Ensemle) Kalman Filtering, because the latter two components are both optimal only for (multi-) Gaussian distributions. At the end of Chapter 11 we demonstrate a straightforward implementation of aPC-based Bayesian updating on the benchmark problem of CO_2 leakage. However, the surrogate model may be very inaccurate and lead to wrong results, when the prior information is strongly offset against reality. This is caused by the fundamental property of all chaos expansion techniques that the error of approximation is lowest where the (prior) probability density is high, i.e., large errors may occur in low-probability regions. For that case, we introduce at the end of Chapter 11 also an advanced iterative approach for aPC-based Bayesian updating. It allows to perform Bayesian updating even in the case where the prior assumptions on model parameters are far from reality, such that the response surface has to be re-iterated in order to be accurate in the relevant regions of high posterior probability. The iteration can be seen as an extension of statistical inversion based on successive linearization to successive high-order expansion.

4. Feasibility of applied environmental tasks

Modeling of environmental systems often demands such tasks as prediction of possible scenarios, design of planned engineering actions, estimation of risks, calibration of the model using available data etc. The common core of all tasks is multi-query evaluation of the underlying model with varying modeling parameters. Regardless of that, the overall modeling procedure should be performed within an acceptable time frame. However, ignoring or neglecting, for the sake of computational efficiency, of some of the components or processes in the system could be very dangerous. Indeed, the modeling process in the field of environmental engineering is a chain of many tasks. Like any chain, it is only as strong as its weakest link. Estimation of uncertainty is part of the chain and underestimation of its influence can weaken that link immensely. An adequate combination of a reliable physical model with advanced stochastic tools offers a powerful base for the forecasting of environmental behavior. The current Chapter focuses on applications of model reduction technique to problems of energy relevant gas storages in geological formations and discuss related to that obstacles. It offers an overview over the works presented in Part IV of this thesis, and addresses the corresponding Chapters 12 to 16 in the following sections.

4.1. Robust design of an environmental project under uncertainty

Environmental systems require statistical concepts to support risk assessment and decision making. The task of finding optimal system design while accounting for uncertainty leads to robust designs. In specific, risk can be controlled to stay below a given risk acceptance level, if robust design is performed under corresponding constraints. The modeling parameters can be separated into two classes: design or control parameters that can be chosen by
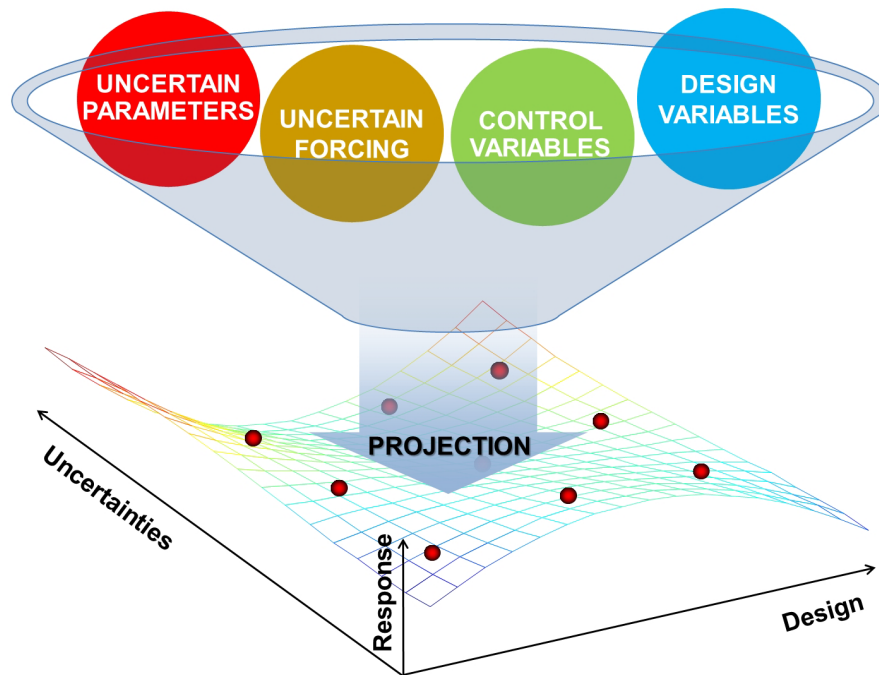
Figure 4.1.: Integrative response surface

the operator of a system, and uncertain parameters that describe our (incomplete) knowledge of the system properties. For example, in the CO_2 injection problem, the latter include permeability, porosity etc. On the other hand, the system's performance and failure probability will also depend on design parameters such as the injection rate or depth. Evidently, the decision making will depend on the interplay between the response to design parameters, system uncertainty and, finally, the probability of failure or some statistical expectation of benefit. Robust design directly accounts for uncertainties and failure probabilities. This way, it can minimize risks during operation.

Chapter 12 proposes to combine the model response to all design variables and uncertain parameters into a single approach based on an integrative (joint) response surface (see Figure 4.1), which allows to reflect the non-linear dependence of the original model on all these parameters. Based on this integrative concept, the design task explicitly includes uncertainty, which leads to robust designs with minimum failure probability. This integrates the design task into the reduced stochastic model, allowing us to find robust designs with controlled risks at low computational costs. Namely, designing system behavior for maximum load at a guaranteed specific safety level can be achieved by quick Monte-Carlo methods on the resulting polynomials. It assures that the known risks remain below a chosen acceptable level. In the example of underground gas storage, one could choose the injection rate and depth

to avoid geomechanical cracking at 99% safety level. Chapter 12 also draws attention to the possible significant impacts of incorporating uncertainty into risk predictions and design tasks, in comparison to deterministic modeling approaches. We demonstrated that neglecting parametric uncertainty in design can be an overly strong and dangerous simplification for the design tasks. Due to the non-linearity of processes, including uncertainty can lead to a systematic and significant shift of the predicted values, affecting both risk estimates and the design of injection scenarios. Thus, design under uncertainty that openly admits uncertainty and seeks for roust solutions is much more reliable in comparison to conventional design.

The subsequent Chapter 13 continues this line of thoughts and applies robust design to assess the pressure increase in a sand channel system during carbon dioxide injection. The pressure increase is important to consider, since too large pressures might lead to caprock failure near the injection well and might result in new leakage pathways. Far away from injection, the pressure increase also has to be taken into account because it might influence the pressure in other storage reservoirs and might reduce the overall storage capacity. Additionally, brine displacement due to pressure increase has to be investigated since infiltration into higher (drinking) water aquifers needs to be avoided. Chapter 13 closes by demonstrating how robust design helps to assess the maximum allowable injection rate in the presence of uncertain structural parameters of the reservoir, and how well-done uncertainty quantification and risk assessment helps to guide further reservoir exploration efforts.

4.2. Data-driven uncertainty quantification in environmental systems

Applied research on real-world environmental projects, like modeling underground reservoir, gas storage, groundwater flow, etc., often faces the problem of immensely limited information about the model parameters involved. Unfortunately, precise information on distribution shapes for all uncertain input parameters is very rare in realistic applications. With only limited data available or even in total absence of data, not even probability density functions representing the lack of knowledge can be easily inferred in a justified manner. Moreover, even if some amount of data is available, the statistical distribution of the corresponding model parameters can be nontrivial, e.g. bounded, skewed, multi-modal or discontinuous. Several examples of real data types can hardly be described by known families of theoretical statistical distributions. Sophisticated approaches to assigning distributions to data include

maximum likelihood or moment-fitting of selected parametric distributions that satisfy the personal perception of the modeler. Due to the typically low numbers of parameters (e.g. two or three), this resembles a subjective and highly restrictive assumption on the distribution shape, its higher-order moments, and its extreme-value tails. Methods of maximum entropy (Jaynes, 1982) and minimum relative entropy (Woodbury and Ulrych, 1993) are often used in the engineering sciences to construct probability distributions from sparse prior information (mostly in the form of a few statistical moments and bounds). The required information may be available from different sites with supposedly similar conditions. Although these two methods are designed to minimize subjectivity, they are heavily debated within the statistics community (e.g. (Sambucini, 2007)) and in fact still impose a specific assumption on distribution shape. Thus, in any case, the attempt to construct probability density functions from samples of limited size or from sparse information introduces additional room for subjectivity into the analysis. Chapter 14 provides a striking example of bias introduced by subjectivity, which can appear in various environmental problems. Apparently, subjectivity in assigning distribution shapes is not only a philosophical problem, but can severely bias or even invalidate all subsequent results, including uncertainty quantification and risk assessment.

Applied tasks demand the direct handling of arbitrary distribution shapes and sparse data sets without additional assumptions. Therefore, Chapter 14 continues by presenting a highly parsimonic and yet purely data-driven description of uncertainty. We apply a very recent polynomial chaos expansion technique that can be applied to arbitrary distributions of data, which has been discussed in detail in Chapter 9. Our derivation explicitly shows that statistical moments are the only source of information that is propagated in all types of polynomial expansion-based stochastic approaches. This avoids the necessity of assuming subjectively or speculating on the exact shapes of probability distributions. The new freedom opens the path to accessing with the chaos expansion technique even those applications where data samples of limited size merely allow the inference of a few moments, and one would not be able to construct a probability density function without introducing subjective assumptions and hence dangerous sources of bias. In a small fictitious example in Chapter 14, we asked independent experts to select and fit parametric distributions to two raw data sets describing reservoir properties. The example demonstrated that subjectivity in data interpretation can easily lead to prediction bias. Chapter 14 offers also validation of the discussed data-driven stochastic approach by Monte Carlo simulation using a common 3D benchmark problem of CO_2 storage. The proposed approach yields a significant computational speed-up compared with Monte Carlo, and provides faster convergence than conventional polynomial

chaos expansions. Even for small degrees of expansion, the data-driven expansion can be very accurate, which can save a lot of computational power for probabilistic risk analysis.

4.3. On risk estimation including various levels of uncertainty

Risk assessment requires dealing with uncertainties. In a broader sense, uncertainty can be defined as the absence of information or knowledge. In the field of uncertainty quantification, one often distinguishes between two categories, i.e. epistemic and aleatory uncertainty (e.g., (Helton, 1994), (Hoffman and Hammonds, 1994), (Wojtkiewicz et al., 2001)). Epistemic uncertainty is also denoted as reducible, subjective, or type B uncertainty and results from a lack of information about the system. Such an uncertainty can be reduced by further research or measurements. Aleatory uncertainty is also known as variability, irreducible, stochastic, or type A uncertainty and results from the fact that a system can react in many different unpredictable ways. Some authors additionally distinguish between dimensions of uncertainties (e.g., (Rowe, 1994) and (Walker et al., 2003)). (Walker et al., 2003) divides uncertainties not only into different natures (epistemic and aleatory) but also into the location and the level of uncertainty. The location identifies where the uncertainty is situated within the model complex, whereas the levels categorise the uncertainty into ranges of knowledge. In this study, we follow the approach of (Walker et al., 2003) and categorize the uncertainties into levels. Levels of knowledge range from determinism to total ignorance. Determinism means that everything is known exactly, which would be an ideal situation. Total ignorance is the other extreme, where nothing is known, and one does not ever know a complete list of uncertainty sources. The levels in between are statistical uncertainty, scenario uncertainty and recognized ignorance.

Statistical uncertainties can still be described by statistical terms (e.g., statistical moments, probability density functions, or histograms). In the case of environmental problems, such uncertainties can be found, for example, in permeability, porosity, or anisotropy values. The next level is scenario uncertainty (or conceptual uncertainty). Such uncertainties, e.g. the geological layering considered in the model or a gap in the caprock, cannot be described sufficiently by statistical terms. Adequate scenarios have to be chosen properly by considering the opinion of experts in the field of interest, e.g., geologists. Measurements and site exploration typically reduce uncertainty. For example, drilling an exploration well together with

a seismic campaign might substantially improve the knowledge of number and orientation of distinct geological features, while still leaving much uncertainty of their specific properties. When considering uncertainty due to geological features, there can be a continuous transition between these two categories of uncertainty. The modeler must consider both the unknown natural conditions and interpretation of available expert opinion. Recognized ignorance is defined by (Walker et al., 2003) as fundamental uncertainty about the mechanisms and functionalities.

Chapter 15 deals with risk estimation including various levels of uncertainty. Namely, we distinguish between scenario and statistical uncertainty that occur during the modeling of geological CO_2 storage. Addressing scenario uncertainty involves expert opinion on relevant geological features such as caprock properties, faults, distinct geological layers, etc. Chapter 15 investigates scenario uncertainty by simulating 6 scenarios with different geological setup and subsequent ranking according to the resulting damage. Accounting for statistical uncertainties is performed using a massive stochastic model reduction via polynomial chaos expansion. The overall approach estimates the probabilities of certain hazardous events, which leads to quantification of risks. We also would like to point out that, dependent on the availability of data, both scenario and statistical uncertainty can be equally significant. The proposed ideas can be easily transferred to other problems such as natural-gas storage or radioactive-waste disposal.

4.4. Large-scale global sensitivity analysis in geological storage

Once one deals with a real-world large-scale environmental problem, the freedom to observe diverse scenarios and handle various uncertainty sources in large time/space scales could be strongly reduced because of high computational demands. However, many question are open even at the stage of site screening (Nordbotten and Celia, 2011). For example, ranking the important model parameters based on their influence on the model responses can support a better understanding of the system, and it can result in a better design of subsequent studies on the stochastic nature of the process. In an approach to quantify the impact of geological heterogeneity on model predictions of multiphase flow in geological formations, a large number of shallow marine depositional realizations has been generated and used in the sensitivity analysis of the impact of geological uncertainties on production forecasting
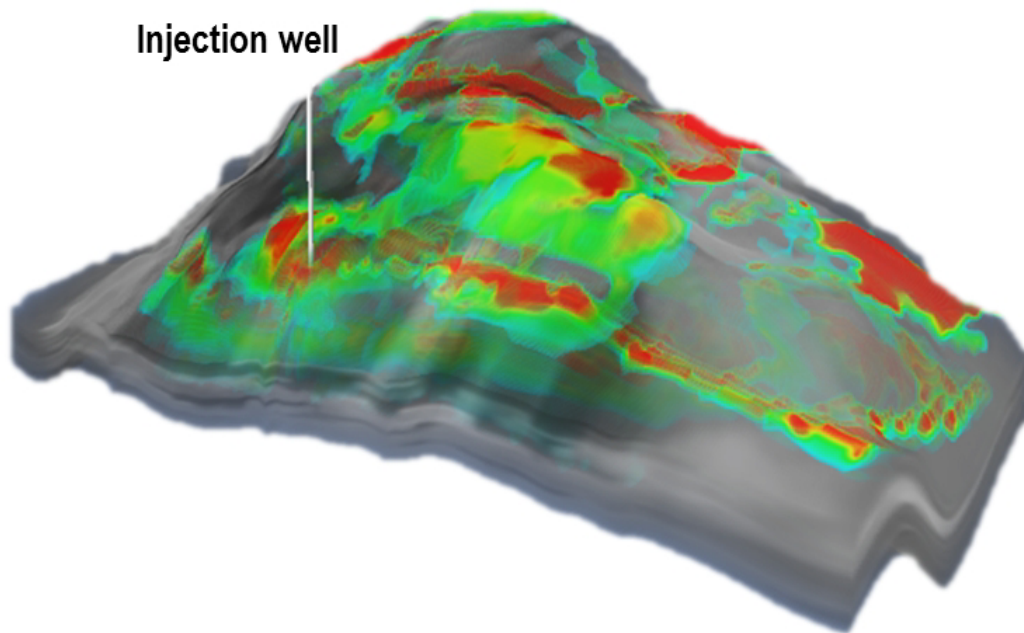
Figure 4.2.: Possible scenario of carbon dioxide displacement in a potential storage site

(SAIGUP), see (Howell et al., 2008; Manzocchi et al., 2008a; Matthews et al., 2008). The main general conclusion of that study is that realistic features of geological uncertainty in modeling (other than typical hydrological parameters) can lead to considerable uncertainties in prediction. Morover, the European Commission and the United States Environment Protection Agency recommend using sensitivity analysis in the context of extended impact assessment for policy making (Commission, 2002).

Chapter 16 tests and demonstrates the applicability of the recent set of methods discussed in the previous chapters to perform a sensitivity analysis and to assess the risks for a realistic large-scale problem of CO_2 storage (Ashraf et al., 2013), which is illustrated in Figure 4.2. We make use of global sensitivity analysis rather than a local one, because local analysis fails to cover the non-linear variation of model responses over the entire range of probability distributions of the input parameters. Namely, we use Sobol indices (Sobol, 2001) constructed from a chaos expansion (Oladyshkin et al., 2011d) for sensitivity analysis. The parameters we consider are the level of barriers presence, aggradation angle, fault transmissibility, and regional groundwater effects. The considered features are the structural and depositional features that dictate the distribution of hydrological parameters such as permeability and porosity, both in terms of value and spatial distribution. Additionally, Chapter 16 discusses and estimates risks which can appear in the analyzed geological formation. The proposed

framework offers a chance to observe the diverse type of behavior of a potential geological storage site and can inform about dangerous behavior already at the stage of site screening. To the best of our knowledge, the current study is the first one that implements the proposed mathematical analysis tools on realistic geological structural parameters at reservoir scale.

Part II.

From complexity to feasibility for environmental modeling

Environmental systems are one of the biggest and most important classes of complex dynamic systems. Many environmental systems are dominated by real-time influences of external driving forces. Unfortunately, a complete picture of the environmental system is not available, especially if one tries to analyze behaviors below the surface, such as groundwater flow, petroleum reservoirs, radioactive waste deposits, storage of gases, etc. Namely, environmental data is hardly available and expensive to acquire. That is why we need adequate and efficient physical model concepts for the reconstruction of a full and complex picture of the environment. In this context, models play a very important role in the understanding of environmental behaviors and offer a unique way to predict behaviors of complex systems.

Flow and transport processes in subsurface porous media often involve several phases, such as water, oil, gas, and so forth. These process are governed by complex and non-linear multi-physics and multi-scale laws which could take into consideration instable multiphase flow, phase transitions, rock deformation, geochemistry, temperature effects etc. Although these laws are traditionally derived for small scales, many modeling tasks require considering impacts on very large spatial and time scales as well. Due to the complexity of the surrounding environment, numerical simulation models for multiphase flow and transport processes are often too expensive for realistic large-scale applications. Even the increasing power of parallel high performance computing offers only limited help and can never completely solve that problem. The big difficulty here is the definition of consistent and feasible environmental models that can provide adequate conceptual descriptions and that can simultaneously maintain a reliable time frame of simulations.

The current Part II focuses on several conceptual particularities of environmental modeling and offers several possibilities to accelerate the modeling process. Chapter 5 briefly discusses different sources of possible error within an entire modeling challenge and illustrates their possible impact using an application to carbon dioxide storage. Chapter 5 underlines the possibly greatest overall challenge which consists in finding a healthy and reasonable compromise between numerical techniques, conceptual physical formulations, estimation of model parameters and their uncertainties, computational effort, etc. Chapter 6 presents an efficient handling of thermodynamic properties of a compressible compositional two-phase flow in porous media which demands an extremely short time of simulation. Chapter 7 presents an effective framework based on the streamline technique for multiphase multi-component flows in large-scale heterogeneous petroleum reservoirs. Chapter 8 deals with hydrogen migration around a storage of site of radioactive waste. This Chapter offers a calibrated analytical form of hydrogen solubility in water (phases densities, viscosities and phase concentrations) which vastly accelerates further modeling.

5. On the impact of modeling error sources: application to CO$_2$ storage

Bibliographic Note: The content of this chapter is based on the following original article:
Oladyshkin S., Class H., Helmig R., Nowak W., A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. Advances in Water Resources, Elsevier, V. 34, P. 1508-1518, 2011.

Model-based uncertainty analysis can help to judge the potentials and hazard in many engineering applications better. This requires to specify the probability distributions of all model parameters, posing a huge demand on data availability or requiring highly subjective assumptions on distribution shapes to compensate for missing data. In a small fictitious example we demonstrate that this subjectivity can easily lead to substantial prediction bias, and that the subjective choice of distribution shapes has a similar relevance as uncertainties due to physical conceptualization, numerical codes and parameter uncertainty.

5.1. Introduction to modeling carbon dioxide storage

CO$_2$ storage in geological formations is currently being discussed intensively as an interim technology with a high potential for mitigating CO$_2$ emissions. In recent years, greate research efforts have been directed towards understanding the physical processes in CO$_2$ storage (e.g. (IPCC, 2005)). The multiphase flow and transport processes involved are strongly non-linear. They include phase changes in the region of th critical point and effects such as gravity-induced fingering and convective mixing as well as geo-chemical and geo-mechanical processes etc. In order to describe the space-time evolution of the CO$_2$ plume and the influence of potentially leaky abandoned wells, (semi-) analytical solutions

have been derived by Nordbotten et al. (Nordbotten et al., 2005a, 2009). A comparison study of various simplifying semi-analytical models with complex simulations models was discussed by Ebigbo et al. (Ebigbo et al., 2007). The analysis in (Birkholzer et al., 2009) focused on the effects of large-scale CO$_2$ leakage through low-permeability layers. Various optimization strategies for monitoring surface leakage using near-surface measurement approaches were developed in (Cortis et al., 2008). Recently, Class et al. (Class et al., 2009) published a benchmark study comparing a number of mathematical and numerical models of varying complexity. In the current chapter, we would like to attract attention to the often ignored or forgotten fact, that efforts invested in improved physical conceptualization, numerical codes and stochastic modeling can easily be overwhelmed by error through human subjectivity in data interpretation and in histogram analysis at a very early stage of modeling (see Section 5.3). The approach developed in this chapter will help overcome this problem, especially in data-sparse situations when only global databases for similar sites and expert opinions are available. A direct application of our proposed framework can be considered best at the stage of site screening, when considering relatively low-parametric large-scale models, prior knowledge and measured data on large scales. In more detailed site-specific studies, more resolved structural models will be more useful, as defined site investigation and new data allow the spatial resolution of geological heterogeneity and Bayesian updating.

Modeling underground CO$_2$ storage involves many conceptual and quantitative uncertainties (Hansson and Bryngelsson, 2009a). The lack of information on subsurface properties may lead, depending on the specific question at hand, to parameter uncertainties up to a level where the uncertainties dominate or even override the influence of secondary physical processes (see (Oladyshkin et al., 2010)). A significant part of the scientific community still refrains from considering uncertainty in modeling, but the corresponding arguments are discussed and rejected one by one in (Pappenberger and Beven, 2006). The current chapter attracts attention to the often ignored or forgotten fact, that efforts invested in improved physical conceptualization, numerical codes and stochastic modeling can easily be overwhelmed by error through human subjectivity in data interpretation and in histogram analysis at a very early stage of modeling.

5.2. Physical problem formulation

We consider the benchmark leakage problem of injected CO$_2$ into overlying formations through a leaky well defined by Class et al. (Class et al., 2009). In the current chapter,

we present an illustrative example for a homogenized system, i.e. we consider spatial heterogeneity only through zonation according to different geological media. However, the technique presented can be extended to many classes of heterogeneous systems, where spatially correlated heterogeneous parameter fields can be decomposed into their uncorrelated principal components using the KL-expansion (e.g. (Li and Zhang, 2007)), if heterogeneity does not span over too many scales. Figure 5.1 illustrates a 2D section of the 3D domain under consideration. The top of the figure is not the ground surface, but the transition to yet another aquitard. CO_2 is injected into a deep aquifer, spreads within the aquifer and, upon reaching a leaky abandoned well, rises to a shallower aquifer. The goal of the simulation is to quantify the leakage rate which depends on the pressure build-up in the aquifer due to injection and on the plume evolution. The leaky well is at the center of the domain and the injection well is 100m away. Both aquifers are 30m thick and the separating aquitard has a thickness of 100m. The leaky well is modeled as a porous medium with a higher permeability than the formation. The CO_2 leakage rate is defined in the benchmark study as the total CO_2 mass flux integrated over a control plane midway between the top and bottom aquifer, divided by the injection rate, in percent.

We will consider a relatively simple benchmark model to illustrate the concept presented in the current chapter. The simplicity of the benchmark model chose is fully motivated by the high computational demand of Monte Carlo simulations, which we use to validate the
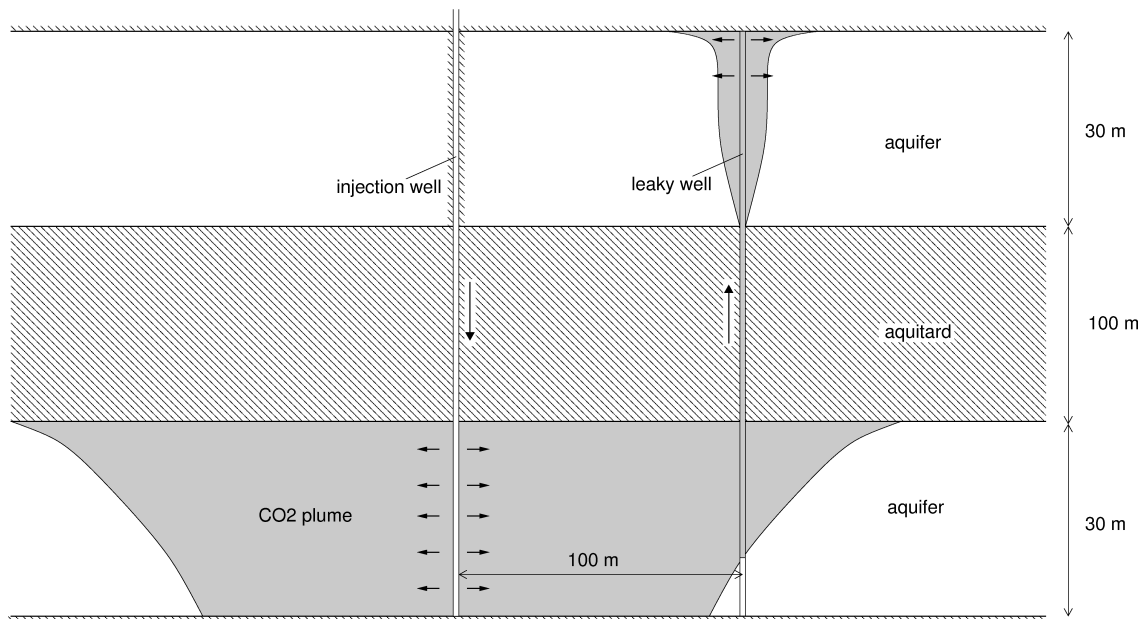
Figure 5.1.: Benchmark problem for CO_2 leakage through an abandoned well.

proposed concept (see details in Section 14.3.2). The benchmark problem assumes that fluid properties such as density and viscosity are constant, all processes are isothermal, CO$_2$ and brine are two separate and immiscible phases, mutual dissolution is neglected, the pressure conditions at the lateral boundaries are constant over time, the formation is isotropic rigid and chemically inert, and capillary pressure is negligible. Within our approach, the physical complexity of the problem can be increased to an arbitrary extent because of the non-intrusive black-box conception of the probabilistic collocation method (Isukapalli et al., 1998), (Li and Zhang, 2007). The initial conditions in the fully saturated domain include a hydrostatic pressure distribution which is dependent on the brine density. The aquifers are initially filled with brine. The initial pressure at the bottom of the domain (at 3000m depth) is 3.086×10^7 Pa. The lateral boundary conditions are constant Dirichlet conditions and equal to the initial conditions. All other boundaries are no-flow boundaries. CO$_2$ is injected at a constant rate of 8.87 kg/s, which corresponds to $1600 m^3/d$ at reservoir conditions. Simulation time is 1000 days. All relevant parameters used for the simulation are given in Table 5.1. Mass balances of the two phases and the multiphase version of Darcy's Law give the following system of

Parameter	Value
CO$_2$ density, ρ_g	479kg/m^3
Brine density, ρ_w	1045kg/m^3
CO$_2$ viscosity, μ_g	$3.950 \cdot 10^{-5}$Pa s
Brine viscosity, μ_w	$2.535 \cdot 10^{-4}$Pa s
Aquifer permeability, K$_A$	$2 \cdot 10^{-14}$m^2
Aquifer thickness	30m
Aquitard thickness	100m
Leaky well permeability, K$_L$	$1 \cdot 10^{-12}$m^2
Porosity, ϕ	0.15
Leaky & injection well radius	0.15m
Injection rate	8.87kg/s (1600m^3/d)
Distance between wells	100m
Dimensions of model domain	$1000m \times 1000m \times 160m$
Simulation time, t	1000days

Table 5.1.: Benchmark parameters

differential equations:

$$-\phi\frac{\partial S_g}{\partial t} - \nabla \cdot \left\{ \frac{k_{rw}}{\mu_w}\mathbf{K} \cdot (\nabla p - \rho_w \mathbf{g}) \right\} - q_w = 0,$$

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

(5.1)

which is constrained by the equation

$$S_w + S_g = 1. \tag{5.2}$$

The subscripts w and g stand for the brine (water) phase, and the CO_2-rich (gas) phase, respectively. The primary variables in equation (5.1) are the gas-phase saturation S_g and pressure p. S_w is the brine-phase saturation. The relative permeabilities k_{rw} and k_{rg} are secondary variables and linear functions of S_w and S_g, ($k_{rw} = S_w = 1 - S_g$; $k_{rg} = S_g$), $\mathbf{g}$ is the gravity vector, $\mathbf{K}$ is the absolute permeability tensor and q_w, q_g are sources/sinks. In the current study, the benchmark problem has simulated using DuMuX, a multi-scale multi-physics toolbox for the simulation of flow and transport processes in porous media (Flemisch et al., 2007).

5.3. Impact of error sources

In the previous section, we discussed the convergence and minimal subjectivity of our data-driven concept. However, stochastic approximation and subjectivity are, of course, not the only sources of possible error within an entire modeling challenge. The possibly greatest overall challenge consists in finding a healthy and reasonable compromise between numerical techniques, physical formulations and model selection, mathematical descriptions, estimation of model parameters and their uncertainties, computational effort, etc. In this section, we would like to set the corresponding magnitudes of possible errors in relation to the possible error range of subjectivity, again using the above benchmark problem as quantitative example. Our intention here is not to reduce the importance of numerical schemes and a proper consideration of conceptual uncertainty. Instead, we wish to draw attention to subjectivity in data interpretation as an equally important source of overall error.

We will subdivide the overall error into the following four categories: I-numerical approximation error; II-conceptual model error; III-error due to parameter uncertainty; IV-error due to subjectivity in data analysis (see Figure 5.2). The shaded areas in Figure 5.2 illustrate the

ranges of variations obtained using different concepts, approaches, techniques and choices within each category.

The first type of error is illustrated in the upper left plot of Figure 5.2, where CO_2 leakage was predicted using different numerical codes within an international code intercomparison study. The correspondence of individual lines to numerical codes is provided in (Class et al., 2009). The shaded area illustrates the difference between the lowest and highest prediction. To illustrate the second type of impact, i.e. conceptual model error, we consider two alternative conceptual models, one with and one without phase transition, see the upper right plot of Figure 5.2. We choose this physical phenomenon for our current illustration, because it has a strong impact on multi-phase flow prediction in underground reservoirs. Moreover, the two lines bounding the shaded area in the upper right plot of Figure 5.2 represent the averaged solution (through different numerical codes), which makes this plot independent of the numerical choice. The upper two plots in Figure 5.2 were reproduced with kind permission of the authors of (Class et al., 2009). The third type of impact is illustrated in the bottom left plot of Figure 5.2, where the consequences of parameter uncertainty (such as reservoir permeability and porosity) are presented. Here, we plot the standard deviation for the mean value of CO_2 leakage over time. The fourth type of impact, i.e. the influence of subjectivity in data interpretation, is presented in the bottom right plot of Figure 5.2.

The current example shows that the influence of different numerical schemes may be less important than different physical formulations, the influence of uncertainty and subjectivity. Also, we can see that parameter uncertainty has a large impact on modeling CO_2 storage even in comparison with the considered refinement of physical phenomena. Therefore, physical refinement is only meaningful if the uncertainty of newly introduced parameters is also accounted for. Also, a balanced compromise between allocating computer resources on refined models, better stochastic approximation and numerical accuracy has to be kept in mind. Moreover, this demonstrates that subjectivity within data interpretation can cause the same order of magnitude in uncertainty or error as numerical, conceptual and stochastic approximation error, at least in the example considered here.

The list of sources for possible errors discussed here is well known. However, to the best of our knowledge, the impact of subjectivity in data interpretation especially for non-linear stochastic problems has not been discussed in the literature before. For generalization and abstraction from the example considered here, we expect that more complex models with more parameters (each one requiring its own assumptions) increase the number of opportunities to introduce subjectivity. Moreover, more complex models often exhibit stronger
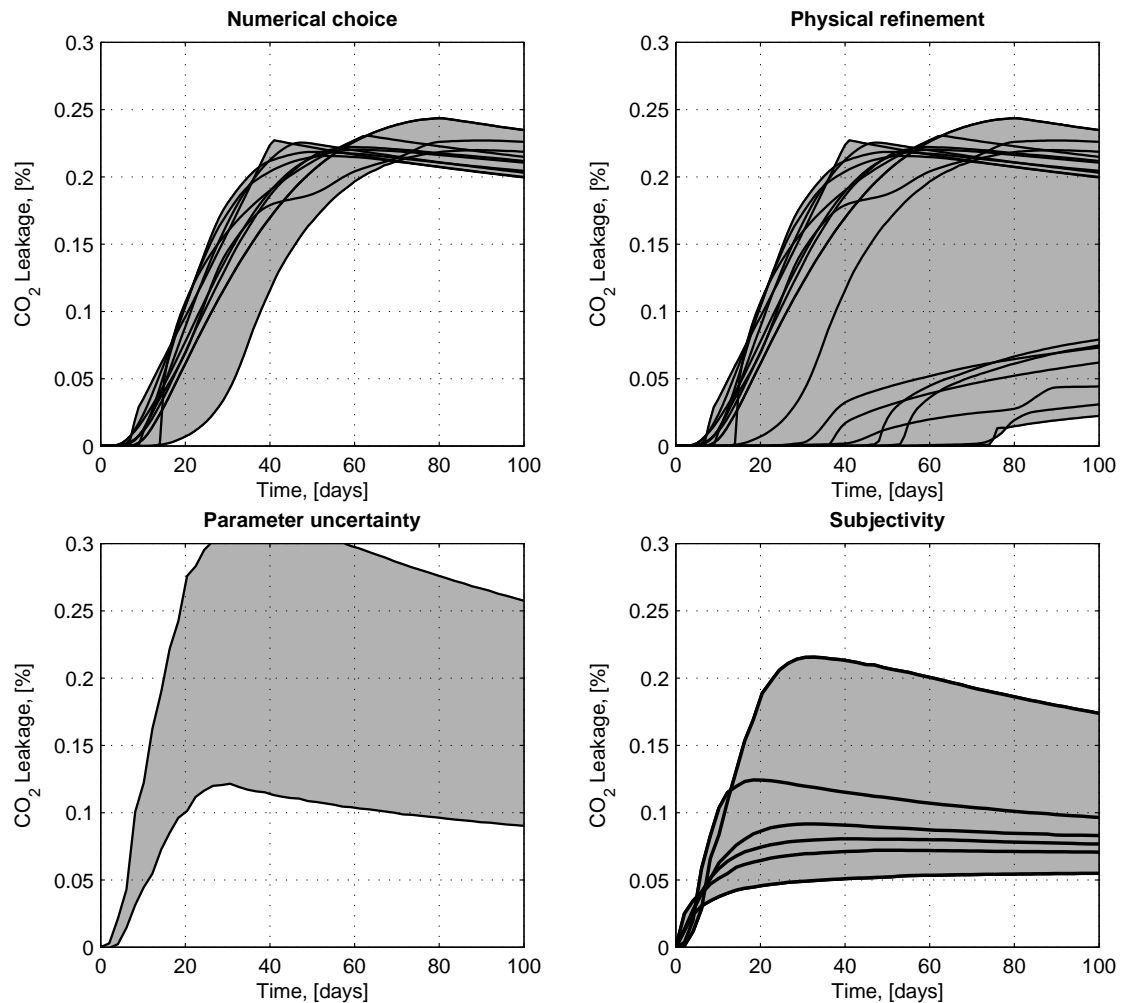
Figure 5.2.: Variation of CO_2 leakage: impact of numerical choice, conceptual model choice, parameter uncertainty and subjectivity

non-linearity, which can amplify the impact of subjectivity especially for assumptions on extreme value tails and higher-order moments.

5.4. Conclusions

We demonstrated that subjectivity in data interpretation can easily lead to prediction bias. With the considered example, we also illustrated that subjectivity can be a source of error that plays in the same league as physical conceptualization, numerical code development, and stochastic modeling. Moreover, the modeling process is a chain of many tasks. Like any chain, it is only as strong as its weakest link. Our results indicate clearly that the statistical

treatment of input data is part of the chain, and that the subjectivity in assuming theoretical curves can weaken that link immensely.

6. Thermodynamic modeling of multicomponent two-phase flow

Bibliographic Note: The content of this chapter is based on the following original article: Oladyshkin S., Panfilov M. Open thermodynamic model for compressible multicomponent two-phase flow in porous media, Journal of Petroleum Science and Engineering, Elsevier, V. 81, P. 41-48, 2012.

The compressible compositional two-phase flow in porous media was considered. This system was characterized by two essential parameters: the perturbation parameter and the relative mobility parameter. The perturbation parameter represents the stabilization time of the pressure field after perturbation. The relative mobility parameter represents the interior characteristic of the system by the ratio between liquid mobility and gas mobility. For the contrast phase mobility and fast pressure relaxation process an open thermodynamic model was obtained by splitting between thermodynamics and hydrodynamics (HT-splitting). This model includes several differential thermodynamic equations for characterization of the transport phenomena (Delta Law), i.e. describes the equilibrium in an open system. A new open thermodynamic simulator (OTS) was developed. The OTS simulations are validated by the full compositional flow simulations using Eclipse. However, compared to the full compositional simulator an evident advantage of the new OTS simulator is the fact that due to the independence from space and time, data construction demands an extremely short time of calculation. Moreover this new approach does not limit the number of chemical components in the system.

The closed thermodynamic system was shown as behaving similarly to the open system for several test cases. In order to characterize the difference between the closed and the open thermodynamics we proposed to use Delta Law for one chemical component and a variation function for the combination of all components. This behaviour depends on mass conservation in an individual volume. The closed description can be realistic if the incoming

gas in an individual volume is not different from the leaving gas; otherwise the use of the open approach (new or classical) is suggested.

6.1. Introduction

We shall consider the compositional fluid which consists of N chemical components. The components are able to form two thermodynamic phases separated from one another by an interface. Usually there are three types of components in the system: light (N_2, CH_4, C_2H_6, ...), heavy (C_5H_{12}, C_6H_{14}, ...) and neutral (C_3H_8, C_4H_{10}, ...). Usually such a model represents a high order nonlinear transcendent equation system for which a constructive analysis is very difficult, so that the only method to analyze this system is the numerical simulation (Aziz and Settari, 1979), (Coats, 1980). Some simplified versions of the compositional model may be obtained by reducing the number of components up to two (Koldoba and Koldoba, 2003), or by assuming a flow stationarity (Whitson et al., 2003), (Dinariev, 1996). An attempt to provide a justification to such a stationary theory was done in (Chopra and D., 1986) by neglecting dispersion, capillarity and gravity.

We will focus on a system without chemical reactions. However the mass exchange between phases will be one of the important mechanisms for the thermodynamic system. Such compositional fluid in porous media takes the form of a finely dispersed system with a characteristic scale of liquid drops or gas bubbles of the order of pore size. The interface between the phases is then highly dispersed and irregular. Under such conditions the examined system may be effectively described in terms of a phenomenological approach when both phases are considered as two interpenetrating continua, so that each of
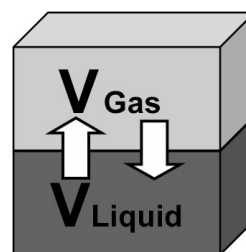
Figure 6.1.: Closed thermodynamic system: phase exchange

them occupies the whole space. Moreover, the system may be assumed to be in local equilibrium, due to a highly developed interface area ensuring fast mass transfers.

There are two typical physical scenarios of thermodynamic behaviors in the compositional system. The first one corresponds to the closed thermodynamic system behaviors. This type of behaviors means that the phase transition between a gas sub-volume V_g and a liquid sub-volume V_l happens in a closed volume $V = V_g + V_l$, llustrated in the Figure 6.1. Physically this type of behaviors corresponds to the thermodynamic experience in closed PVT bomb. The

thermodynamic theory of these systems is well developed in fundamental works: (Sedov, 1976), (Nigmatulin, 1987), as well as in reservoir engineering: (Coats, 1980), (Nikolaevski et al., 1968), (Firoozabadi, 1999), (Danesh, 1998), (Batalin et al., 1992).

The second type of behaviors corresponds to the open thermodynamic system behaviors. In this case there is not only phase exchange between the two individual sub-volumes of gas V_g and liquid V_l, but also the transport of one relatively another. Practically this type of system is a very common one, and can be represented by a petroleum reservoir of a compositional fluid with mass exchange between phases, where gas moves relatively liquid. Due to the transport in the open system, the individual sub-volume of

Figure 6.2.: Open thermodynamic system: phase exchange & transport

gas V_g is replaced by a new one V_g^{new}, which is illustrated in the Figure 6.2.

Unfortunately, in order to understand the thermodynamic behaviors in an open system it is not sufficient to consider the classical thermodynamic relations for a closed system. Usually for this kind of analysis a full compositional model consisting of classical thermodynamic relations as well as hydrodynamic equations is used. Actually, the mathematical description of this model consists of N equations of mass balance for each component, two equations of phase state and $N + 2$ equations of phase equilibria relating the chemical potentials of each component in both phases, phase pressures and phase temperatures. Usually such a model represents a high order nonlinear transcendent equation system which may be solved numerically by using various iteration procedures. Finally, due to hydrodynamic coupling, the analysis of behaviors in an open system becomes very complicated. In this chapter we propose a new type of thermodynamic model and a corresponding simulator that describes phase equilibrium in a thermodynamically open system.

6.2. Mathematical model of two-phase flow in porous media

6.2.1. Hydrodynamic system formulation

We shall consider the full compositional model of two-phase flow consisting of N chemical components. The components form two thermodynamic phases: gas and liquid, which moved in porous medium. We shall consider the phases as the interpenetrating continua, according to the phenomenological approach, (Nikolaevski et al., 1968), (Nigmatulin, 1987), (Sedov, 1976). The general dynamic of compositional flow can be described by the mass balance equation for each chemical component k and momentum balance equation for each phase (the Darcy law):

$$\phi \frac{\partial}{\partial t}\left(\rho_l c_l^{(k)} s + \rho_g c_g^{(k)}(1-s)\right) = \mathrm{div}\left(\rho_l c_l^{(k)} V_l + \rho_g c_g^{(k)} V_g\right), k = 1,...,N \qquad (6.1)$$

$$V_l = -\frac{Kk_l}{\mu_l}\mathrm{grad}P_l, \quad V_g = -\frac{Kk_g}{\mu_g}\mathrm{grad}P_g$$

where superscript k refers to k-th chemical component ($k = 1,...,N$), indexes g and l to gas and liquid; ϕ is the porosity; $c_g^{(k)}$ or $c_l^{(k)}$ is the mass concentration of k-th component in gas or liquid $[kg/kg]$; s is the volume saturation of pores by liquid; ρ_g and ρ_l are the phase densities; k_g and k_l are the relative permeabilities; μ is the dynamic phase viscosity; K is the absolute permeability; P is the phase pressure.

6.2.2. Thermodynamic system formulation

The thermodynamic part describes the local equilibrium in the two-phase system and consists from N equilibrium equations (6.2a) for each chemical component, two equations of phase state (6.2b), two normalizing equations for concentrations (6.2c) and the equilibrium

equation for phase pressures in the capillary field:

$$v_g^{(k)}\left(P_g, \left\{c_g^{(q)}\right\}_{q=1}^N\right) = v_l^{(k)}\left(P_l, \left\{c_l^{(q)}\right\}_{q=1}^N\right), \qquad k=1,...,N \tag{6.2a}$$

$$\rho_g = \rho_g\left(P_g, \left\{c_g^{(q)}\right\}_{q=1}^N\right), \quad \rho_l = \rho_l\left(P_l, \left\{c_l^{(k)}\right\}_{q=1}^N\right) \tag{6.2b}$$

$$\sum_{k=1}^N c_g^{(k)} = 1, \quad \sum_{k=1}^N c_l^{(k)} = 1 \tag{6.2c}$$

$$P_g - P_l = P_c(s) \tag{6.2d}$$

here $v_i^{(k)}$ is the chemical potential of kth component in i^{th} phase; $P_c(s)$ is the capillary pressure function. Equations of state (6.2b) are written in general form. Functions $v_g^{(k)}(P,...)$ are given.

6.2.3. Thermodynamic variance in two-phase systems

We will note that in the case of a gas-condensate or an oil-gas system, the capillary pressure is always large as the surface tension is low due to familiar phase compositions (Panfilova and Panfilov, 2004): $P_c =\sim 0.01 - 0.1$ bar. Thus, we will assume that the capillary pressure P_c can be neglected with respect to the characteristic pressure difference over the domain, ΔP, i.e. the capillary number Ca is large or the flow velocity is rather important. So, the system has a unique pressure:

$$P_l = P_g \equiv P \tag{6.3}$$

Hence, the thermodynamic system (6.2a), (6.2b), (6.2c), (6.3) does not contain saturation s, as the phase densities and concentrations are independent of s.

The hydrodynamic system consists of N equations (6.1) written with respect to pressures P and saturation s and $N-2$ independent concentrations. Such concentrations represents the basis and can be denoted as $\chi^{(k)}$:

$$\chi^{(j)} = c_g^{(j)}, \qquad j=1,...,N-2 \tag{6.4}$$

On the other hand, other concentrations and phase densities are determined from the thermodynamic system (6.2a), (6.2b), (6.2c), (6.3) which includes $(N+4)$ equations with respect to $(2N+3)$ variables: P, ρ_g, ρ_l, $\left\{c_g^{(q)}\right\}_{q=1}^N$, $\left\{c_l^{(q)}\right\}_{q=1}^N$. According to the Gibbs rule of phases, the difference between the number of variables $2N+3$ and the number of equations $N+4$

represents the thermodynamic variance of the system $v = N - 1$, which determines the number of independent variables. We will select the pressure and $N - 2$ basic concentrations as the independent variables, and hence another variable (phase densities and other concentrations) can be find from the thermodynamic system. Basis concentration $\chi^{(j)}$ and pressure P can be computed with help of the hydrodynamic system (6.1). This mean that the hydrodynamic system strongly couplet with thermodynamic one. Thus, the classical approach for analysis of an open system requires solving of the full compositional model, which represents a huge nonlinear hydro-thermo dynamic problem. The current chapter presents the study on compositional two-phase flow and provides important simplification which helps to overcome direct solution of the couplet thermo-hydro system.

6.3. Flow model properties for an open contrast two-phase system

6.3.1. Open system characterization

Evidently such natural system as underground petroleum reservoirs or another one non isolate from exterior exchange represent class of open system. To characterize this type of system we will introduce two essential parameters: the perturbation parameter and the relative mobility parameter.

For that, first of all, let us introduce the the characteristic values of the length, time, pressure, viscosities and densities as L, t_*, ΔP, μ_g^0, μ_l^0, ρ_g^0 and ρ_l^0. Let $\langle K \rangle$ and $\langle \phi \rangle$ be the mean values of permeability and porosity. Hence, the compositional system (6.1) can be transformed to a dimensionless canonical one ((Panfilova and Panfilov, 2004)), where each equation has a strict physical meaning: first equation describes the total flow, the second one describes the liquid flow, while the remaining $N - 2$ equations will describe the transport of basic chemical

components:

$$\varepsilon \frac{\partial}{\partial \tau} \left(\varphi_l \overline{\rho} s + \varphi_g (1-s) \right) = \text{div} \left(\left[\psi_g k_g + \omega \psi_l k_l \right] \text{grad} p \right), \tag{6.5a}$$

$$\varepsilon \left(\overline{\rho} \frac{\partial (\varphi_l s)}{\partial \tau} + \overline{\rho} \varphi_l s \frac{\partial \zeta_l^{(N)}}{\partial \tau} + \varphi_g (1-s) \frac{\partial \zeta_g^{(N)}}{\partial \tau} \right) = \tag{6.5b}$$

$$\omega \text{div} \left(\psi_l k_l \text{grad} p \right) + \omega \psi_l k_l \text{grad} p \cdot \text{grad} \zeta_l^{(N)} + \psi_g k_g \text{grad} p \cdot \text{grad} \zeta_g^{(N)},$$

$$\varepsilon \left[\overline{\rho} \varphi_l s \frac{\partial}{\partial \tau} \left(\zeta_l^{(k)} - \zeta_l^{(N)} \right) + \varphi_g (1-s) \frac{\partial}{\partial \tau} \left(\zeta_g^{(k)} - \zeta_g^{(N)} \right) \right] = \tag{6.5c}$$

$$\omega \psi_l k_l \text{grad} p \cdot \text{grad} \left(\zeta_l^{(k)} - \zeta_l^{(N)} \right) + \psi_g k_g \text{grad} p \cdot \text{grad} \left(\zeta_g^{(k)} - \zeta_g^{(N)} \right), k = 1, ..., N-2$$

Here the operations div and grad are performed in dimensionless space coordinates; other definitions are the following:

$$p \equiv \frac{P}{\Delta P}, \quad \varphi_i \equiv \frac{\phi \rho_i}{\langle \phi \rangle \rho_i^0}, \quad \psi_i \equiv \frac{K \rho_i \mu_i^0}{\langle K \rangle \rho_i^0 \mu_i}, \quad \tau \equiv t/t_*, \quad \overline{\rho} \equiv \frac{\rho_l^0}{\rho_g^0}, \quad t^* \equiv \frac{L^2 \mu_g^0 \langle \phi \rangle}{\langle K \rangle \Delta P}$$

The parameter t^* is called the global reservoir relaxation time and is equal to the time of propagation of the perturbation caused by a pressure variation ΔP. The functions $\zeta_g^{(k)}$ and $\zeta_l^{(k)}$ is the component neutralities, which describes how much the concentration variation within a fixed phase $(dc_i^{(k)})$ is higher/lower than the concentration variation between two phases $(\Delta c^{(k)} = c_l^{(k)} - c_g^{(k)})$:

$$d\zeta_l^{(k)} \equiv \frac{1}{\Delta c^{(k)}} dc_l^{(k)}, \qquad d\zeta_g^{(k)} \equiv \frac{1}{\Delta c^{(k)}} dc_g^{(k)} \tag{6.6}$$
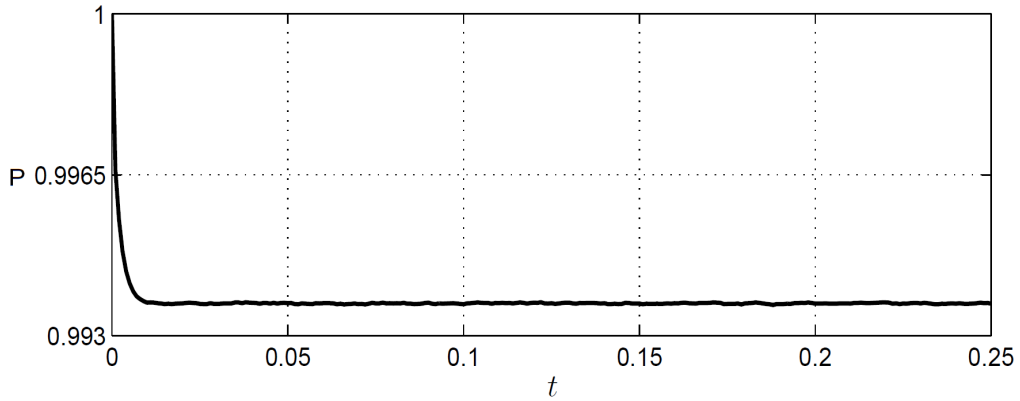
Figure 6.3.: Pressure stabilization: dimensionless pressure p and time t

Finally, the perturbation ε and relative mobility ω characteristic parameters for the considered system can be presented in the following form:

$$\varepsilon \equiv \frac{t^*}{t_*}, \qquad \omega \equiv \frac{\rho_l^0 \mu_g^0}{\rho_g^0 \mu_l^0} \tag{6.7}$$

Physically the parameter ε characterize the time of stabilization of the pressure field after perturbation, i.e. the ratio of perturbation propagation time to the external process time t_*. Usually this time of stabilization quite small relatively the characteristic time of processes in the system, (Oladyshkin and Panfilov, 2005). Such physical phenomena assure the small value of the parameter ε. One illustration for pressure behaviors after the well creation in several gas-condensate reservoir presented on the Figure 6.3 ($t^* = 0.005$ - dimensionless stabilization time). Thus, the perturbation parameter ε is an exterior characteristic for the considered system. Another interior characteristic for the system can be presented by the ratio between liquid mobility and gas mobility, parameter ω. Usually the liquid is less mobile relatively to the gas in systems as gas-condensate and gas-liquid, i.e. characterized by the small parameter ω.

6.3.2. Splitting thermodynamics and hydrodynamics

For the contrast phase motilities and fast stabilizing pressure field the limit form of the gas-liquid flow model can be obtained (Oladyshkin and Panfilov, 2007a). Duo to the limit behaviors of the system (6.5) when $\omega \rightarrow 0$ and $\varepsilon \rightarrow 0$ the last equation characterizing the transport of basic components can be transformed (Oladyshkin and Panfilov, 2007a) into the following sub-system (Delta Law) along the streamlines:

$$\frac{1}{\Delta c^{(k)}} \frac{dc_g^{(k)}}{dp} = \frac{1}{\Delta c^{(N)}} \frac{dc_g^{(N)}}{dp}, \quad k = 1, ..., N-2 \tag{6.8}$$

This sub-system is a time-independent and space-independent system of ordinary differential equations, and this sub-system has a thermodynamic character. Hence the group of transport equations (6.5c) can be used in thermodynamic formulation (6.8), and finally we obtained the splitting between the thermodynamics and the hydrodynamics:

Hydrodynamic part determines pressure p and saturation s:

$$\operatorname{div}\left(\psi_g k_g \operatorname{grad} p\right) = 0, \tag{6.9a}$$

$$\overline{\rho}\,\varphi_l \frac{\partial s}{\partial \tau} + \overline{\rho}\,\varphi_l s \frac{\partial \zeta_l^{(N)}}{\partial \tau} + \varphi_g(1-s)\frac{\partial \zeta_g^{(N)}}{\partial \tau} =$$

$$\frac{\omega}{\varepsilon}\operatorname{div}\left(\psi_l k_l \operatorname{grad} p\right) + \frac{1}{\varepsilon}\psi_g k_g \operatorname{grad} p \cdot \operatorname{grad} \zeta_g^{(N)} \tag{6.9b}$$

Thermodynamic part determines $N-2$ basic concentrations $\chi^{(k)}$, $N+2$ other concentrations, and two phase densities ρ_l and ρ_g as the functions of pressure p:

$$v_g^{(k)}\left(p,\left\{c_g^{(q)}\right\}_{q=1}^{N}\right) = v_l^{(k)}\left(p,\left\{c_l^{(q)}\right\}_{q=1}^{N}\right), \quad k=1,...,N \tag{6.10a}$$

$$\rho_g = \rho_g\left(p,\left\{c_g^{(q)}\right\}_{q=1}^{N}\right), \quad \rho_l = \rho_l\left(p,\left\{c_l^{(k)}\right\}_{q=1}^{N}\right) \tag{6.10b}$$

$$\sum_{k=1}^{N} c_g^{(k)} = 1, \quad \sum_{k=1}^{N} c_l^{(k)} = 1 \tag{6.10c}$$

$$\frac{1}{\Delta c^{(k)}}\frac{dc_g^{(k)}}{dp} = \frac{1}{\Delta c^{(N)}}\frac{dc_g^{(N)}}{dp}, \quad k=1,...,N-2 \tag{6.10d}$$

The new thermodynamic system is represented by the group of classical equations and the new differential thermodynamic equations (Delta Law). Finally, the new thermodynamic system depends on pressure only. This means that for known pressure all thermodynamic variable (concentrations, phases densities) can be obtained.

The interpretation of the new delta law consists in the description of chemical components transport in an open two-phase system (illustration on the Figure 6.2). This means the obtained thermodynamic system is the open one, in which two types of principals processes are described: phase equilibrium (classical relations) and transport (Delta Law). Physically there are phase transitions between individual volume of gas and individual volume of liquid, and the gas volume replaced by the new one due to a faster gas flow relatively liquid. Thus, this open thermodynamic model presents a new field for development of an alternative solvers in relation to the classical approach.

6.4. Open thermodynamic model

The full mathematical formulation of the open thermodynamic model consists of the equilibrium equations for chemical potentials (6.10a), the equations of phase state (6.10b), the concentration normalizing relations (6.10c), the total composition variation in an open system (Delta Law, 6.10d) and thermodynamic closure relations (see details in Section 6.4.3).The details of formulation for phase equilibrium conditions of two-phase fluid and equations of state for the multicomponent systems described in Section 6.4.3 and Section 6.4.1 correspondently.

Based on the described model (6.10) the open thermodynamic simulator was developed for the open multicomponents two-phase system. Very important to note that such a type of simulator produces the data on phase equilibrium for the class of stabilized and contrast systems (discussed in the section 6.3.1).

6.4.1. Equations of state for the multicomponent systems

Equations of state are basically developed for pure components, and applied to multicomponent systems by using some mixing rules which determine the averaged parameters for a mixture (Peng and Robinson, 1976), (Orbey and Sandler, 1998). The mixing rules take into account the prevailing forces between molecules of different substances which form the mixture.

The EOS like Peng-Robinson is usually formulated in a form with respect to the z-factor z as an implicit function of P, T and concentrations:

$$z_i^3 - (1 - B_i)z_i^2 + (A_i - 3B_i^2 - 2B_i)z - (A_iB_i - B_i^2 - B_i^3) = 0,\ i=g,l \tag{6.11}$$

where z_i is z-factor of ith phase; coefficients A_i, B_i depend on pressure P, temperature T and mixture composition:

$$A_i = \bar{a}_i \frac{P}{R^2T^2}, \quad B_i = \bar{b}_i \frac{P}{RT}, \quad i=g,l \tag{6.12}$$

Parameters $\bar{a}_i$ and $\bar{b}_i$ describe the mixture composition and the mixing rules:

$$\bar{a}_i = \sum_{k=1}^{N} \sum_{j=1}^{N} x_i^k x_i^j a^{kj}, \quad a^{kj} = \sqrt{a^k a^j}(1 - \delta^{kj}), \quad \bar{b}_i = \sum_{k=1}^{N} x_i^k b^k, \quad i=g,l \tag{6.13}$$

where δ^{kj} is known as a binary interaction parameter (Danesh, 1998); x_i^k is the mole fraction of kth component in ith phase.

Parameters a^k and b^k for kth chemical component are defined by the following equations:

$$a^k = 0.4274800232 \frac{R^2 T_{cr}^{k\,2}}{P_{cr}^k} \left[1 + m^k \left(1 - \sqrt{\frac{T}{T_{cr}^k}} \right) \right]^2, \tag{6.14a}$$

$$m^k = 0.37464 + 1.5422\omega^k - 0.26992\omega^{k\,2}, \tag{6.14b}$$

$$b^k = 0.086640350 \frac{R T_{cr}^k}{P_{cr}^k}, \quad k = \overline{1,N} \tag{6.14c}$$

where T_{cr}^k and P_{cr}^k are the critical fluid temperature and pressure for component k; ω^k is the acentric Pitcer factor.

Cubic equation (6.11) for the z-factor may be solved for liquid and vapor phases. Generally three solutions are obtained. The smallest root corresponds to liquid, while the largest root describes vapor.

6.4.2. Phase equilibrium conditions for two-phase fluid

The equality of chemical potentials of each component in all the co-existing phases at equilibrium ((Danesh, 1998),(Batalin et al., 1992)) is described by equation (6.10a). The chemical potential for gas $v_g^{(k)}$ and liquid $v_l^{(k)}$ phases can be defined as:

$$v_i^{(k)} \left(P, \left\{ c_i^{(q)} \right\}_{q=1}^N \right) = v_{ip}^{(k)} \left(P, T, \left\{ c_i^{(q)} \right\}_{q=1}^N \right) + RT \ln \left(f_i^{(k)} c_i^{(k)} \right), \tag{6.15}$$

$$k = \overline{1,N}; i = g,l$$

where R is the universal gas constant; T and P are the reservoir temperature and pressure; $f_i^{(k)}$ is the fugacity of component k in phase i; $v_{ip}^{(k)}$ is the chemical potential of the ideal gas:

$$v_{ip}^{(k)} \left(P, \left\{ c_i^{(q)} \right\}_{q=1}^N \right) = G_p^{(k)} (P,T) + RT \ln \left(c_i^{(k)} \right), \quad k = \overline{1,N}; i = g,l \tag{6.16}$$

herein $G_p^{(k)} (P,T)$ is the chemical potential of the pure component k.

Therefore, using (6.16) and (6.16), the phase equilibrium conditions (6.10a) lead to a balance between the fugacities:

$$f_g^{(k)} \left(P, T, \left\{ c_g^{(q)} \right\}_{q=1}^N \right) = f_l^{(k)} \left(P, T, \left\{ c_l^{(q)} \right\}_{q=1}^N \right), \quad k = \overline{1,N} \tag{6.17}$$

The fugacities will be described the attraction/repulsion between molecules. The ratio of the fugacity to the pressure is called the fugacity factor $\phi_i^{(k)}$:

$$\phi_i^{(k)} = \frac{f_i^{(k)}}{Px_i^k}, \quad k = \overline{1,N}; \quad i = g,l \tag{6.18}$$

The fugacity factors are calculated using:

$$\ln\left(\phi_i^{(k)}\right) = \frac{b^k}{\bar{b}_i}(z_i - 1) - \ln(z_i - B_i) -$$

$$\frac{A_i}{2\sqrt{2}B_i}\left(\frac{2\sum\limits_{j=1}^{N} x_i^k a^{kj}}{\bar{a}_i} - \frac{b^k}{\bar{b}_i}\right)\ln\left(\frac{z_i + (\sqrt{2}+1)B_i}{z_i - (\sqrt{2}-1)B_i}\right) \tag{6.19}$$

where z_i, A_i, B_i, $\bar{a}_i$, $\bar{b}_i$ and a^{kj} are defined by equations above ($i=g,l$; $k=\overline{1,N}$).

Thus, the fugacity factor of a component depends on the composition and z-factor of the gas mixture (z_i), on the additional functions ($\bar{a}_i$ and $\bar{b}_i$), on the properties of the pure components (T_{cr}^k, P_{cr}^k, and ω^k), and on the cross parameters δ^{kj}.

6.4.3. Thermodynamic closure relations

The closure relations are needed to describes some other parameters of the equation of state. Actually, the fluid densities can be calculated from following relations:

$$\rho_i = \frac{PM_i}{zRT}, \qquad i = g,l \tag{6.20}$$

where M_i is the molar masse for ith phase, $M_i = \sum\limits_{k=1}^{N} x_i^k M^k$.

6.4.4. Open thermodynamic simulator

Conversion of thermodynamic differential equations. The open thermodynamic system has two types of equations: nonlinear and differential equations. We shall convert the system of differential equations into the finite-difference recurrent algebraic equations. Let us replace the differential equations by their finite difference analogues, using the definition

of pressure as $p^i = p^{i-1} + \Delta p$ (for a step i=0,1,...), with $0 < |\Delta p| << 1$. Hence the relations (6.10d) can be presented in the following form:

$$\left(\frac{1}{\Delta c^{(k)}}\right)^{(i)} \left[(c_g^{(k)})^{(i)} - (c_g^{(k)})^{(i-1)}\right] = \left(\frac{1}{\Delta c^{(N)}}\right)^{(i)} \left[(c_g^{(N)})^{(i)} - (c_g^{(N)})^{(i-1)}\right], \quad (6.21)$$

$$k = 1, ..., N - 2$$

Open thermodynamic simulator numerical algorithm. The algorithm of developed simulator consists from the following principal steps:

Step 1. The initial pressure p^0 is imposed as an input parameter.

Step 2. The zero approximation for initial conditions $\left\{c_{l*}^{(q)}\right\}_{q=1}^N$ and $\left\{c_{g*}^{(q)}\right\}_{q=1}^N$ of the recurrent equations (6.21) calculated for the pressure p^0 using the classical closed thermodynamic methods.

Step 3. The new pressure is imposed as $p^1 = p^0 + \Delta p$ for the given pressure step Δp. The system of phase equilibrium equations, phase state equations and the recurrent phase composition equations are nonlinear transcendent system, and can be solved simultaneously using the iterative Newton-Raphson method for the current pressure p^1.

Actually, we shall present the relations (6.10) in the form of linear system relatively the unknown vector $\mathbf{X}$:

$$\mathbf{AX} = \mathbf{B} \qquad (6.22)$$

Here the vector $\mathbf{X}$ represents correction for the nonlinear system relatively the previous solution:

$$c_l^{(i)} = c_{l*}^{(i)} + X_{(i)}, \quad i = 1, ...N - 2, N \qquad (6.23a)$$

$$c_g^{(i)} = c_{g*}^{(i)} + X_{(N-1+i)}, \quad i = 1, ...N - 2, N \qquad (6.23b)$$

The kernel of numerical algorithm consists on the definition of the matrix $\mathbf{A}$ and $\mathbf{B}$ for considered physical problem. The matrix $\mathbf{A}$ composed on two matrixes $\mathbf{A_g}$ and $\mathbf{A_l}$: $\mathbf{A} = \mathbf{A_g} \bigcup \mathbf{A_l}$. The full definition of the matrixes presented in general forms below.

$$\mathbf{A_g} =$$

$$
\mathbf{A_g} =
\begin{array}{c}
1 \\ \\ \dots \\ \\ N \\ N+1 \\ \\ \dots \\ \\ 2N-3 \\ \\ 2N-2
\end{array}
\begin{pmatrix}
\dfrac{\partial \phi_l^{(1)}}{\partial c_l^{(1)}} & \dots & \dfrac{\partial \phi_l^{(1)}}{\partial c_l^{(N-2)}} & \dfrac{\partial \phi_l^{(1)}}{\partial c_l^{(N)}} \\[2mm]
\dots & \dots & \dots & \dots \\[2mm]
\dfrac{\partial \phi_l^{(N)}}{\partial c_l^{(1)}} & \dots & \dfrac{\partial \phi_l^{(N)}}{\partial c_l^{(N-2)}} & \dfrac{\partial \phi_l^{(N)}}{\partial c_l^{(N)}} \\[2mm]
-\dfrac{1}{\left(\Delta c^{(1)}\right)^2}\dfrac{\partial c_g^{(1)}}{\partial p} & \dots & 0 & 0 \\[2mm]
\dots & \dots & \dots & \dots \\[2mm]
0 & \dots & -\dfrac{1}{\left(\Delta c^{(N-2)}\right)^2}\dfrac{\partial c_g^{(N-2)}}{\partial p} & 0 \\[2mm]
0 & \dots & 0 & -\dfrac{1}{\left(\Delta c^{(N)}\right)^2}\dfrac{\partial c_g^{(N)}}{\partial p}
\end{pmatrix}
$$

with top column labels: $1, \dots, N-2, N-1$.

$$\mathbf{A_l} =$$

$$
\mathbf{A_l} =
\begin{array}{c}
1 \\ \\ \dots \\ \\ N \\ N+1 \\ \\ \dots \\ \\ 2N-3 \\ \\ 2N-2
\end{array}
\begin{pmatrix}
-\dfrac{\partial \phi_g^{(1)}}{\partial c_g^{(1)}} & \dots & \dfrac{\partial \phi_g^{(1)}}{\partial c_g^{(N-2)}} & \dfrac{\partial \phi_g^{(1)}}{\partial c_g^{(N)}} \\[2mm]
\dots & \dots & \dots & \dots \\[2mm]
-\dfrac{\partial \phi_g^{(N)}}{\partial c_g^{(1)}} & \dots & \dfrac{\partial \phi_g^{(N)}}{\partial c_g^{(N-2)}} & \dfrac{\partial \phi_g^{(N)}}{\partial c_g^{(N)}} \\[2mm]
\dfrac{1}{\left(\Delta c^{(1)}\right)^2}\dfrac{\partial c_g^{(1)}}{\partial p} & \dots & 0 & 0 \\[2mm]
\dots & \dots & \dots & \dots \\[2mm]
0 & \dots & \dfrac{1}{\left(\Delta c^{(N-2)}\right)^2}\dfrac{\partial c_g^{(N-2)}}{\partial p} & 0 \\[2mm]
0 & \dots & 0 & \dfrac{1}{\left(\Delta c^{(N)}\right)^2}\dfrac{\partial c_g^{(N)}}{\partial p}
\end{pmatrix}
$$

with top column labels: $N, \dots, 2N-3, 2N-2$.

$$\mathbf{B} =$$

$$
\begin{array}{c}
1 \\
\\
... \\
\\
N-1 \\
N \\
\\
N+1 \\
\\
\\
... \\
\\
2N-2
\end{array}
\left(
\begin{array}{c}
-\left(\phi_l^{(1)} + ln(x_l^{(1)}) + \phi_g^{(1)} + ln(x_g^{(1)})\right) \\
\\
... \\
\\
-\left(\phi_l^{(N-2)} + ln(x_l^{(N-2)}) + \phi_g^{(N-2)} + ln(x_g^{(N-2)})\right) \\
-\left(\phi_l^{(N)} + ln(x_l^{(N)}) + \phi_g^{(N)} + ln(x_g^{(N)})\right) \\
\dfrac{1}{\Delta c^{(N)}}\dfrac{\partial c_g^{(N)}}{\partial p} - \dfrac{1}{\Delta c^{(1)}}\dfrac{\partial c_g^{(1)}}{\partial p} \\
\\
... \\
\\
\dfrac{1}{\Delta c^{(N)}}\dfrac{\partial c_g^{(N)}}{\partial p} - \dfrac{1}{\Delta c^{(N-2)}}\dfrac{\partial c_g^{(N-2)}}{\partial p}
\end{array}
\right)
$$

In the current pressure point p_1 the iteration process for the roots vector $\mathbf{X}$ is stopped when the giving precision ζ was attend:

$$
\sum_{i=1}^{N}\left|\frac{\phi_l^{(i)} + ln(x_l^{(i)})}{\left(\phi_g^{(i)} + ln(x_g^{(i)}) - 1\right)^2}\right| < \zeta
\tag{6.24}
$$

On the other hand if the precision ζ was not obtained the linear system (6.22) need to be solved for new value of the approximation $c_{l*}^{(i)} = c_l^{(i)}$ and $c_{g*}^{(i)} = c_g^{(i)}$ (i=1,...,N).

Step 4. The algorithm is repeated for the new pressure $p^2 = p^1 + \Delta p$. The initial approximation for the solution $\left\{c_{l*}^{(q)}\right\}_{q=1}^{N}$ and $\left\{c_{g*}^{(q)}\right\}_{q=1}^{N}$ can be taken from the previous point p_l. The algorithm repeats until the defined bottom limit for pressure is reached.

6.4.5. Thermodynamic system simulation

Based on the described approach the open thermodynamic simulator OTS was developed. The last version of this noncommercial simulator was developed for an unlimited number of chemical components in fluid. Let us show several test cases and compare to the results of the commercial software Eclipse.
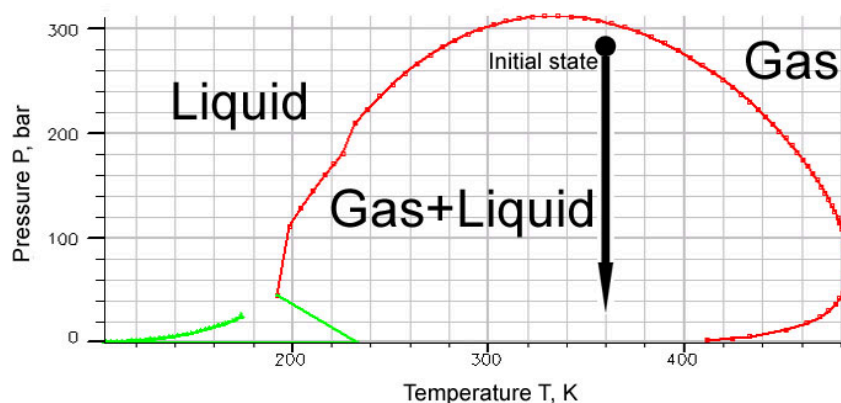
Figure 6.4.: Phase plot for the 8-components fluid

We will consider the multicompositional mixture consisting of 8 chemical components $(CO_2, N_2, CH_4, C_2H_6, C_3H_8, C_4H_{10}, F1, F2)$. We will focus on the domain of coexistence of gas and liquid phases - illustration presented in the Figure 6.4. We will perform the simulations for this 8-components mixture using the OTS simulator and also the compositional simulator Eclipse. However, to obtain the same data using the classical approach (Eclipse), a full compositional system with a hydrodynamic space and time depending part needs to be simulated. The results of Eclipse simulation are based on the model of radial flow and are illustrated after the pressure stabilization in the Figure 6.5 by circles. In terms of mole fractions, the liquid as well as the gas are strongly represented by CH_4, however the liquid phase is also presented by the heavy component $F2$. On the other hand, OTS data construction demands an incomparably small time of calculation, which is represented by lines in the Figure 6.5.

The open thermodynamic model describes correctly the two-phase behaviors for an open system moving through porous media. Even in the case in which the initial conditions are far from real state, OTS comes fast to the solution that describes the open behaviors. An illustration is presented in the Figure 6.6 for a 9-components mixture. Here both phases are also strongly represented by mole fractions of component CH_4.

Let us note, that for some pressure domains the closed thermodynamics can behave similarly to the open system, while for other pressures the behavior can be totally different. We will show an example for a 4-components mixture that was obtained by the simulation of a classical closed thermodynamic system using the commercial software Eclipse PVTi. For the same fluid OTS data was gained. Figure 6.7 illustrates the difference between the two
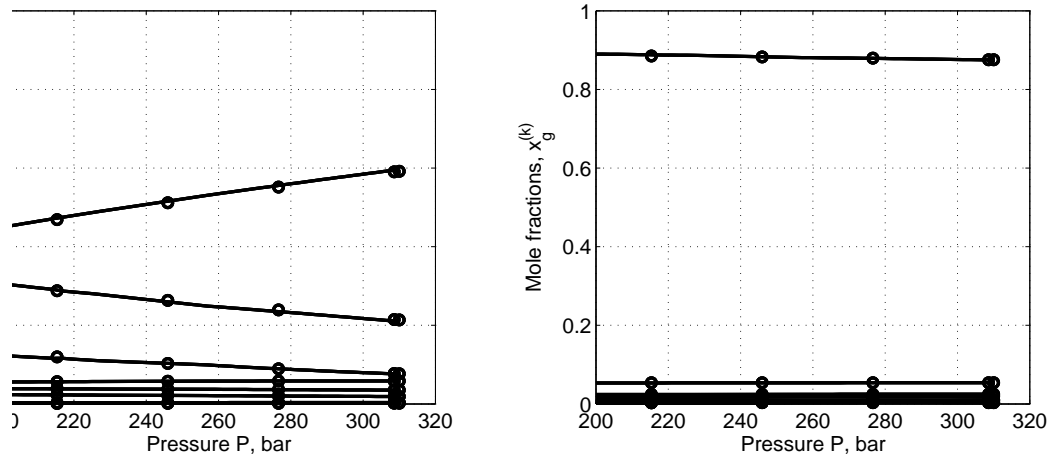
Figure 6.5.: Liquid and gas mole fractions for the 8-components fluid: solid lines - OTS; circles - Eclipse
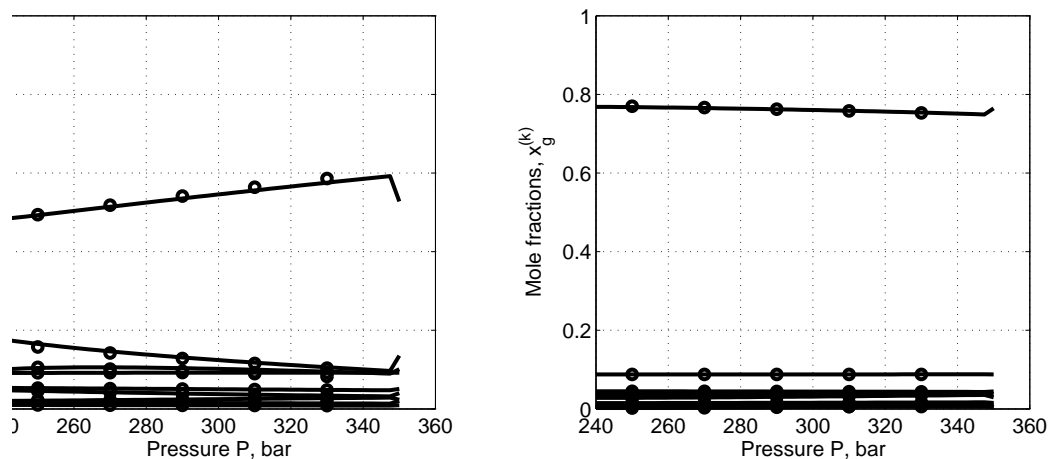
Figure 6.6.: Liquid and gas mole fractions for the 9-components fluid: solid lines - OTS; circles - Eclipse

kinds of simulation - open (solid lines) and closed (dashed lines). Even in the case when the initial approximation for the open simulations are taken from the closed model (Figure 6.7) the result become enormously diverse. In the next section we will concentrate on the analysis of this difference.

One of the evident advantages of the described open thermodynamic simulation is that the system is not calculated in space and time points any more, which can be useful for advanced simulation techniques (e.g. (Thiele et al., 1996), (S. et al., 2008)). Moreover this approach does not limit the number of chemical components in the system.
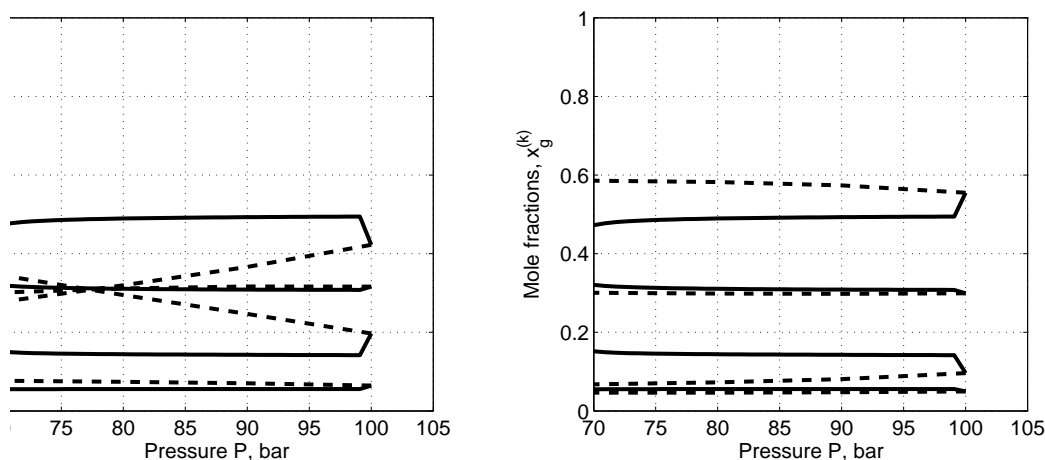
Figure 6.7.: Liquid and gas mole fractions for the 4-components fluid: solid lines - OTS; dashed lines - closed Eclipse PVTi

6.4.6. Behavior of thermodynamic systems

The open thermodynamic model presents the existence of motion through porous media in terms of Delta Law that was obtained from transport equations. Physically this means that the mass of each component is not conserved in some individual volume (Figure 6.1), while for a closed system mass conservation is evident. Actually, the difference between closed and open thermodynamic description is represented by Delta Law for chemical components:

$$\frac{1}{\Delta c^{(k)}}\frac{dc_g^{(k)}}{dp} = \frac{1}{\Delta c^{(N)}}\frac{dc_g^{(N)}}{dp}, \quad k = 1,...,N-2 \tag{6.25}$$

 Delta Law means that the variation of component concentration in gas along the streamline divided by the concentration difference in phases is the same for any component in the examined class of systems. This law may be considered as a necessary condition ensuring a thermodynamic behavior to an open gas-liquid system. Now let us show an example of the behavior for one chemical component by calculating the thermodynamic variables for the 8-components fluid using the classical closed approach (Eclipse PVTi). From the obtained data we can construct the left and right sides of transport Delta Law formula (6.25) for different chemical components. An illustration of this difference is presented in the Figure 6.8 for the component CH_4. This means that behaviors of CH_4 look the same in closed and open systems for the domain of high pressure. However for a large pressure domain the behavior of this component is not the same.

To understand and classify the global behavior in the system by means of the combination
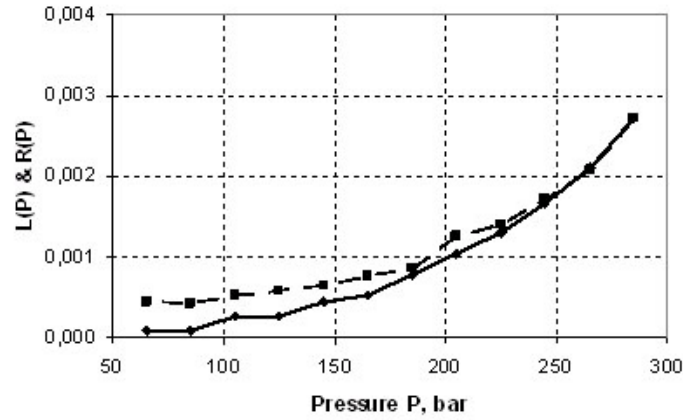
Figure 6.8.: Thermodynamic behaviour: Delta Law for the component CH_4 in the 8-components fluid

of all components let us introduce the following function:

$$V(p) = \frac{1}{N-2} \sum_{k=1}^{N-2} \frac{\dfrac{1}{\Delta c^{(k)}} \dfrac{dc_g^{(k)}}{dp}}{\dfrac{1}{\Delta c^{(N)}} \dfrac{dc_g^{(N)}}{dp}} \qquad (6.26)$$

This function $V(p)$ represents the variation of the closed model from the open one. If $V(p)$ tends to the value 1 the system behaves like the open one. An example is presented in the Figure 6.9 for the 8-components fluid. It is easy to see that the closed thermodynamic system behaves in the described way only for a certain pressure domain. Nevertheless there is a domain where both open and closed systems behave totally different for the same mixture. Actually, we can imagine one individual volume through which gas flows fast while the liquid stays almost immobile. The closed description will be more realistic when the incoming gas in such a volume will be not different from the leaving gas.

6.5. Conclusions

The general case of the compressible compositional two-phase flow in porous media was considered. For this kind of system the mass exchange between phases is one of the important mechanisms, however it is not unique as there is also the transport influence.
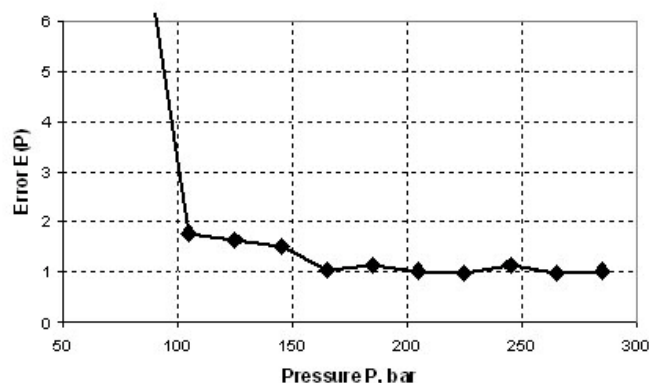
Figure 6.9.: Thermodynamic behaviour: Delta Law for the 8th components fluid

The multicomponent two-phase system moving through porous media was characterized by two essential parameters: the perturbation parameter and the relative mobility parameter. The perturbation parameter represents the stabilization time of the pressure field after perturbation. The relative mobility parameter represents the interior characteristic of the system by the ratio between liquid mobility and gas mobility.

For the contrast phase mobility and fast pressure relaxation process an open thermodynamic model was obtained by HT-splitting. This model includes several differential thermodynamic equations for characterization of the transport phenomena (Delta Law), i.e. describes the equilibrium in an open system.

The new open thermodynamic simulator (OTS) was developed as an alternative to the classical approach. The OTS simulations are validated by the full compositional flow simulations using Eclipse. However, compared to the full compositional simulator an evident advantage of the new OTS simulator is the fact that due to the independence from space and time, data construction demands an extremely short time of calculation. Moreover this new approach does not limit the number of chemical components in the system.

In several cases the closed thermodynamics can behave similarly to the open system. This behavior depends on mass conservation in an individual volume. In order to characterize the difference between the closed and the open thermodynamic systems we proposed to use Delta Law for one chemical component and a variation function for the combination of all components. The closed description can be realistic if the incoming gas in an individual volume is not different from the leaving gas; otherwise the use of the open approach (new or classical) is suggested.

7. Effective streamline-based modeling of compositional flow

Bibliographic Note: The content of this chapter is based on the following original article: Oladyshkin S., Royer J.-J., Panfilov M. Effective solution through the streamline technique and HT-splitting for the 3D dynamic analysis of the compositional flows in oil reservoirs. Transport in Porous Media, Springer, V. 74, N. 3, P. 311-329, 2008.

Streamline approach are often used as an alternative effective method to classical finite difference technique for solving large heterogeneous fluid flow models in petroleum reservoirs. In the case of complex multi-component fluid system, this approach is scarcely used because the hydrodynamic and thermodynamic flow equations are strongly coupled through non independent variables including the pressure, the saturation and the species concentrations. It has been shown recently (Oladyshkin and Panfilov, 2007a), (Oladyshkin and Panfilov, 2007b) that assuming quasi steady state for the pressure field, the hydrodynamic and thermodynamic parts can be split into a set of equations that is referred as HT-split compositional model. In this work, the HT-split model is combined with the streamline technique. This approach has been implemented in gOcad using the StreamLab plugin. The pressure field and the streamlines are computed using the finite volume flow simulator. The equations that govern the equilibrium between phases are solved separately using a classical non-linear solver. A multi-component 1D solver has been implemented using the HT-split equations along the streamlines. Tools for visualizing the time evolution of species compositions have been also developed. Finally a simple case study illustrating the technique is presented.

It is shown that the HT-splitting method coupled to the streamline technology provides an effective tool to solve complex problems involving multi-compositional flow for any 3D reservoir geometry and for any gas-liquid system. The advantage of such a technology is that the number of components is not limited.

7.1. Introduction

One of the most challenging problems in petroleum reservoir engineering is the modelling of flows in complex reservoirs accounting for geology and petro physical properties. This modelling should take into account the simulation of the permeability field that matches history dynamic production data in order to predict pressure and saturation changes in the reservoir during exploitation. In this context, streamlines are often used to find an approximation of the flow and saturation equations in order to accelerate dynamic simulations.

The main idea underlying the streamline-based concept is to reduce a multidimensional problem to a set of $1D$ flow problems along streamlines. Streamline-based flow simulation is often presented as an alternative method to the classical traditional flow modelling approaches based on finite differences or finite element techniques. Moreover, the streamline simulation approach is an effective technology to simulate fluid flows in large heterogeneous geological models of petroleum reservoirs. Extended works have been done on black oil models in heterogeneous systems (Thiele et al., 1996) and on $3D$ Compositional Reservoir Simulation (Thiele et al., 1997). The streamline approach is often based on the assumption that pressure is stationary at least during short period of time. This assumption may not be verified during a long-term exploitation period when the time variation of pressure becomes significant. However, in some cases as in gas-condensate reservoirs (Oladyshkin and Panfilov, 2005), (Oladyshkin and Panfilov, 2007b) pressure may be considered as stationary with a good approximation, on the contrary to saturation. Hence, this work suggests applying streamline technology to gas-condensate reservoirs.

Streamline-based methods have been successfully applied to multi-phase flow in porous media, in particular in hydrogeology and in reservoir simulation, (Batycky, 1997) and (Batycky et al., 1996), while treating gravity but neglecting capillarity. Water-oil flow problems were treated in (Wang and Kovscek, 2000), (Thiele and Batycky, 2001), (Yeh, 1986) where efficient solutions to the pressure equation were suggested. In these cases the streamline methods use the fractional flow formulation when the model represents an elliptic/parabolic pressure equation and a hyperbolic saturation transport equation. The sensitivity analysis shows that the fractional flow rate at the production wells is directly related to the time-of-flight along each individual streamline and to the pressure at the grid cells (Fetel, 2007).

The basis of any streamline method is the simulation of the pressure field. A typical technique is presented in paper (Aarnes, 2004) which uses a multi scale mixed finite-element method to compute the pressure on a coarse grid, while the velocity field is estimated on an

underlying finer grid using base functions fitted on the coarse grid. This numerical methodology together with streamlines gives an efficient and robust method to solve detailed flow patterns on the underlying fine grid.

A significant element of any streamline technique is the introduction of the time-of-flight, used instead of the longitudinal curvilinear coordinate along each streamline. The parameterization of streamlines with the time-of-flight was suggested in (Thiele and Batycky, 2001), (Wang and Kovscek, 2000), (Lialin and Silniov, 2005), (Batycky, 1997), (Batycky et al., 1996). In particular Pollock (Pollock, 1988) suggested a simple analytical formula to calculate the streamline trajectory crossing a $3D$ cell. Assuming a piece-wise linear approximation of the velocity field in the cell, the Pollock algorithm determines the exit point and the time of flight for a specific streamline from any entry point.

The streamline technology gives an alternative choice to the traditional approach for multi phase flow problems in heterogeneous reservoirs and rises attention to the special case of compositional flow (Christie and Clifford, 1997). In several field-scale compositional studies, serious numerical difficulties occur in conventional methods. A common solution often used is to reduce the number of cells and/or to simplify the geological complexity which affects directly the numerical performances. In opposition, the streamline approach remains unrivaled in the ability to efficiently model the transport of species along the flow paths, even in the presence of extreme permeability/porosity values. In the very specific case of multi-component and multi-phase flow characterized by an extreme degree of complexity, the reduction of a three-dimensional problem to the series of one-dimensional solutions is so attractive in controlling the numerical dispersion that it cannot be overlooked, as shown in (Thiele and Edwards, 2001).

For large fields exploited by gas injection, streamlines unlike traditional technologies offer some advantages to model full-field scenarios with a reasonable inter-well spatial resolution in acceptable run times on affordable computers. In the present chapter we examine one particular case of a compositional flow that corresponds to gas-condensate systems which contain N chemical components like hydrocarbons from CH_4 to $C_{10}H_{22}$ and even heavier ones. The flow of such a mixture is described by the *full compositional model* that consists in a coupled system involving the momentum balance equations for each phase, the mass balance equations for each chemical component present in the mixture, and the thermodynamic closure relations that describe the phase equilibrium and the phase state. Usually such a model represents a high order nonlinear transcendent equation system for which a constructive analysis is very difficult, so that the only method to analyze this system is the

numerical simulation (Aziz and Settari, 1979), (Coats, 1980). Some simplified versions of
the compositional model may be obtained by reducing the number of components up to two
(Koldoba and Koldoba, 2003), or by assuming a flow stationarity (Whitson et al., 2003),
(Dinariev, 1996). An attempt to provide a justification to such a stationary theory was done
in (Chopra and D., 1986) by neglecting dispersion, capillarity and gravity.

Another, much less coarse approach was developed in (Oladyshkin and Panfilov, 2007b)
where the thermodynamic block was totally split from the hydrodynamic sub-system along
streamlines, whatever the number of components. Such a decoupling procedure was named
the *HT-splitting*.

In the present work, the HT-model is combined with the streamline technique. This approach
has been implemented in gOcad using the StreamLab plugin. The pressure field and the
streamlines are computed using the StreamLab finite volume flow simulator. The equations
that govern the equilibrium between phases are solved separately using a classical non-linear
solver. A multi-component 1D solver has been implemented as a StreamLab plug-in to solve
the HT-split equations along the streamlines. Tools for visualizing the time evolution of
species compositions have been linked to the StreamLab plug-in. Finally, a simple case
study illustrates the technique presented here.

7.2. Compositional flow model

This work considers a mixture consisting of N chemical components able to form two ther-
modynamic phases (namely, gas and liquid) separated from one to another by an interface.
Each component may be dissolved in both phases at a variable solvability degree under var-
ious thermodynamic conditions.

7.2.1. Full compositional model

According to the phenomenological approach (Nigmatulin, 1987), (Sedov, 1976), the gas,
liquid and solid are considered as three interpenetrating continuums. This implies a fine
dispersed system which justifies a local equilibrium assumption. We will consider isothermal
process in the reservoir. The general compositional model involves N mass balance equations
for each chemical component and two momentum balance equations for each phase, (Coats,

1980), (Aziz and Settari, 1979) :

$$\phi \frac{\partial}{\partial t}\left(\rho_l c_l^{(k)} S_l + \rho_g c_g^{(k)}\left[1-S_l\right]\right) + \mathrm{div}\left(\rho_l c_l^{(k)} \mathbf{v}_l + \rho_g c_g^{(k)} \mathbf{v}_g\right) = 0, \quad k = 1,...,N \quad (7.1a)$$

$$\mathbf{v}_l = -\frac{k_l \mathbf{K}}{\mu_l}\,\mathrm{grad}(P_l - \rho_l gz), \quad \mathbf{v}_g = -\frac{k_g \mathbf{K}}{\mu_g}\,\mathrm{grad}(P_g - \rho_g gz) \quad (7.1b)$$

where superscript k refers to the kth chemical component ($k = 1, \cdots, N$), and indexes g and l to gas and liquid, respectively; ϕ is the porosity; ρ is the phase density; $c_g^{(k)}$ or $c_l^{(k)}$ are the mass concentrations of the kth component in gas or liquid; S_l is the volume saturation of pores by liquid; k_g and k_l are the relative permeabilities; μ is the phase dynamic viscosity; $\mathbf{K}$ is the tensor of the absolute permeability; P is the phase pressure, $\mathbf{v}$ is the Darcy velocity. The components are assumed to be sorted in ascending order of molecular weights, such that component 1 is the lightest, whilst component N is the heaviest.

The momentum balance equations (7.1b) are written in the Darcy form for each phase. The phase dynamic viscosity μ is a given function of the pressure P and of the phase composition $c_i^{(k)}$. The absolute permeability K and porosity ϕ are given as functions of the space coordinates and of the pressure P. The system of $N+2$ equations (7.1) is written with respect to the pressure P, the liquid saturation S_l, the phase velocities $\mathbf{v_l}$ and $\mathbf{v_g}$, and the $N-2$ independent concentrations which will be denoted as $\chi^{(k)}$, $k = 1,...,N-2$ and called the *concentration basis* according to (Oladyshkin and Panfilov, 2007b).

Other concentrations and phase densities are determined from the thermodynamic closure relationships:

$$v_g^{(k)}\left(P,\left\{c_g^{(q)}\right\}_{q=1}^N\right) = v_l^{(k)}\left(P,\left\{c_l^{(q)}\right\}_{q=1}^N\right), \quad k = 1,...,N \quad (7.2a)$$

$$\rho_g = \rho_g\left(P,\left\{c_g^{(q)}\right\}_{q=1}^N\right), \quad \rho_l = \rho_l\left(P,\left\{c_l^{(k)}\right\}_{q=1}^N\right) \quad (7.2b)$$

$$\sum_{k=1}^N c_g^{(k)} = 1, \quad \sum_{k=1}^N c_l^{(k)} = 1 \quad (7.2c)$$

where $v_i^{(k)}$ is the chemical potential of the kth component in the ith phase which is a given function.

Equation (7.2a) describe the local phase equilibrium (equivalence of the chemical potentials for each component in both phases), equation (7.2b) are the equations of phase state formulated in a general form (various versions of these equations may be found in (Danesh, 1998)), Equation (7.2c) normalize the concentrations.

For sake of simplicity, the capillary pressure P_c and the gravity head $\rho_l gh$ with h being the characteristic reservoir height, are neglected with respect to the characteristic pressure difference ΔP over the studied domain. Under these assumptions, gas and liquid pressures are identical: $P_l = P_g = P$. Thus, the liquid and gas velocities take the form :

$$\mathbf{v}_l = -\frac{k_l\,\mathbf{K}}{\mu_l}\,\mathrm{grad}P, \quad \mathbf{v}_g = -\frac{k_g\,\mathbf{K}}{\mu_g}\,\mathrm{grad}P \tag{7.3}$$

It is classical to define the *total velocity* $\mathbf{v}_t$ as the sum of liquid and gas velocities $\mathbf{v}_t = \mathbf{v}_g + \mathbf{v}_l$, and to introduce the *fractional flows* f_l and f_g for liquid and gas, respectively. The fractional flow $f_j = \iota_j/\iota_t$ of phase j is defined as the ratio of the phase mobility $\iota_j = k_j/\mu_j$ to the total mobility $\iota_t = \iota_l + \iota_g$, such that $\mathbf{v}_j = f_j\,\mathbf{v}_t$. Then, the full compositional model can be simplified into :

$$\phi\frac{\partial}{\partial t}\left(\rho_l c_l^{(k)}S_l + \rho_g c_g^{(k)}\left[1 - S_l\right]\right) + \mathrm{div}\left(\left[\rho_l c_l^{(k)}f_l + \rho_g c_g^{(k)}f_g\right]\mathbf{v}_t\right) = 0, \quad k = 1,...,N \tag{7.4}$$

where $\mathbf{v_t}$ is the total velocity defined by :

$$\mathbf{v_t} = -\iota_\mathbf{t}\,\mathbf{K}\,\mathrm{grad}\mathbf{P} \tag{7.5}$$

7.2.2. HT-split mompositional model

In (Oladyshkin and Panfilov, 2007b), (Oladyshkin and Panfilov, 2005) it has been shown that for gas-liquid systems the compositional model can be written in a HT-split form in which the thermodynamic part is totally separated from the hydrodynamic one. This is possible to applied the same procedure along each streamline if the following properties of the system are verified :

- the contrast between gas and liquid mobilities ($\iota_g \ll \iota_l$) is very high in magnitude, liquid mobility is much lower than that of gas;

- the pressure field is quasi-stationary ($\partial P/\partial t \approx 0$), i.e. the reservoir is studied for times greater than the characteristic propagation time of the pressure perturbation;

Within this framework, the global system of the N mass balance equations can be reduced to two hydrodynamic equations which characterize the saturation S_l and pressure P evolutions of the two-phase multi component fluid system, which is called the split hydrodynamic

model (Oladyshkin and Panfilov, 2007b):

$$\text{div}\left(\rho_g \mathbf{v}_g\right) = 0 \tag{7.6a}$$

$$\phi \rho_l \frac{\partial S_l}{\partial t} + \phi \left\{ \rho_l \lambda_l S_l + \rho_g \lambda_g (1 - S_l) \right\} \frac{\partial P}{\partial t} = -\left[\text{div}\left(\rho_l \mathbf{v}_l\right) + \rho_g \lambda_g \mathbf{v}_g \cdot \text{grad} P \right] \tag{7.6b}$$

where

$$\lambda_l = \frac{1}{\Delta c^{(N)}} \frac{dc_l^{(N)}}{dP}, \quad \lambda_g = \frac{1}{\Delta c^{(N)}} \frac{dc_g^{(N)}}{dP}, \quad \Delta c^{(N)} = c_l^{(N)} - c_g^{(N)}$$

All the other thermodynamic variables (mass densities and phase concentrations) are determined through the split thermodynamic model that consists of equations (7.2) and the new differential thermodynamic equations:

$$\sum_{q=1}^{N-2} \left(\frac{1}{\Delta c^{(q)}} \delta_{qk} - \frac{1}{\Delta c^{(N)}} \frac{\partial c_g^{(N)}}{\partial \chi^{(q)}} \right) \frac{d\chi^{(q)}}{dP} = \frac{1}{\Delta c^{(N)}} \frac{\partial c_g^{(N)}}{\partial P}, \quad k = 1, ..., N-2 \tag{7.7}$$

They are obtained from the original compositional system (7.1) assuming the splitting hypothesis, see (Oladyshkin and Panfilov, 2007b). These last differential thermodynamic equations describe the variation of the total composition in an open system, see (Oladyshkin and Panfilov, 2007a). The boundary-value conditions impose the pressure value at the starting point of each streamline and at the well-bore (or at the reservoir boundary depending on the flow direction).

System (7.2), (7.7) is closed if the pressure is known. In other words, this system determines all the concentrations and phase densities as functions of the pressure. Due to this, the coefficients of the hydrodynamic model (7.6b) depend on pressure only.

Note that the thermodynamic model (7.7) is valid only along the streamlines for any 3D geometry and for any gas-liquid system.

The advantage of the HT-split model is that the number of components is not limited.

7.2.3. HT-split model expressed through the total velocity

Equation (7.6) can be rewritten using the total velocity $\mathbf{v}_t = \mathbf{v}_l + \mathbf{v}_g$ introduced in §7.2.1 in a more compact form which can be easily used then to be applied in to the streamline approach. The liquid and gas velocities expressed through the total velocity are: $\mathbf{v}_l = f_l \mathbf{v}_t$ and

$\mathbf{v}_g = f_g \mathbf{v}_t$, where f_l and f_g are the fractional flows for liquid and gas, respectively. Reporting these expressions in equation(7.6), it comes :

$$\mathrm{div}\,(\rho_g f_g \mathbf{v_t}) = 0 \tag{7.8a}$$

$$\phi \rho_l \left[\frac{\partial S_l}{\partial t} + \lambda \frac{\partial P}{\partial t} \right] = -\rho_g f_g \left(\mathrm{grad}\,\pi_g^l + \lambda_g \mathrm{grad}P \right) \cdot \mathbf{v}_t \tag{7.8b}$$

where

$$\lambda \equiv \lambda(P, S_l) \equiv \lambda_l(P)S_l + \frac{\rho_g(P)}{\rho_l(P)} \lambda_g(P)(1 - S_l) \tag{7.9}$$

$$\pi_g^l = \pi_g^l(P, S_l) = \frac{\rho_l f_l}{\rho_g f_g} = \frac{\rho_l \mu_g k_l(S_l)}{\rho_g \mu_l k_g(S_l)} \tag{7.10}$$

π_g^l is the *fractional mass flow ratio*.

Indeed, using Eq. (7.8a), the first term in the right-hand side of Eq. (7.6b) may be transformed in the following way:

$$\mathrm{div}\,(\rho_l \mathbf{v}_l) = \mathrm{div}\,(\rho_l f_l \mathbf{v}_t) = \mathrm{div} \left(\rho_g f_g \mathbf{v}_t \, \frac{\rho_l f_l}{\rho_g f_g} \right) =$$

$$= \rho_g f_g \mathbf{v}_t \cdot \mathrm{grad} \left(\frac{\rho_l f_l}{\rho_g f_g} \right) = \rho_g f_g \mathbf{v}_t \cdot \mathrm{grad} \left(\frac{\rho_l \mu_g k_l}{\rho_g \mu_l k_g} \right)$$

The second term in the right-hand side of Eq. (7.6b) can be also transformed by replacing the pressure gradient by the flow velocity from Darcy's law, but this operation appears to be useless for the streamline technique.

7.3. Compositional flow problem solved along streamlines

7.3.1. Framework and assumptions

The underlying idea of the streamline simulation method is to decouple the governing flow equations from the full $3D$ problem into a set a multiple $1D$ transport problems solved independently along the streamlines. The streamlines are simulated using the $3D$ equation for pressure (7.8a) in which the saturation field is assumed to be known. After this, the saturation

field is calculated using (7.8b) as a $1D$ equation along each streamline. The advantage of the HT-split method consists in that the concentrations transport equations are not calculated at all. Whatever the number of components, the phase concentrations are calculated from the thermodynamic split model.

The general procedure is to :

- Simulate the mono-variant thermodynamic system for pressures ranging in a realistic interval defined by the exploitation conditions of the reservoir under study;

- Compute the pressure and the velocity fields, separately, by assuming the saturation field be known;

- Track the streamlines defined as lines tangent to the flow velocity vector;

- Solve the saturation equations in $1D$;

Then, once both the boundary conditions and the saturation field within the reservoir have changed significantly, the pressure and the streamlines are updated.

Such a streamline technique is based on the following assumptions :

- the *pressure field is steady-state* or *quasi steady-state*;

- the *total velocity depends weakly* on the saturation (and concentration) at least during sufficient periods of time;

- the *capillary pressure is neglecting*;

- the *permeability is isotropic*, so the pressure gradient and the flow velocity are collinear;

The neglected capillary pressure implies that the velocities of both pases are collinear being proportional to the same pressure gradient. This means that the system of streamlines is common for gas and liquid (even if the absolute values of flow velocities are different).

The assumption concerning the medium is isotropy means that the flow velocities for both phases are collinear to the pressure gradient. In an anisotropic medium, the components of the pressure gradient orthogonal to the streamline are non zero, due to that, implementation of the streamline technology becomes much more complicated.

Within the framework of the above mentioned assumptions, the saturation equation can then be rewritten in a $1D$ form along streamlines as it was done in (Oladyshkin and Panfilov,

2005). This method provides an extremely fast solution to the compositional system. In some very specific cases, these one-dimensional equations can be even solved analytically (Panfilova and Panfilov, 2004).

7.3.2. Saturation equation along a streamline

In the following, it is assumed that the set of streamlines is *compact*, i.e. any point of the studied domain is crossed by one and only one streamline, and *uniquely defined*, i.e. the streamlines do not cross over one each other. So, at any location in the studied domain, it is always possible to find a *unique* streamline passing through that point.

Let $\mathscr{S}$ be an arbitrary streamline and Y an orthogonal curvilinear coordinate system $Y = (y^1, y^2, y^3)$ with unit vectors $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ in such a way that the vector $\mathbf{e}_1$ is tangent to the streamline $\mathscr{S}$ at each point, then :

$$\mathbf{e}_1 = \mathbf{v}_t / \|\mathbf{v}_t\|$$

where $\mathbf{v}_t$ is the total Darcy velocity. The two other coordinate vectors, $\mathbf{e_2}$ and $\mathbf{e_3}$, are located on the curvilinear iso-pressure surface orthogonal to all the streamlines, as shown in Figure 7.1.

The Riemann metrics in the introduced curvilinear space is defined by the metric tensor $\mathbf{g} = [g_{ij}]$ whose components are defined as the scalar products of the base vectors $g_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j$. It determines the metric properties of the space such as length, angles, areas and volumes in the vicinity of any space point. In an orthogonal frame, only the diagonal elements of this tensor are non nil.

Figure 7.1.: Curvilinear frame (y^1, y^2, y^3) associated to the streamlines

To reformulate the saturation equation (7.8b) along a streamline, we need to use the general relationship for the gradient operator of any arbitrary function Φ in an orthogonal curvilinear coordinate system, (Korn and Korn, 1968):

$$\mathrm{grad}\Phi = \frac{1}{\sqrt{g_{11}}} \frac{\partial \Phi}{\partial y^1} \mathbf{e}_1 + \frac{1}{\sqrt{g_{22}}} \frac{\partial \Phi}{\partial y^2} \mathbf{e}_2 + \frac{1}{\sqrt{g_{33}}} \frac{\partial \Phi}{\partial y^3} \mathbf{e}_3$$

As seen, in the frame Y associated to the streamline, the total velocity vector becomes: $\mathbf{v}_t = v_t \mathbf{e}_1$. Due to this, the scalar product of the total velocity vector and the gradient of any function Φ becomes mono-dimensional in the frame Y:

$$\mathbf{v}_t \cdot \operatorname{grad}\Phi = v_t \, \mathbf{e}_1 \cdot \operatorname{grad}\Phi = \frac{v_t}{\sqrt{g_{11}}} \frac{\partial \Phi}{\partial y^1}$$

Instead of y^1, let introduce l the curvilinear abscise (length) measured along the current streamline. It is easy to show that (Korn and Korn, 1968):

$$dl = \sqrt{g_{11}} dy^1$$

Due to this, the saturation equation (7.8b) formulated along a streamline becomes:

$$\phi \rho_l \left[\frac{\partial S_l}{\partial t} + \lambda \frac{\partial P}{\partial t} \right] = -\rho_g f_g v_t \left(\frac{\partial \pi_g^l}{\partial l} + \lambda_g \frac{\partial P}{\partial l} \right) \tag{7.11}$$

where v_t is the magnitude of the total flow velocity: $v_t \equiv \|\mathbf{v}_t\|$.

7.3.3. Saturation equation in terms of time-of-flight

The *time of flight* τ (TOF) is the time required by a transported particle to reach a given point located at a known distance l along a streamline (Lialin and Silniov, 2005). The TOF is defined by :

$$d\tau(l) = \frac{\phi(l)}{v_t(l)} dl \tag{7.12}$$

where ϕ is the medium porosity, so that v_t/ϕ is the true flow velocity.

Then,

$$v_t \frac{\partial}{\partial l} = \phi \frac{\partial}{\partial \tau}$$

Using the time of flight equation (7.12) instead of the curvilinear abscise l, the saturation equation Eq.(7.11) takes the following simplified fundamental form :

$$\frac{\partial S_l}{\partial t} + \lambda \frac{\partial P}{\partial t} = -\frac{\rho_g f_g}{\rho_l} \left(\frac{\partial \pi_g^l}{\partial \tau} + \lambda_g \frac{\partial P}{\partial \tau} \right) \tag{7.13}$$

Noting that the *fraction mass flow ratio* is a function of the saturation and of the pressure, it comes

$$\frac{\partial \pi_g^l}{\partial \tau} = \frac{\partial \pi_g^l}{\partial S_l} \frac{\partial S_l}{\partial \tau} + \frac{\partial \pi_g^l}{\partial P} \frac{\partial P}{\partial \tau}$$

Then, within the streamline approach, in the last equation 7.13, the pressure field is known *a priori*, so the saturation equation can be reduced to the following non-linear convective transport equation with a source term caused by phase transitions:

$$\frac{\partial S_l}{\partial t} + a(S_l, \tau, t)\, \frac{\partial S_l}{\partial \tau} = b(S_l, \tau, t) + c(S_l, \tau, t) \tag{7.14}$$

where

$$a(S_l, \tau, t) \equiv \frac{\mu_g}{\mu_l} f_g(S_l) \frac{d}{dS_l}\left(\frac{k_l}{k_g}\right),$$

$$b(S_l, \tau, t) \equiv -\frac{\rho_g f_g}{\rho_l}\left[\lambda_g + \left(\frac{k_l}{k_g}\right)\frac{d}{dP}\left(\frac{\rho_l \mu_g}{\rho_g \mu_l}\right)\right]\frac{\partial P}{\partial \tau},$$

$$c(S_l, \tau, t) \equiv -\lambda \frac{\partial P}{\partial t}$$

The dependence of all the coefficients on τ and t is due to the presence in their structure of the pressure P that varies in space and time.

In a particular case when the pressure field is totally steady-state and the phase viscosities are constant, the coefficient c is zero and we obtain the following equation:

$$\frac{\partial S_l}{\partial t} + a(S_l)\, \frac{\partial S_l}{\partial \tau} = b(S_l, \tau) \tag{7.15}$$

Equation for saturation should be completed with the initial condition:

$$S_l(\tau, t = 0) = S_{l_0}(\tau)$$

where S_l^0 is a known function.

The structure of Eq. (7.15) reveals two basic mechanisms which determine the saturation variation in time: the *liquid flow* which is described by the convective term, and the *phase transition* that corresponds to the source-term b.

The convective term is close to the respective term in the Buckley-Leverett equation which describes the flow of two immiscible liquids without phase transitions. The source-term, b, is responsible for the mass exchanges between gas and liquid and determines the continuous evolution of the liquid saturation in time at each space point. When the relative liquid mobility is low then the phase transition term is dominant, which will lead to a continuous non-stationary process of liquid accumulation or evaporation in the overall space.

Unfortunately, the term b in Eq. (7.15) contains explicitly the time of flight, which makes it impossible to develop an analytical approach in the general case. In particular for cases of metrics which correspond to a radial or a plan-parallel flow, the analytical methods for solving this equation, based on the singular perturbation technique, were developed in (Panfilova and Panfilov, 2004).

7.3.4. Simulating the phase composition

The compositions of the liquid and gas phases result from the thermodynamic equilibrium described by the thermodynamic split equations (7.2), (7.7).

As mentioned, Eq. (7.7) takes into account the fact that any medium volume is an open system due to the difference between the gas and liquid velocities. Indeed, any mobile volume element strictly related with the individual particles of gas is open with respect to liquid and vice versa. The open character of the system does not influence its phase behavior if the concentrations are uniform throughout all this volume: indeed, the entering fluid has the same composition as the fluid leaving the volume, so the composition in the volume is invariable. In contrast, if the concentration gradients are non zero, then the open character of the system influence its phase behavior. By this way, Eq. (7.7) contains the concentration gradients in the pressure space.

The practical simulation of mixed systems of algebraic and differential equations (7.2), (7.7) is not simple and requires a numerical solution for the differential equations, which renders the overall thermodynamic algorithm more complicated than in the classical cases. It is then natural to analyze whether this split thermodynamic model can be replaced by the classical thermodynamic model.

Let remind that the full compositional model (7.1), (7.2) is usually not calculated directly in this form, but is also split: in this case the thermodynamic part is completed by additional relations describing the mass balance in a closed system, after which one solves the hydrodynamic part (7.1). These additional thermodynamic relations are:

$$\eta^{(k)} = \frac{\rho_g c_g^{(k)}(1-\sigma) + \rho_l c_l^{(k)}\sigma}{\rho_g(1-\sigma) + \rho_l\sigma}, \qquad k = 1,2,...,N-1 \tag{7.16}$$

where $\eta^{(k)}$ is the total composition of the fluid which is assumed to be known, σ is the liquid saturation which is reached in a quasi-static process.

System (7.1), (7.16) of $2N+3$ equations contains $2N+2$ variables: $\{c_g^{(k)}\}_{k=1}^N$, $\{c_l^{(k)}\}_{k=1}^N$, P and σ. So this system also determines the phase concentrations as the functions of pressure. This system is classic and is implemented into a number of commercial simulators.

We have performed a comparative simulation of the phase concentration variation as the function of pressure for a standard gas-condensate mixture described in section 7.4.1, by using two approaches: i) the split thermodynamic model (7.1), (7.2) for an open system, and ii) the classic thermodynamic model (7.1), (7.16) for the closed system. The results

of the numerical simulations are shown in Figure 7.2, and presented as a mean relative error between model (i) and (ii) in term of variation of the total composition (VTC), equation (7.7). In other words, the Figure 7.2 shown the order of non-conservation of the full derivative (7.7) in classical thermodynamics relatively open one with respect to pressure along the streamlines, (Oladyshkin and Panfilov, 2007a).
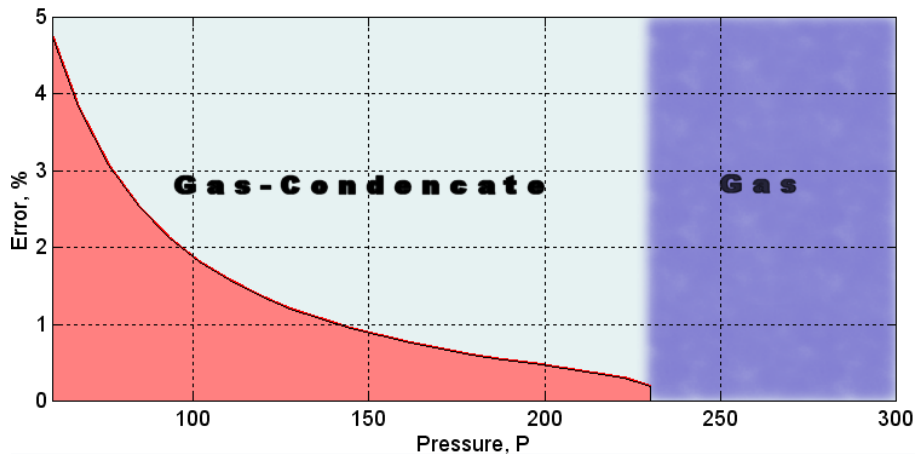
Figure 7.2.: Difference between the simulation results obtained by the split thermodynamic model and the classic thermodynamic model

This means that the open character of any volume is a negligible fact for the thermodynamic behavior of gas-condensate systems.

Due to this result, the classic thermodynamic model will be used in the streamline simulations described below.

7.4. Streamline simulation of a compositional flow

7.4.1. 3D reservoir and fluid

In this section, a simple case study will illustrate the Streamline Compositional Flow Simulation. We examine the exploitation of a $3D$ gas-condensate heterogeneous reservoir by three production wells at natural drive (see the Figure 7.3). The reservoir temperature was constant and equal to $T = 350K$.

At the initial state the reservoir was in single-phase state. The reservoir depletion leads to a pressure decrease in the overall reservoir, which leads in turn to the appearance of the

retrograde liquid and the continuous growth of liquid saturation. The saturation behavior in a gas-condensate system significantly depends on the current composition of each phase and the character of mass exchanges between gas and liquid.

As an example, let consider the mixture constituted of 9 chemical components at the following initial mass concentrations:

$$N_2 = 0.047, C_1 = 0.791, C_2 = 0.0672, C_3 = 0.036, C_{4N} = 0.0149,$$

$$C_5 = 0.0196, C_6 = 0.0162, C_{10} = 0.0071, CO_2 = 0.001$$

As explained in the previous section 7.3.4, the thermodynamics of the mixture was simulated using the classical approach adapted to a closed system. As the result, all the thermodynamic variables were determined as functions of pressure for a realistic range of pressure variation. This enables to calculate all the coefficients of the saturation transport equation (7.14) which depend on pressure.

7.4.2. Simulation of the pressure and velocity fields

The pressure and the velocity fields were computed by solving the $3D$ Eq. (7.6a). The example of the total velocity field in a reservoir is shown in Figure 7.3, together with the pressure around the production wells (Figure 7.4).

7.4.3. Streamline tracking

From the obtained $3D$ pressure field, the streamlines can be built in $3D$ for a fixed number of lines using the technique described in (Fetel, 2007). Figure 7.4 illustrates an example of the pressure field around the production wells with 50 streamlines.

The streamlines are defined as tangent lines to the total velocity vector and are parameterized with the time-of-fight. A Pollock analytical approach is used to determine the trajectory while crossing a $3D$ cell. Assuming a piece-wise linear approximation of the velocity field in the cell, the Pollock algorithm determines the exit point and the time of flight for a specific streamline given at any entry point.

The results of simulations are shown in Fig. 7.4.
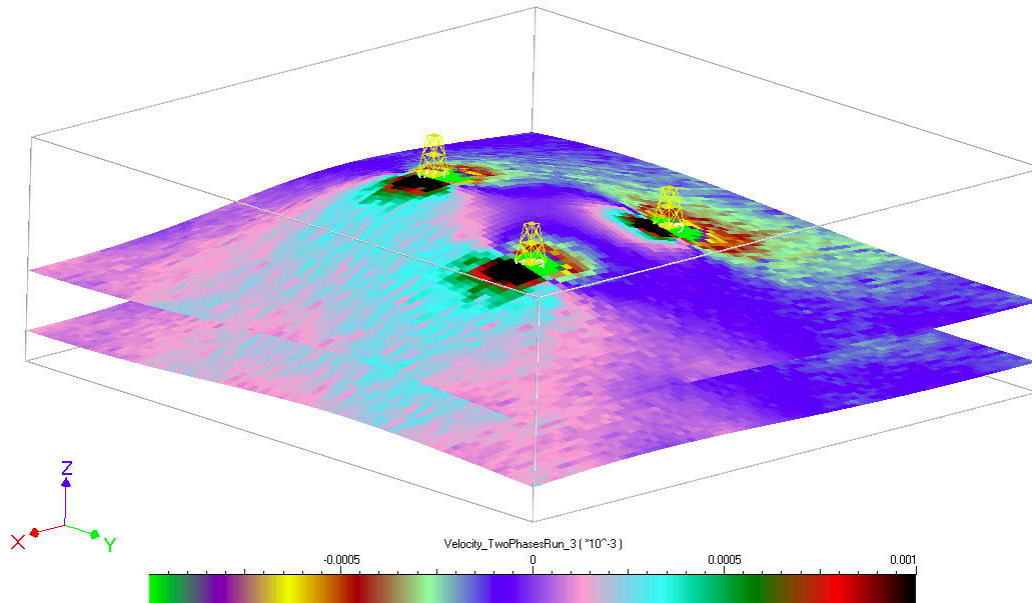
Figure 7.3.: Velocity distribution in the reservoir
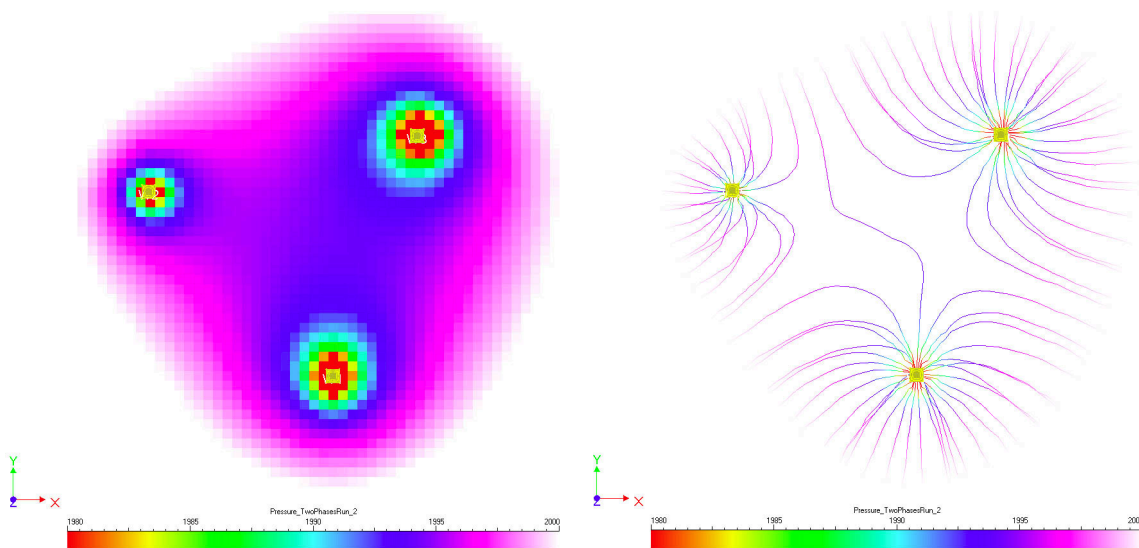
Figure 7.4.: Pressure fields (left) and Streamlines (right) in the vicinity of wells

7.4.4. Computing the saturation along streamlines

After computing the pressure and the velocity fields, the derivative of pressure with respect to the time-of-flight τ was simulated along each streamline, in order to compute parameter b in the saturation equation (7.15).

The solution to (7.15) was obtained by applying the discretization scheme that was explicit in time and backward in space. The nonlinearity if the convective term a was treated by iterations, without linearizing. In the examined problem of the natural reservoir depletion, the saturation and concentration behavior is continuous without arising of shocks, as shown in (Panfilova and Panfilov, 2004), so there was no needed to apply some special procedure of grid refinement or a priori shock detection. A typical variation of the saturation variation along a streamline is presented in Figure 7.5b.

The numerical grid along each streamline was refined in the vicinity of the production wells, by using a grid refinement near the wells through linking to the pressure gradient. This was caused by the fact that the general streamline metrics tends to the radial one near each well.

The time step was variable and was selected by verifying the Courant criterium of stability.

The results of saturation simulations along all the streamlines are shown in Figure 7.5.

7.4.5. Mapping the saturation in the reservoir

To visualize the results simulated along each streamline, it is necessary to perform the mapping of the results obtained on each $1D$ grid built along each streamline to the $3D$ grid used to simulate pressure. The influence of each streamline on the 3D cell depends on two factors: (i) the time of flight of the particles in the cell trajectory and (ii) the flow rate associates with the cell. This is performed by means of standard mathematical relationships which define the coordinate transformation between two different coordinate frames. Actually, if a cell C crossed by streamline $n_{sl,C}$, the saturation in the cell $S_{l,C}$ can be obtained as:

$$S_{l,C} = \sum_{i=0}^{n_{sl,C}} \frac{q_i \Delta \tau_{i,C}}{\sum_{j=0}^{n_{sl,C}} q_j \Delta \tau_{j,C}} S_{l,i}$$

here $S_{l,i}$ is saturation, q_i is total flow rate and $\Delta \tau_{i,C}$ is time of flight on the cell C for the streamline i.

The presented approach has been implemented in gOcad using the StreamLab plugin. Special tools for visualizing the time evolution of species compositions have been linked to the StreamLab plug-in.

Figure 7.6 illustrates the saturation dynamics in the reservoir, which shows a significant liquid saturation increases through time in the vicinity of the production wells, by forming the well-known condensate banks.
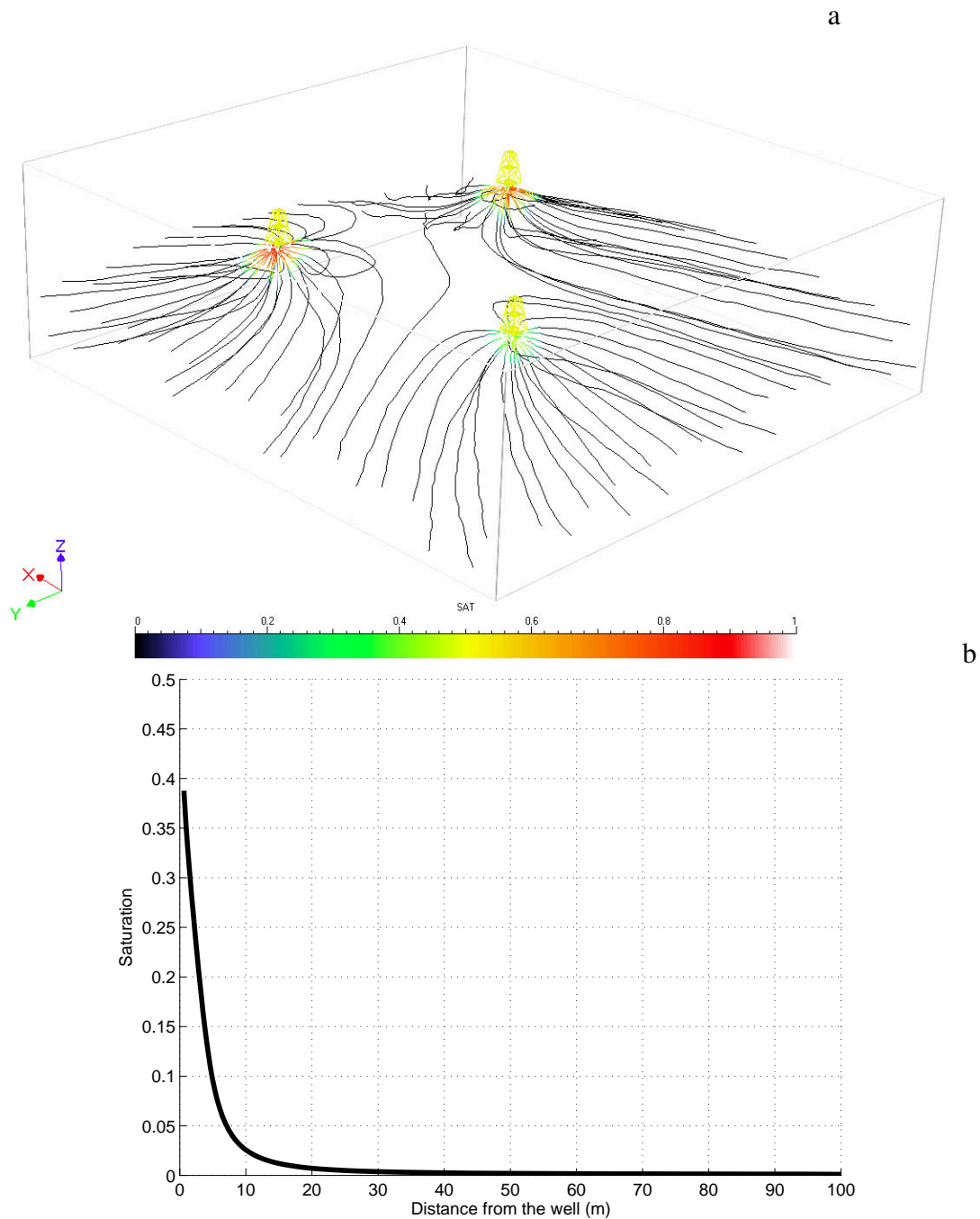
Figure 7.5.: Saturation variation during the natural depletion of a gas-condensate reservoir:
            (a) in the overall 3D system, and (b) along a streamline

7.4.6. Comparison with classical approach

In the compositional reservoirs, like a gas-condensate reservoirs, the significant liquid satura-
tion increases through time observed in the vicinity of the production wells, Figure 7.6. The
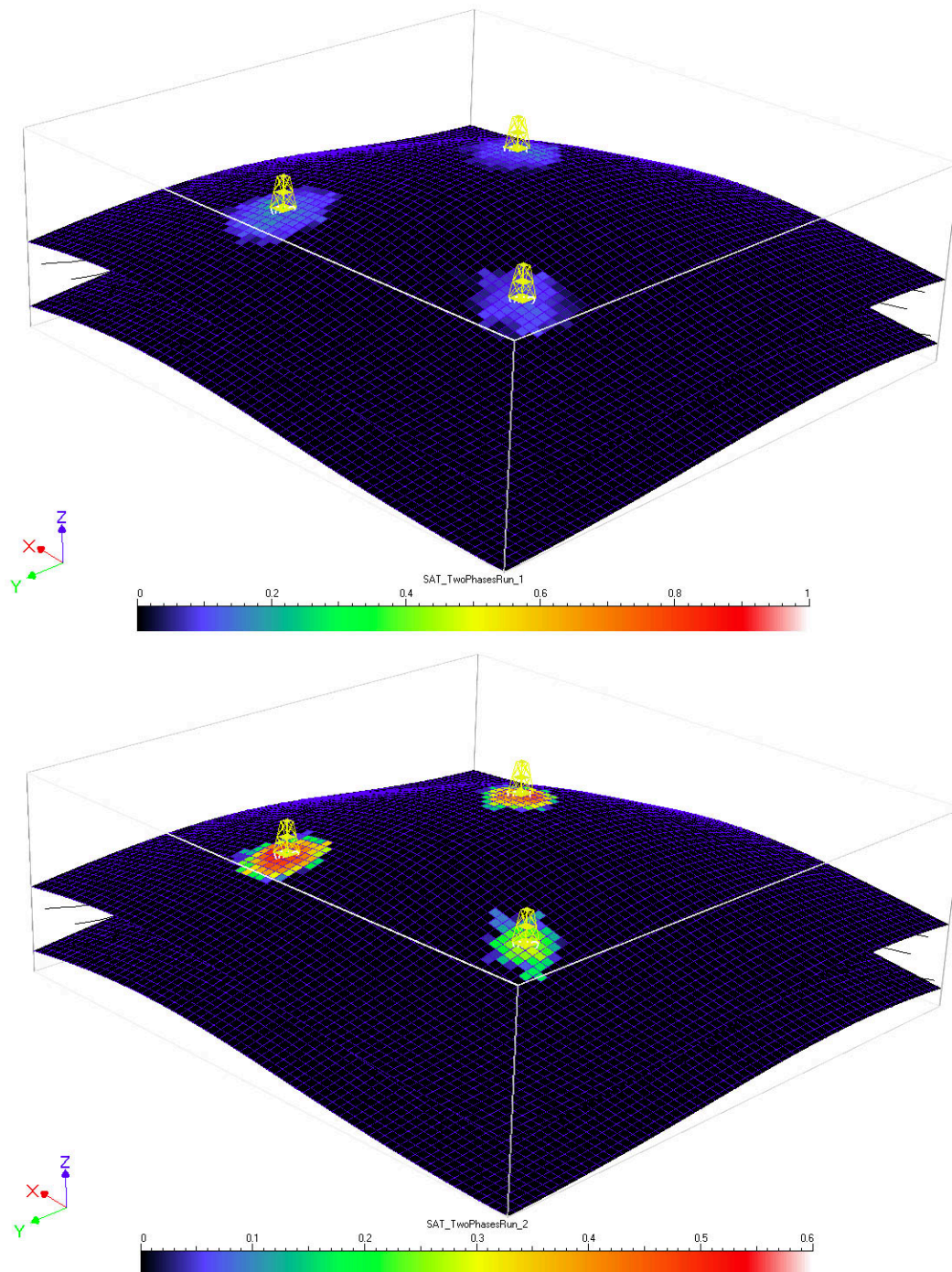
Figure 7.6.: Evolution of saturation against time in a gas-condensate reservoir: 3 month (upper plot) and 3 years (lower plot).

big pressure gradient on the wells (Fig. 7.4) provide the intensive phase exchange processes, which actually provoke the saturation increase around the wells, so-called condensate banks.

Obviously, that in this case the flow around the wells is very close to radial flow type, while far from the wells the flow is not really important (Fig. 7.4). This fact give an opportunity to perform a series of numerical tests using the radial flow approximation for each of three wells with the objective to verify the Streamline HT-splitting approach presented below.

Thus, the simulation of classical full compositional flow model with mass exchange between the phases (section 7.2.1) was done by the radial flow approach without any splitting between thermodynamics and hydrodynamics, Figure 7.7. This kind of tests based on the 9-components mixture and the reservoir description presented in the section 7.4.1. However, one approximation for the radial flow was done: the absolute permeability in each of three domain is the constant value (the average value). The example of comparison between Streamline HT-splitting approach and direct numerical simulation of radial flow around the wells presented on Figure 7.7 for 3 years. The upper, middle and lower plots on the Figure 7.7 correspond to the wells W1, W2 and W3 on the Figure 7.6, counted from the left well counter-clockwise. The lower plot of the Figure 7.6 give the qualitatively similar results for the saturation evolution around the each production well: W1, W2 and W3, presented on the Figure 7.7.

7.5. Conclusions

This work presents an original combination of the HT-split compositional model with the streamline technique, by using the fact that both these approaches are based on the hypothesis of a quasi-stationary pressure field. The HT-split model was developed in (Oladyshkin and Panfilov, 2007b) and used only for the one-dimensional problems. In the present work the HT-split model is applied for a $3D$ flow case for the first time.

The technology of splitting the thermodynamics and the hydrodynamics of the compositional model reduces significantly the complexity of the hydrodynamic part. Due to such a splitting, the hydrodynamics system for the two-phase compositional mixture is simplified up to two equations only. Within this approach the phase compositions are simulated from the independent thermodynamic system which determines the phase concentrations as the function of pressure only.

The pressure field and the streamlines were computed using the StreamLab finite volume flow simulator developed by the research group GOCAD for two-phase flow problems in an arbitrary $3D$ heterogeneous reservoir.
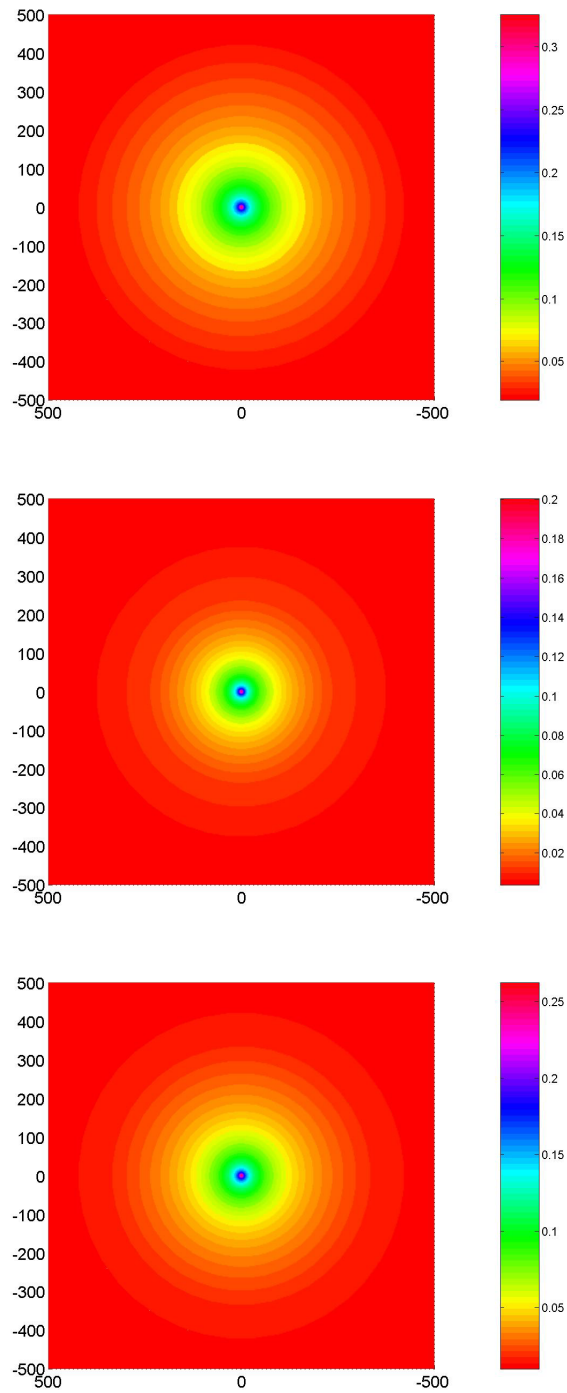
Figure 7.7.: Radial flow approach: saturation on the three production wells of a gas-condensate reservoir after 3 years: W1 (upper plot), W2 (middle plot), and W3 (lower plot).

At present, the adaptation of the HT-split model to the streamline technique was performed for the case of a neglected capillary pressure, an isotropic permeability field, and vanishing gravity effects.

The problem of natural depletion of an as-condensate reservoirs contained three production wells was simulated. The original mixture contained 9 chemical components. The reservoir was heterogeneous.

The saturation distribution in the reservoir was computed by solving the set of $1D$ saturation transport problems along each streamline. The saturation transport equation was formulated in terms of the time-of-flight, in an optimal form for application to the streamline numerical technique.

We have shown that, in contrast to the problem of immiscible two-phase flow, the coefficients of the saturation equation depend both on the pressure though the thermodynamic sub-system and on the pressure space derivatives along streamlines. These derivatives should be calculated once the global $3D$ pressure field is simulated.

The structure of the saturation equation is determined by the competition between the term responsible for the liquid flow and another term responsible for the mass exchange between liquid and gas. The nonlinear character of this equation does not permit to develop analytical solutions in general case of the space metrics.

The obtained numerical results illustrate an adequate scenario of the condensate saturation evolution in space and in time, by predicting the formation of condensate banks in the vicinity of the productors in which the liquid saturation behaves as a highly non-stationary variable and undergoes a fast growth in time.

The developed new simulation technique is much faster that the traditional compositional flow simulation based on the streamline approach but a full (non-split) compositional flow model.

8. Hydrogen penetration in water through porous medium

Bibliographic Note: The content of this chapter is based on the following original article: Oladyshkin S., Panfilov M. Hydrogen penetration in water through porous medium: application to a radioactive waste storage site. Environmental Earth Sciences, Springer, V. 64, N. 4, P. 989-999, 2011.

Hydrogen penetration in water through a porous medium was analyzed in this chapter. A two-phase compositional model approach was considered. The first part of the work deals with the thermodynamic analysis of a hydrogen-water system. The thermodynamic model was calibrated using experimental data of hydrogen solubility in water. The phase densities, viscosities and concentrations were presented in an analytical form. Moreover, the domain of validity of analytical laws for the estimation of phase properties (such as Henry's, Raoult's and Kelvin's laws) was presented for the analyzed system. The second part deals with two-phase hydrodynamic behavior. An analytical solution for a non-compressible flow was constructed. For a general case the influence of relative permeabilities on the flow regimes was analyzed numerically. The notion of pseudo-saturation was introduced to define the appearance of the phases. Low mobile gas created a time displaced front that was relatively slower than the mobile gas flow. Diffusion becomes really important for a low mobile gas case, as the penetration accelerates for a large range of saturation. In contrast to this, the mass exchange phenomenon has a small influence on the flow type. Thus, the regimes of hydrogen penetration in liquid were shown as really sensitive to the relative permeability form.

8.1. Introduction

This work deals with the study of gas migration around storage sites of radioactive waste. Important quantities of hydrogen (H_2) are produced due to radiolysis and corrosion. Other

gases, such as CH_4, C_2H_6, C_2H_2 etc., can also be produced, however, quantities of these gases are very small in comparison to H_2. The document (Andra, 2005) provides detailed information about the mechanisms of hydrogen generation around a radioactive waste storage site. Because of its small molecular size hydrogen has a strong ability of transport in porous media even with almost impermeable properties. Hence, the danger consists in the transport of radioactive elements from the waste storage site by the hydrogen molecule. Let us notice, that water usually exists in porous media around the storage site. That is why the analysis of gas transfer (actually hydrogen transfer) in porous media saturated by water is a non-trivial problem and demands better understanding of transport processes.

The current chapter contains research initiated in scope of the MoMaS project and is based on data of French National Radioactive Waste Management Agency (ANDRA) (see (Andra, 2005). Modeling of gas transport around a radioactive waste storage site asks for very large time scale simulations and, as consequence, demands high computational power. That is why the questions of model complexity describing flow in porous media and possible assumptions are still open. In this chapter we consider a full multiphase compositional model containing gas and liquid phases, which consist of two main components: hydrogen and water. These components can be present in all phases. In this chapter we provide clarification about thermodynamic model complexity and we illustrate errors of simplified thermodynamic relations in comparison to a full compositional formulation (see the section 8.3). Section 8.4 demonstrates the influence of compositional effects on hydrogen transport in porous media. Also, we analyze the influence of relative permeability on the regimes of hydrogen penetration using the presented compositional approach.

8.2. Hydrogen-water compositional model of two-phase flow in a porous medium

We will consider a compositional mixture consisting of 2 chemical components, H_2 and H_2O, capable of forming two thermodynamic phases, gas and liquid, that are separated from one another by an interface. The permanent mass exchange of hydrogen and water components between liquid and gas phases was determined by thermodynamic conditions. We will suppose that there are no chemical reactions, but each component may be dissolved in both phases.

8.2.1. Compositional two-phase flow model

We shall consider a fine dispersed system which is in local equilibrium. According to the phenological approach we will examine gas, liquid and solid as three interpenetrating continuums, (Nikolaevski et al., 1968), (Nigmatulin, 1987), (Sedov, 1976). As accepted in underground thermo and hydrodynamics, we consider an isothermal process, because the overall mass of earth rocks surrounding the examined reservoir plays the role of a huge calorimeter with a huge integral calorific capacity which maintains a constant temperature.

To describe a hydrogen-water flow in a porous medium we will use the general compositional model which is composed in this case of 2 equations of mass balance for each chemical component, H_2 and H_2O, and two equations of momentum balance for each phase, gas and liquid (see (Nigmatulin, 1987), (Sedov, 1976), (Coats, 1980) and (Aziz and Settari, 1979)):

$$\phi \frac{\partial}{\partial t} \left(\rho_l c_l^{(k)} s + \rho_g c_g^{(k)} [1-s] \right) + \mathrm{div} \left(\rho_l c_l^{(k)} \mathbf{V}_l - J_l^{(k)} + \rho_g c_g^{(k)} \mathbf{V}_g - J_g^{(k)} \right), \quad k = 1, 2 \quad (8.1\mathrm{a})$$

$$\mathbf{V}_l = -\frac{K k_l}{\mu_l} \mathrm{grad} P_l, \quad \mathbf{V}_g = -\frac{K k_g}{\mu_g} \mathrm{grad} P_g \qquad (8.1\mathrm{b})$$

where superscript k refers to k-th chemical component (H_2 denoted as w or H_2O denoted as h), indexes g and l to gas and liquid; ϕ is porosity; ρ is the phase density; $c_g^{(k)}$ or $c_l^{(k)}$ is the mass concentration of k-th component in gas or liquid; s is the volume saturation of pores by liquid; k_g and k_l are the relative permeabilities; μ is the phase viscosity; K is the absolute permeability; P is the phase pressures, V is the Darcy velocity.

The momentum balance equations 8.1b are written in the form of the Darcy law for each phase, thus we neglect the momentum exchange at phase transitions. The phase viscosity is a given function of pressure and phase composition. The absolute permeability K and porosity ϕ are given functions of space coordinates. The structures of relative permeabilities may be arbitrary, in particular they may be examined as functions of velocity, as shown in (Henderson et al., 1995), but not only of saturation.

Due to the capillarity, a phase pressure difference appears in the compositional model:

$$P_g - P_l = P_c(s) \qquad (8.2)$$

P_c - is the given function of the effective capillary pressure. Let us note, that we will us the notation P for the gas pressure (P_g) below.

The diffusive flow of dissolved hydrogen can be expressed by applying Fick's Law:

$$J_l^h = \rho_l D_l^h \mathrm{grad} c_l^h \qquad (8.3)$$

$$J_g^h = \rho_g D_g^h \text{grad} c_g^h \tag{8.4}$$

here the diffusion coefficients, (Andra, 2005), D_l^h and D_g^h are:

$$D_l^h = D_l^0 s \frac{\mu_l^0}{\mu_l}, \quad D_l^0 = 1.57 \cdot 10^{-14} \frac{\phi T}{\theta^2 \mu_l^0}$$

$$D_g^h = D_g^0 (1-s) \frac{P_0}{P}, \quad D_g^0 = \frac{D_0 \phi}{\theta^2} \frac{P_{at}}{P_0} \left(\frac{T}{T_0} \right)^{1.75}$$

θ is tourtosity and $T_0 = 303K$, $P_{at} = 1.01 \cdot 10^5 Pa$, $D_0 = 9.5 \cdot 10^{-5} m^2/s$; P_0 - characteristic values of the pressure.

The system of 4 equations 8.1 is written with respect to pressure P, saturation s, velocities V_l and V_g. Thus, for two components, h and w, the system 8.1 does not contain independent concentrations. This means that all concentrations and phase densities are determined by additional thermodynamic relationships. In other words, for a mixture consisting of two chemical components it is possible to totally split the general compositional model into a thermodynamic system and a hydrodynamic system which may be resolved independently of one another. Let us note that such splitting is not possible for general class of multi-compositional systems, however, can be performed for a particular case along streamlines (Oladyshkin and Panfilov, 2007a).

8.2.2. Thermodynamic closure relationships

The closure relationships for system 8.1 describe the local equilibrium thermodynamic behavior. They consist of 2 equilibrium equations in terms of chemical potentials 8.5a for chemical component h and w, two equations of phase state 8.5b and two normalizing equations for concentrations 8.5c:

$$v_g^h \left(P, c_g^h, c_g^w \right) = v_l^h \left(P, c_g^h, c_g^w \right), \tag{8.5a}$$

$$v_g^w \left(P, c_g^h, c_g^w \right) = v_l^w \left(P, c_g^h, c_g^w \right),$$

$$\rho_g = \rho_g \left(P, c_g^h, c_g^w \right), \tag{8.5b}$$

$$\rho_l = \rho_l \left(P, c_g^h, c_g^w \right),$$

$$c_g^h + c_g^w = 1, \tag{8.5c}$$

$$c_l^h + c_l^w = 1$$

where v_i^h and v_i^w are the chemical potential of components h and w in i^{th} phase ($i = g, l$). Equations of state 8.5b are written in a general form. Functions $v_g^{(k)}(P, ...)$ are given. Various versions of these functions may be found in (Danesh, 1998), (Batalin et al., 1992). A detailed description of the thermodynamic system will be presented below.

The system 8.5a – 8.5c includes 6 equations with respect to 7 variables: P, ρ_g, ρ_l, c_g^h, c_g^w, c_l^h, c_l^w. The difference between the number of variables and the number of equations, $v = 7 - 6 = 1$, is called the *thermodynamic variance* and determines the number of independent variables. We select the pressure as an independent variable. As a result the phase densities and concentrations depend only on the pressure P, which constitute the Gibbs rule of phase.

8.2.3. Model of relative permeabilities

The flow of a two-phase system in a porous medium strongly depend on the petrophysic closure relations. We will analyze the influence of relative permeability on the regimes of hydrogen penetration in water in a porous medium. In this chapter we shall consider two principal models of gas penetration: the model of mobile gas and the model of low mobile gas.

Mobile gas model:

$$k_g(s) = 1 - s^{\alpha}, \tag{8.6a}$$

$$k_l(s) = \gamma(s - s_r)^{\beta} \tag{8.6b}$$

Low mobile gas model:

$$k_g(s) = (1 - s)^{\alpha}, \tag{8.7a}$$

$$k_l(s) = \gamma(s - s_r)^{\beta} \tag{8.7b}$$

8.3. Hydrogen-water thermodynamic behavior

In the previous section we showed that hydrodynamics and thermodynamics may be split for a hydrogen-water mixture. As the result of such a splitting we obtain a hydrodynamic model 8.1 consisting of two equations (for two variables - pressure and saturation) and a thermodynamic model 8.5. The independent thermodynamic model for the hydrogen-water

mixture consists of two equilibrium equations, two equations of phase state and two normalizing equations for concentrations. In this section we will describe the mathematical model for hydrogen-water ($h-w$) two-phase system through thermodynamic relations.

8.3.1. Phase equilibrium conditions

The equality of chemical potentials of each component in all the co-existing phases at equilibrium is described by the equation (8.5a). The chemical potentials for gas v_g^h, v_g^w and liquid v_l^h, v_l^w phases of components h and w can be defined as:

$$v_i^h = v_{ip}^h \left(P, T, c_i^h, c_i^w \right) + RT \ln \left(f_i^h c_i^h \right), \tag{8.8}$$

$$v_i^w = v_{ip}^w \left(P, T, c_i^h, c_i^w \right) + RT \ln \left(f_i^w c_i^w \right), i = g, l$$

where R is the universal gas constant; T and P are the reservoir temperature and pressure; f_i^h and f_i^w are the fugacities in phase i; v_{ip}^h and v_{ip}^w are the chemical potentials of the ideal gas:

$$v_{ip}^h = G_p^h (P, T) + RT \ln \left(c_i^h \right), \tag{8.9}$$

$$v_{ip}^w = G_p^w (P, T) + RT \ln \left(c_i^w \right), \quad i = g, l$$

herein $G_p^h (P, T)$ and $G_p^w (P, T)$ are the chemical potential of the pure components h and w correspondently.

Therefore, using (8.8) and (8.9) the phase equilibrium conditions (8.5a) lead to a balance between the fugacities:

$$f_g^h \left(P, T, c_g^h, c_g^w \right) = f_l^h \left(P, T, c_l^h, c_l^w \right), \tag{8.10}$$

$$f_g^w \left(P, T, c_g^h, c_g^w \right) = f_l^w \left(P, T, c_l^h, c_l^w \right)$$

The fugacities will be introduced through the equation of state for gas-liquid systems, section 8.3.2, and will be described in section 8.3.3.

8.3.2. Equations of state for gas-liquid system

Equations of state are basically developed for pure components (Redlich and Kwong, 1949), (Walas Stanley, 1984), but may be applied to multicomponent systems by using mixing

rules which determine the averaged parameters for a mixture (Peng and Robinson, 1976), (Orbey and Sandler, 1998). The mixing rules take into account the prevailing forces between molecules of different substances which form the mixture.

An equation of state like Peng-Robinson is usually formulated with respect to the z-factor z (compressibility) as an implicit function of P, T and concentrations:

$$z_i^3 - (1 - B_i)z_i^2 + (A_i - 3B_i^2 - 2B_i)z - (A_iB_i - B_i^2 - B_i^3) = 0, \quad i=g,l \qquad (8.11)$$

where z_i is the z-factor of i-th phase; coefficients A_i, B_i depend on pressure P, temperature T and mixture composition:

$$A_i = \bar{a}_i \frac{P}{R^2 T^2}, \quad B_i = \bar{b}_i \frac{P}{RT}, \quad i=g,l \qquad (8.12)$$

Parameters $\bar{a}_i$ and $\bar{b}_i$ describe the mixture composition and the mixing rules:

$$\bar{a}_i = x_i^h x_i^h a^{h,h} + x_i^h x_i^w a^{h,w} + x_i^w x_i^h a^{w,h} + x_i^w x_i^w a^{h,w}, \qquad (8.13)$$

$$a^{kj} = \sqrt{a^k a^j}(1 - \delta^{kj}), \quad k,j = h,w,$$

$$\bar{b}_i = x_i^h b^h + x_i^w b^w, \quad i=g,l$$

where δ^{kj} is known as a binary interaction parameter (Danesh, 1998); x_i^h and x_i^w are the mole fractions of components h and w in the i-th phase (gas or liquid).

Parameters a^k and b^k for k-th chemical component (h or w) are defined by the following equations:

$$a^k = 0.4274800232 \frac{R^2 T_{cr}^k{}^2}{P_{cr}^k} \left[1 + m^k \left(1 - \sqrt{\frac{T}{T_{cr}^k}} \right) \right]^2, \qquad (8.14\text{a})$$

$$m^k = 0.37464 + 1.5422\omega^k - 0.26992\omega^k{}^2, \qquad (8.14\text{b})$$

$$b^k = 0.086640350 \frac{RT_{cr}^k}{P_{cr}^k}, \quad k=h,w \qquad (8.14\text{c})$$

where T_{cr}^k and P_{cr}^k are the critical fluid temperature and pressure for component k (h, w); ω^k is the acentric Pitcer factor.

The cubic equation (8.11) for the z-factor may be solved for liquid and vapor phases. Generally, three solutions of equation (8.11) are obtained. Finally, the smallest root corresponds to liquid, while the largest root describes vapor.

8.3.3. Fugacities of chemical components

Let us introduce the fugacities to describe the attraction/repulsion between molecules in the hydrogen-water mixture.

The ratio of the fugacity to the pressure is called the fugacity factor ϕ_i^h, ϕ_i^w:

$$\phi_i^h = \frac{f_i^h}{Px_i^h}, \quad \phi_i^w = \frac{f_i^w}{Px_i^w}, \quad i = g, l \tag{8.15}$$

The fugacity factors are calculated using following relation:

$$
\ln\left(\phi_i^h\right) = \frac{b^h}{\bar{b}_i}(z_i - 1) - \ln(z_i - B_i) -
$$

$$
\frac{A_i}{2\sqrt{2}B_i}\left(\frac{2\left(x_i^h a^{h,h} + x_i^h a^{h,w}\right)}{\bar{a}_i} - \frac{b^h}{\bar{b}_i}\right) \cdot \ln\left(\frac{z_i + (\sqrt{2}+1)B_i}{z_i - (\sqrt{2}-1)B_i}\right), \tag{8.16}
$$

$$
\ln\left(\phi_i^w\right) = \frac{b^w}{\bar{b}_i}(z_i - 1) - \ln(z_i - B_i) -
$$

$$
\frac{A_i}{2\sqrt{2}B_i}\left(\frac{2\left(x_i^w a^{w,h} + x_i^w a^{w,w}\right)}{\bar{a}_i} - \frac{b^w}{\bar{b}_i}\right) \cdot \ln\left(\frac{z_i + (\sqrt{2}+1)B_i}{z_i - (\sqrt{2}-1)B_i}\right)
$$

where z_i, A_i, B_i, $\bar{a}_i$, $\bar{b}_i$ and a^{kj} are defined by equations above ($i = g, l$).

Thus, the fugacity factor of a component depends on the composition and the z-factor of the mixture (z_i), on the additional functions ($\bar{a}_i$ and $\bar{b}_i$), on the properties of the pure components (T_{cr}^k, P_{cr}^k, and ω^k) and on the cross parameters δ^{kj}.

8.3.4. Thermodynamic closure relations

To close the thermodynamic system we will introduce the relation below.

Let us introduce a relation between the mole fractions x_i^h, x_i^w and the mass concentrations c_i^h, c_i^w:

$$c_i^h = \frac{m^h}{V}, \quad c_i^w = \frac{m^w}{V} \tag{8.17}$$

where V is the volume; m^h and m^w are the mass of the chemical components h and w, i.e. m^k is the product of the mole number n^k and the molar mass M^k:

$$m^h = n^h M^h, \quad m^w = n^w M^w \tag{8.18}$$

Therefore:

$$n^h = \frac{c_i^h V}{M^h}, \quad n^w = \frac{c_i^w V}{M^w} \tag{8.19}$$

On the other hand, the mole fractions x_i^h and x_i^w in phase i are the ratio of the mole number for components h and w to the total number of moles:

$$x_i^h = \frac{n^h}{n}, \quad x_i^w = \frac{n^w}{n} \tag{8.20}$$

Taking into account the fact that the total number of moles is the sum of mole numbers for each component, $n = n_h + n_w$, we obtain:

$$x_i^h = \frac{\dfrac{c_i^h}{M^h}}{\dfrac{c_i^h}{M^h} + \dfrac{c_i^w}{M^w}}, \quad x_i^w = \frac{\dfrac{c_i^w}{M^w}}{\dfrac{c_i^h}{M^h} + \dfrac{c_i^w}{M^w}}, \quad i = g, l \tag{8.21}$$

We also need to use the inverse relation for the concentrations c_i^h and c_i^w:

$$c_i^h = \frac{x_i^h M^h}{x_i^h M^h + x_i^w M^w}, \quad c_i^w = \frac{x_i^w M^w}{x_i^h M^h + x_i^w M^w}, i = g, l \tag{8.22}$$

The fluid densities can be calculated as:

$$\rho_i = \frac{P M_i}{zRT}, \quad i = g, l \tag{8.23}$$

where M_i is the molar mass for i-th phase, $M_i = x_i^h M^h + x_i^w M^w$.

8.3.5. Calibrated thermodynamic model for hydrogen-water system

The described mathematical model is based on a compositional approach depending on the numbers of a binary interaction and pseudo individual parameters. Thus, such a theoretical model describing the phase behavior needs to be calibrated with experimental data.

Usual thermodynamic conditions for the underground storage of radioactive waste are a pressure of 50 bars and a temperature about 20^oC. For the calibration of the hydrogen-water
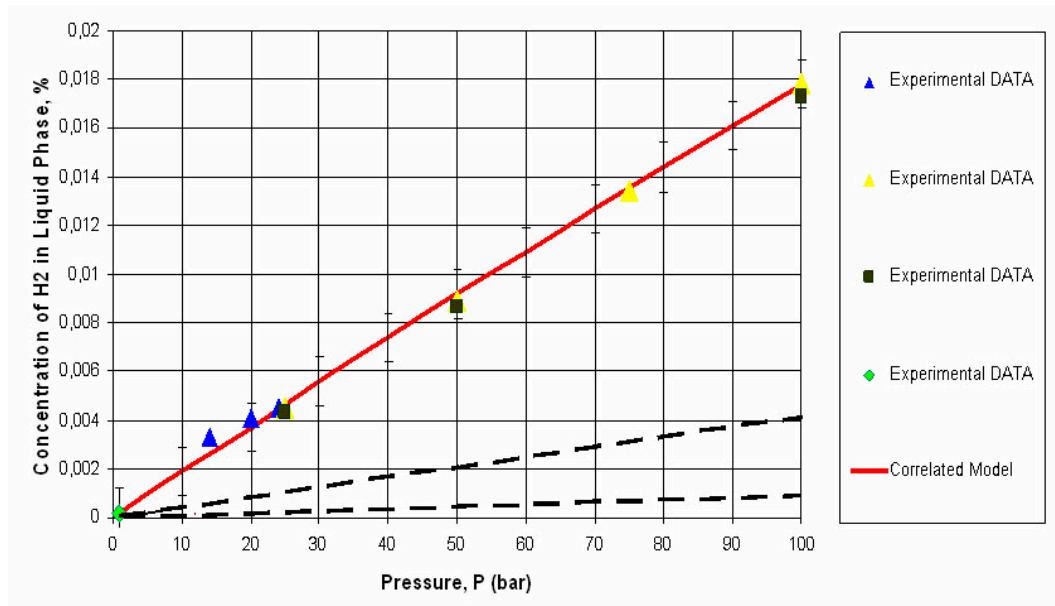
Figure 8.1.: Calibration of H_2 solubility in H_2O: non-correlated model (dashed curve) and correlated model (solid curve)

system we will consider a large pressure range, but constant temperature. This constant approximation is due to the small temperature variation from 19.6^oC up to 23.3^oC around the storage site.

The calibration of a mathematical model for compositional thermodynamics is based on the fitting of experimental data. The most common representation of experimental data is known in terms of hydrogen solubility in water: (Pray et al., 1952), (Wiebe and Gaddy, 1934), (Wiebe et al., 1932), (Morrison and Billet, 1952). The calibration of a PVT-model consists in corrections of the binary interaction parameters and the pseudo individual parameters. Applying of the phase equilibrium law to gases like H, He and Ne requires the adaptation of the critical constant of temperature T_c and pressure P_c. For hydrogen the modifications of such pseudo individual parameters are obtained in (Gunn et al., 1966):

$$T_c = \frac{43.6}{1+10.8/T}, \quad P_c = \frac{20.2}{1+21.9/T}$$

Here temperature T is measured in K and pressure P is measured in atm. On the other hand, the research of the binary interaction parameter is similar to the iterative process of the root search for an nonlinear system.

The calibrated theoretical model is shown on Figure 8.1 by solid a line. Unfortunately, the non-calibrated theoretical model can increase the error and be totally non-realistic (dashed

curves on Figure 8.1). Finally, the binary interaction parameter for the $h - w$ mixture $\delta^{h,w}$ is -0.121. The individual parameters for a calibrated PVT-model are presented in Table 8.1.

Component	h	w
Molar Weight	2.016	18.015
Critical Presssure (bar)	18.82	220.48
Critical Temperature (K)	42.08	647.3
Omega A	0.45724	0.45724
Omega B	0.077796	0.077796
Acentric Factor	-0.218	0.344
Parachors	34.0	53.1
V Critical (m3 /kg-mole)	0.065	0.22942
Z Critical	0.34965	0.056
Boil Temperature (K)	20.3	373.2

Table 8.1.: Individual parameters

8.3.6. Polynomial form for the thermodynamic variables

In this section we propose illustrations of hydrogen-water thermodynamic behavior using the calibrated data in the range of pressure from 1 up to 100 bars.

The dissolution of hydrogen in water shows linear behavior relative to pressure and is presented in Figure 8.2 (left plot) in terms of hydrogen concentration in the liquid phase. On the other hand, phase transitions leads to a the non-linear dependency of the hydrogen concentration from pressure in the gas phase, see right plot in Figure 8.2. Let us note that thermodynamic properties of the liquid such as density and viscosity show pseudo-constant behavior. The thermodynamic properties of gas keep a linear behavior relative to pressure.

Finally, the calibrated mathematical model of a $h - w$ system can be presented in the simple polynomial form of thermodynamic variables in respect to pressure. The concentration of h in % for the liquid and the gas phases, the phases densities in KG/M^3 and the phases
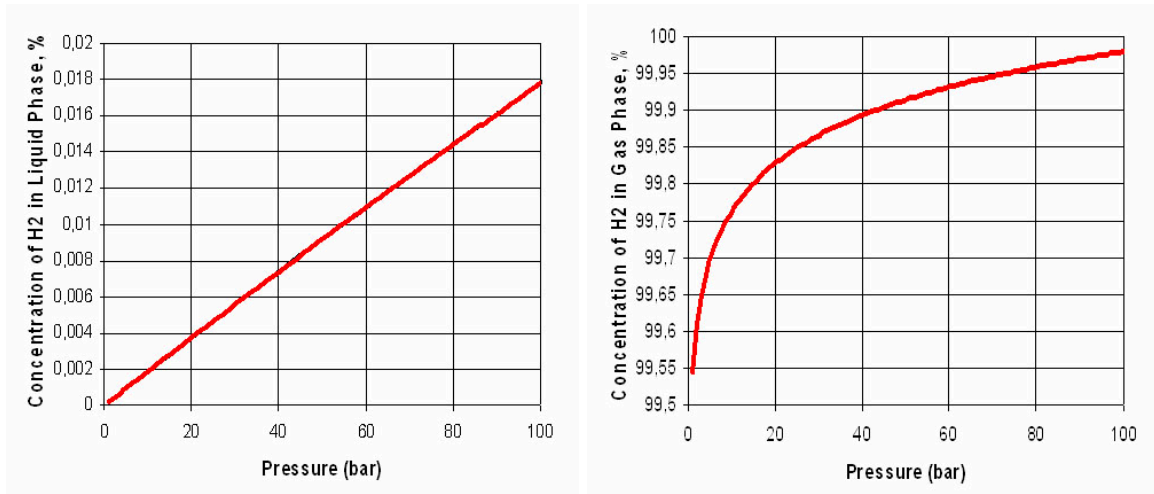
Figure 8.2.: Hydrogen concentration in liquid (left) and in gas (right) phases

viscosity in CPOISE is presented below relative to the pressure in bar:

$$c_l^h = 0.0002P + 10^{-5} \quad (\%), \tag{8.24a}$$

$$c_g^h = 0.0944\ln(P) + 99.545 \quad (\%), \tag{8.24b}$$

$$\rho_l = 0.0089P + 858.74 \quad (KG/M^3), \tag{8.24c}$$

$$\rho_g = 0.0817P + 0.0155 \quad (KG/M^3), \tag{8.24d}$$

$$\mu_l = 3 \cdot 10^{-5}P + 0.3258 \quad (CPOISE), \tag{8.24e}$$

$$\mu_g = 8 \cdot 10^{-9}P^2 + 10^{-6}P + 0.0082 \quad (CPOISE) \tag{8.24f}$$

Thus, the total splitting of hydrodynamics and thermodynamics allows to present the thermodynamic properties of a hydrogen-water two-phase system in an analytical form. In other words, the thermodynamic behavior can be introduced through the coefficients in the hydrodynamic equations, and hence the general compositional model can be analysed.

8.3.7. Error estimation for simplified analytical laws

The properties of a two-phase system can be estimated using simplified analytical laws. Analytical expressions allowing to calculate the pressure exist in form of the laws of ideal gas, such as Henry's, Raoult's, Kelvin's laws. Unfortunately, this kind of analytical simplified laws can also lead to errors in estimation of the system's properties. In this section we

propose to analyze the domain of validity of such laws for the presented hydrogen-water two-phase system.

Primarily, the phase equilibrium analysis was started with the ideal gas law using the equation of state of a hypothetical ideal gas. For the gaseous mixture an empirical law related to the ideal gas laws was observed by Dalton. Mathematically, the pressure of a mixture of gases can be defined as the sum of the partial pressures of each individual component in a gas mixture:

$$P_g = P_g^h + P_g^w \qquad (8.25)$$

where P_g^h and P_g^w are the partial pressures of hydrogen and water in the gas phase:

$$P_g^w = \frac{\rho_g^w}{M^w} RT, \quad P_g^h = \frac{\rho_g^h}{M^h} RT \qquad (8.26)$$

The statement of Henry's law is that the amount of a gas dissolved by a liquid is proportional to the pressure of the gas upon the liquid. This law was originally formulated by William Henry in 1803 for dilute solutions and low gas pressures:

$$M^h H(T) P_g^h = \rho_l^h \qquad (8.27)$$

here $H(T)$ is the Henry constant depending only on the temperature.

The Raoult's law is based on the following formulation: the vapor pressure of a solution of a non-volatile solute is equal to the vapor pressure of the pure solvent at this temperature multiplied by its mole fraction. The vapor pressure of the pure solvent depends only on the temperature and therefore is a constant:

$$P_g^w = \overline{p}_g^w \frac{\rho_{st}^w}{\rho_l^w + \frac{M^w}{M^h} \rho_l^h} \qquad (8.28)$$

where $\overline{p}_g^w$ is the vapor pressure of the pure solvent.

Further, the influence of capillary forces can be introduced using Kelvin's formulation:

$$P_g^w = \overline{p}_g^w \frac{\rho_{st}^w}{\rho_l^w + \frac{M^w}{M^h} \rho_l^h} e^{-M^h \frac{P_c}{RT\rho_l}} \qquad (8.29)$$

However, in the current chapter we considered a real mixture and it is obviously that laws based on the ideal gas approach can not be applied for a large pressure range. Thus, we propose an estimation of error for the described analytical laws in application to the hydrogen-water two-phase system.
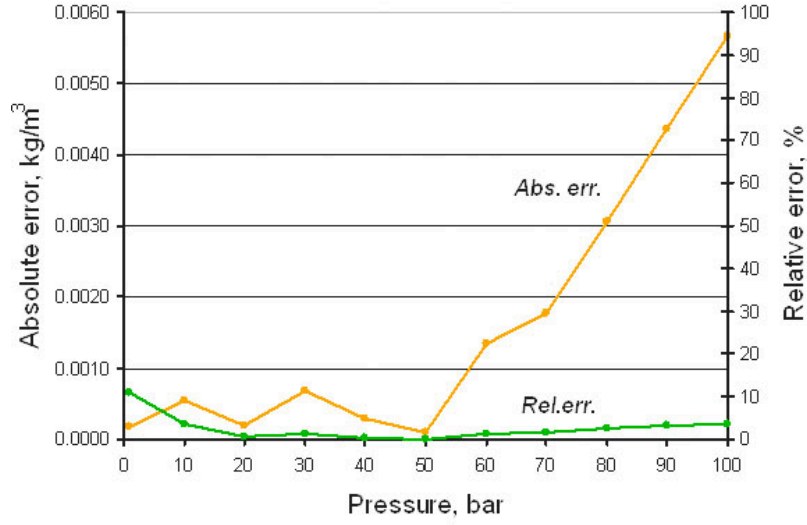
Figure 8.3.: Absolute end relative errors of Henry's law for the hydrogen-water two-phase system

We will use the calibrated data obtained from the compositional approach to calculate the absolute and the relative error of Henry's law:

$$\varepsilon_{abs}^H = \left| M^h H(T) P_g^h - \rho_l^h \right|, \tag{8.30}$$

$$\varepsilon_{rel}^H = 2\varepsilon_{abs}^H / \left(M^h H(T) P_g^h + \rho_l^h \right)$$

As we can see in Figure 8.3, the relative error of Henry's law for the hydrogen-water two-phase system is not more than 15%.

The errors of Raoult's and Kelvin's laws can be defined in the same way:

$$\varepsilon_{abs}^K = \left| P_g^w - \overline{p}_g^w \frac{\rho_{st}^w}{\rho_l^w + \frac{M^w}{M^h}\rho_l^h} e^{-M^h \frac{P_c}{RT\rho_l}} \right|,$$

$$\varepsilon_{rel}^K = 2\varepsilon_{abs}^R \left(P_g^w + \overline{p}_g^w \frac{\rho_{st}^w}{\rho_l^w + \frac{M^w}{M^h}\rho_l^h} e^{-M^h \frac{P_c}{RT\rho_l}} \right)^{-1} \tag{8.31}$$

where ε_{abs}^R and ε_{rel}^R can be calculated for $P_c = 0$.

The relative errors of Kelvin's and Raoult's approximations essentially increase from a pressure of 30 bar, Figure 8.4. Evidently, Kelvin's and Raoult's laws can not be applied for a high pressure range, not even as an approximation. On the other hand, Henry's law gives a more accurate analytical approximation for a hydrogen-water two-phase mixture.
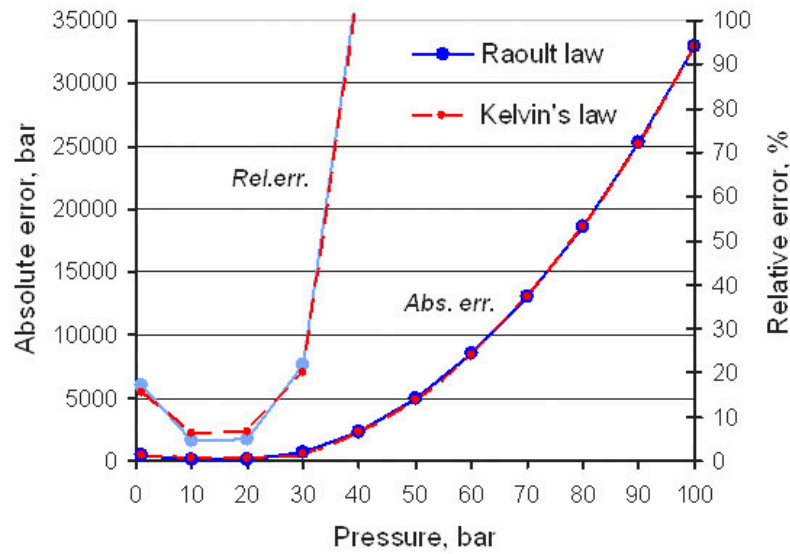
Figure 8.4.: Absolute and relative errors of Raoult's and Kelvin's laws for the hydrogen-
water two-phase system

8.4. Hydrogen-water hydrodynamic behavior

This section is devoted to the analysis of hydrodynamic behavior in a hydrogen-water two-phase system. The hydrodynamic model was formulated for a general case of a 3D hydrogen-water flow with phase transitions (the equation 8.1). The influence of relative permeabilities on the flow behavior will be analysed in this section as well as the influence of a phase exchange effect and the diffusion phenomenon. The thermodynamics was introduced through coefficients of the hydrodynamic equations. Such thermodynamics was shown to be monovariant so that all thermodynamic variables can be presented in analytical form as functions of pressure.

8.4.1. Dimensionless form of flow problem

Let us replace the flow velocities in equation 8.1 by Darcy's law, then the compositional model takes the following form:

$$\phi \frac{\partial \rho}{\partial t} = \text{div}\left([\Psi_l + \Psi_g] \text{grad} P \right) - \text{div}\left(\Psi_l \text{grad} P_c \right), \tag{8.32a}$$

$$\phi \frac{\partial \rho^{(k)}}{\partial t} = \text{div}\left(\left[\Psi_l c_l^{(k)} + \Psi_g c_g^{(k)} \right] \text{grad} P \right) - \text{div}\left(\Psi_l c_l^{(k)} \text{grad} P_c \right) \tag{8.32b}$$

$$+ \text{div}\left(D_l^0 s \rho_l \text{grad} c_l^{(k)} + D_g^0 (1-s) \rho_g \frac{P_0}{P} \text{grad} c_g^{(k)} \right), \quad k = 2$$

here $\Psi_i \equiv \rho_i K k_i / \mu_i$ is the hydraulic phase conductivity $(i = l, g)$; $\rho \equiv \rho_l s + \rho_g (1-s)$ is the total density; $\rho^{(k)} \equiv \rho_l c_l^{(k)} s + \rho_g c_g^{(k)} (1-s)$ is the total partial density of the component k.

We will introduce the characteristic values of length, time, gas pressure and capillary pressure, viscosities and densities by: L, t_*, P_0, $\widehat{P}_c$, μ_g^0, μ_l^0, ρ_g^0 and ρ_l^0; and we will perform the operations div and grad in dimensionless space coordinates. Hence, we obtaine the following dimensionless definitions:

$$p \equiv \frac{P}{P_0}, p_c \equiv \frac{P_c}{\langle P_c \rangle}, \varphi_i \equiv \frac{\phi \rho_i}{\langle \phi \rangle \rho_i^0}, \psi_i \equiv \frac{K \rho_i \mu_i^0}{\langle K \rangle \rho_i^0 \mu_i}, \overline{\rho} \equiv \frac{\rho_l^0}{\rho_g^0}, \tau \equiv t / t_*,$$

$$\omega \equiv \frac{\rho_l^0 \mu_g^0}{\rho_g^0 \mu_l^0}, \varepsilon \equiv \frac{t^*}{t_*}, \frac{1}{Ca} \equiv \frac{\langle P_c \rangle}{P_0}, t^* \equiv \frac{L^2 \mu_g^0 \langle \phi \rangle}{\langle K \rangle P_0}, Pe \equiv \frac{\langle K \rangle P_0}{D_g^0 \mu_g^0}, \sigma \equiv \frac{D_l^0}{D_g^0}$$

Here the parameter t^* is the characteristic time of perturbation propagation caused by the pressure variation. The parameter ε is the ratio of the perturbation propagation time t^* to the characteristic process time t_*, while the parameter ω is the ratio of the liquid mobility to the gas mobility.

Thus, the full compositional model 8.1 takes the following dimensionless form:

$$\varepsilon \frac{\partial}{\partial \tau} \left(\varphi_l \overline{\rho} s + \varphi_g (1-s) \right) = \text{div}\left([\psi_g k_g + \omega \psi_l k_l] \text{grad} p \right) - \frac{\omega}{Ca} \text{div}\left(\psi_l k_l \text{grad} p_c \right), \quad (8.33a)$$

$$\varepsilon \frac{\partial}{\partial \tau} \left(\varphi_l \overline{\rho} s c_l^{(N)} + \varphi_g (1-s) c_g^{(N)} \right) = \text{div}\left(\left[\psi_g k_g c_g^{(N)} + \omega \psi_l k_l c_l^{(N)} \right] \text{grad} p \right) \tag{8.33b}$$

$$- \frac{\omega}{Ca} \text{div}\left(\psi_l k_l c_l^{(N)} \text{grad} p_c \right) + \frac{1}{Pe} \text{div}\left(\frac{\varphi_g (1-s)}{p} \text{grad} c_g^{(N)} + \sigma \omega \psi_l s \text{grad} c_l^{(N)} \right)$$

The first equation 8.33a describes the pressure behavior and the second equation 8.33b describes the saturation evolution. In fact, the saturation evolution is caused by the following

effects: convection, capillary influence and diffusion in liquid:

$$\frac{\varepsilon}{\omega}\frac{\partial s}{\partial \tau} \sim \frac{\partial}{\partial x}\left(\left[\psi_l k_l c_l^{(N)}\right]\frac{\partial p}{\partial x}\right) - \frac{1}{Ca}\frac{\partial}{\partial x}\left(\psi_l k_l c_l^{(N)}\frac{\partial p_c}{\partial x}\right) + \frac{\sigma}{Pe}\frac{\partial}{\partial x}\left(\psi_l s\frac{\partial c_l^{(N)}}{\partial x}\right)$$

In order to analyse hydrogen migration around storage site of radioactive waste we will consider the 1D flow problem on a dimensionless domain $[x_* : 1]$. Hydrogen is produced due to corrosion of metal components of the storage facility and can be described by the given function of the flow rate $q(\tau)$ at the source x_*. At the same time, the pressure at the exterior bound of the domain $(x = 1)$ can be the function of time $p_*(\tau)$. We will suppose that at the initial stage there is no gas in the porous medium, i.e. s=1. However, due to the hydrogen production at the source $x = x_*$ the liquid saturation equal to 0, i.e. s=0. Finally, the formulation 8.33 can be completed by the following initial and boundary-value conditions:

$$s|_{\tau=0} = 1, \tag{8.34a}$$

$$p|_{x=1} = p_*(\tau), \tag{8.34b}$$

$$(\psi_g k_g + \omega\psi_l k_l)\frac{\partial p}{\partial x}\bigg|_{x=x_s} = -\frac{\varepsilon q(\tau)}{2} \tag{8.34c}$$

$$s|_{x=x_s} = 0 \tag{8.34d}$$

The hydrogen-water two-phase system corresponds to the flow around the radioactive waste storage site. The essential properties of this system are the following: average absolute permeability $K = 0.00005mD$, average reservoir porosity $\phi = 0.15$, temperature $T = 298.1500K$ and initial pressure is 100 bars. The thermodynamic properties of the system are taken from the correlated model presented in section 8.3.

8.4.2. Analytical solution for a non-compressible flow

An analytical solution can be constructed for the case of a non-compressible flow, i.e. $\psi_l = 1$, $\psi_g = 1$ and $\varphi_l = 1$, $\varphi_g = 1$. For the 1D flow the system 8.33 takes the form:

$$\varepsilon\frac{\partial}{\partial \tau}(\overline{\rho}s + 1 - s) = \frac{\partial}{\partial x}\left([k_g + \omega k_l]\frac{\partial p}{\partial x}\right), \tag{8.35a}$$

$$\varepsilon\overline{\rho}\frac{\partial s}{\partial \tau} = \omega\frac{\partial}{\partial x}\left(k_l\frac{\partial p}{\partial x}\right) - \frac{\omega}{Ca}\frac{\partial}{\partial x}\left(k_l\frac{\partial p_c}{\partial x}\right) + \frac{1}{Pe}\frac{\partial}{\partial x}\left(\frac{1-s}{p}\frac{\partial c_g^{(N)}}{\partial x} + \sigma\omega s\frac{\partial c_l^{(N)}}{\partial x}\right). \tag{8.35b}$$
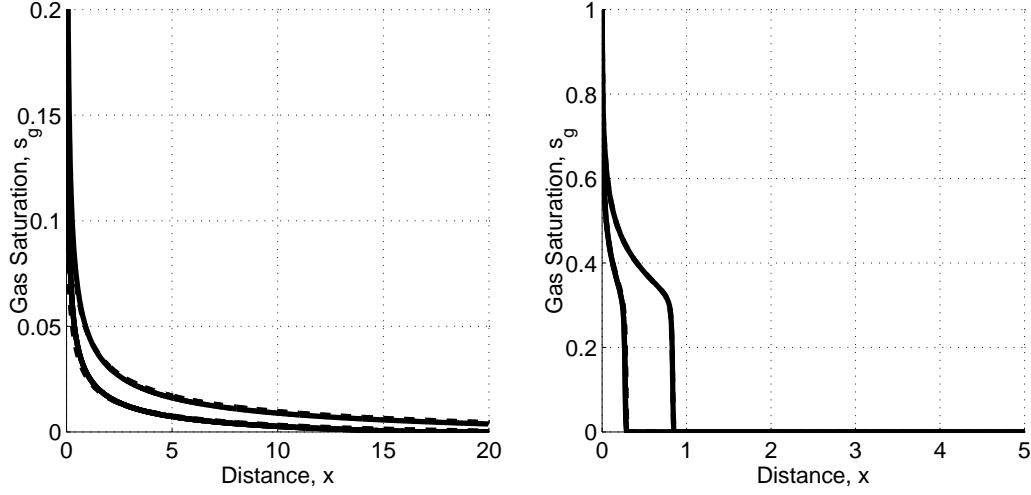
Figure 8.5.: Analytical (solid line) and numerical (dashed line) solutions for a non-compressible flow: left plot - mobile gas; right plot - low mobile gas

Let us substitute equation 8.35b by equation 8.35a:

$$-\varepsilon\frac{\partial s}{\partial \tau} = \frac{\partial}{\partial x}\left(k_g\frac{\partial p}{\partial x}\right) + \frac{\omega}{Ca}\frac{\partial}{\partial x}\left(k_l\frac{\partial p_c}{\partial x}\right) - \frac{1}{Pe}\frac{\partial}{\partial x}\left(\frac{1-s}{p}\frac{\partial c_g^{(N)}}{\partial x} + \sigma\omega s\frac{\partial c_l^{(N)}}{\partial x}\right) \quad (8.36)$$

Let us add the equation 8.35b divided by $\overline{\rho}$ to the last equation 8.36, so that we obtain the following relation:

$$\frac{\partial}{\partial x}\left(\left[k_g + \frac{\omega}{\overline{\rho}}k_l\right]\frac{\partial p}{\partial x}\right) = \left(1-\frac{1}{\overline{\rho}}\right)\frac{1}{Pe}\frac{\partial}{\partial x}\left(\frac{1-s}{p}\frac{\partial c_g^{(N)}}{\partial x} + \sigma\omega s\frac{\partial c_l^{(N)}}{\partial x}\right)$$

$$-\left(1-\frac{1}{\overline{\rho}}\right)\frac{\omega}{Ca}\frac{\partial}{\partial x}\left(k_l\frac{\partial p_c}{\partial x}\right)$$

In the case when diffusion and capillary effects can be neglected we obtain the following pseudo-stationary equation for pressure:

$$\frac{\partial}{\partial x}\left(\left[k_g + \frac{\omega}{\overline{\rho}}k_l\right]\frac{\partial p}{\partial x}\right) = 0 \quad (8.37)$$

Finally, substituting the term $\dfrac{\partial p}{\partial x}$ form the equation 8.37 to 8.36 we obtain:

$$\frac{\partial s}{\partial \tau} - \frac{\omega q}{2\overline{\rho}^2}\frac{\partial}{\partial x}\left(\frac{k_l}{k_g + \frac{\omega}{\overline{\rho}}k_l}\right) = 0 \quad (8.38)$$

This means that saturation can be presented in form of a Buckley-Leverett equation with flow velocity $U = \dfrac{\omega q}{2\bar{\rho}^2}$ and pseudo-fractional flow function $F = k_l/(k_g + \dfrac{\omega}{\bar{\rho}}k_l)$. Thus, the solution to the hydrogen-water two-phase problem can be presented in analytical form for a non-compressible flow without capillarity end diffusion.

As a first attempt of analysis we propose to compare the numerical simulation to the analytical solution (equation 8.38), for the case of a non-compressible flow without mass transfer between the phases. Illustrations of several time instances for low mobile gas and mobile gas are presented in Figure 8.5. Thus, the analytical approach match very well the numerical one.

8.4.3. Full concentration approach and pseudo saturation

Such type of flow as gas penetration in liquid requires a non-trivial analysis. One of the most important questions is how to define the separation of the one-phase domain from the two-phase domain. Initially, the system contains one single phase (liquid phase), and in the analyzed problem this liquid phase mainly consists of water. Actually, if the liquid saturation tends to 1 (not exactly 1), then it is formally considered as a two-phase system, while physically it can be a two-phase system as well as a one-phase system. The question is how to define this separation, i.e. what is the numerical error of calculation. This problem is very relevant for the penetration of mobile gas, especially in the simple case of a non-compressible flow (see the left plot in Figure 8.5). Further questions are how to interpret the fact that the asymptotical curve tends to 0 for gas saturation. and whether it is a physical reality or a numerical error. To answer these questions the classical approach for the definition of saturation is not sufficient. In this chapter we propose to introduce a pseudo-saturation function into the system. To reach this goal we will first introduce the notion of full concentration of hydrogen:

$$c^h = \frac{\rho_l c_l^h s + \rho_g c_g^h (1-s)}{\rho_l s + \rho_g (1-s)} \tag{8.39}$$

Hence, the saturation for the liquid phase can be written as:

$$s = \frac{\rho_g (c_g^h - c^h)}{\rho_l (c_l^h - c^h) + \rho_g (c_g^h - c^h)} \tag{8.40}$$

If for a fixed pressure the full concentration of hydrogen c^h is less than the concentration of hydrogen in the liquid phase c_l^h then the system is a one-phase system (liquid phase).
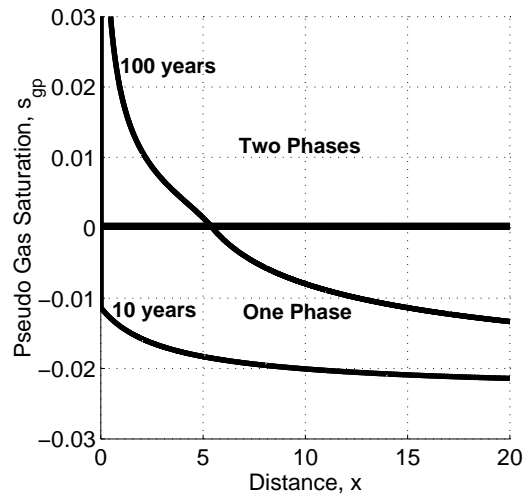
Figure 8.6.: Pseudo-saturation of gas: one phase and two-phase domains

The situation is similar when $c^h > c_g^h$, i.e. the system is a one-phase system (gas phase). The system is in two-phase state for a fixed pressure when the following condition is valid: $c_l^h < c^h < c_g^h$. However, according to relation 8.40 the saturation s is outside of the domain $[0:1]$ if the full concentration of hydrogen c^h is in a non-equilibrium state, i.e. outside of the domain $[c_l^h : c_g^h]$. This gives us the opportunity to introduce the new notion of pseudo-saturation s_p which has the same value as the saturation s in the two-phase domain $[0:1]$, but can also be less than 0 and more than 1 in the one-phase domain (see Figure 8.6). Physically, notions of full concentration and pseudo-saturation indicate clearly the state of the system, i.e. indicate how far the one-phase system is situated from the two-phase state. Thus, this approach defines exactly the separation of the one-phase domain from the two-phase domain, which is very tipical for mobile gas penetration (Figure 8.6).

8.4.4. Gas penetration in liquid through a porous medium

The general compositional flow model of gas penetration in liquid through a porous medium was analyzed numerically (Fortran Code). Using the notion of pseudo-saturation the influence of diffusion on mobile gas penetration in liquid was analyzed. The typical evolution of gas saturation (pseudo-saturation) for a hydrogen-water flow around a radioactive storage site was presented in Figure 8.7. The influence of diffusion increases gas penetration with time, for example gas penetration can exceed the value of 1 meter in 200 years (especially in a low saturated domain). Even without diffusion pure gas penetration is still fast for low
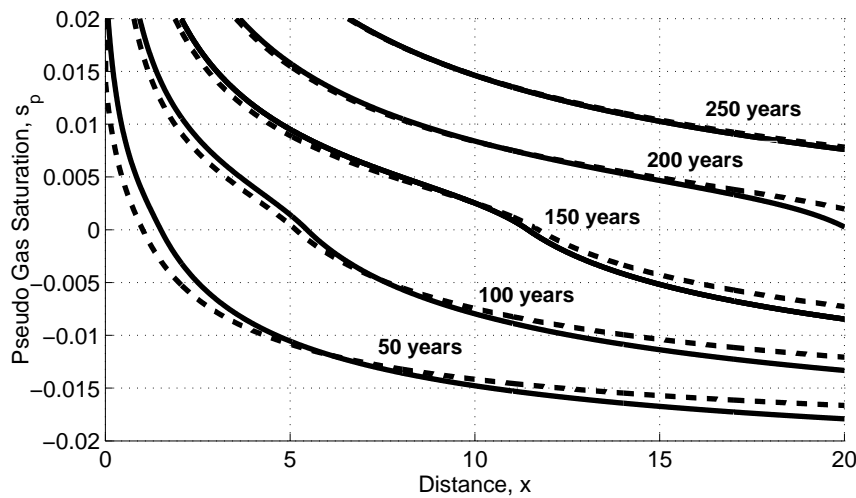
Figure 8.7.: Evolution of gas saturation for mobile gas: solid line - without diffusion, dashed
            line - with diffusion

mobile gas. In the case of low mobile gas penetration in liquid through a porous medium
the flow has a clear front, displacing very slowly.Evidently, diffusion is one of the essential
drives of the flow for a large range of saturation.

The mass exchange phenomenon has a small influence on the flow type but essential influ-
ence on the phase properties (phase compressibilities). The impact of phase transition can
be illustrated by the difference between two saturations: saturation with a diffusion effect
and saturation without this effect, see Figure 8.8. It is necessary to note that the influence of
phase transitions was also introduced non-directly, actually through the diffusion term. Thus,
we can conclude that influence only of the direct mass exchange phenomenon has insignifi-
cant impact on gas displacement in liquid of not more than one millimetre for the presented
example in Figure 8.8.

8.5. Conclusions

Hydrogen penetration in water through a porous medium was analysed in this chapter. A
two-phase hydrogen-water compositional model approach was proposed. The present work
consists of two principal parts:

The first part deals with the thermodynamic analysis of a hydrogen-water system. The ther-
modynamic model was calibrated using experimental data of hydrogen solubility in water.
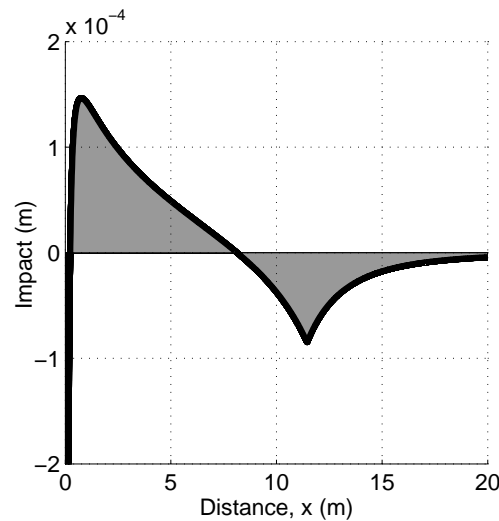
Figure 8.8.: Phase exchange influence: compressibility impact on the displacement of mobile
           gas (simulation time is 150 years)

The phase densities, viscosities and concentrations were presented in analytical form. More-
over, the domain of validity of simplified analytical laws for the estimation of phase proper-
ties (such as Henry's, Raoult's and Kelvin's laws) was presented for the analysed system.

The second part of the work deals with hydrogen-water two-phase hydrodynamic behavior.
An analytical solution for a non-compressible flow was constructed. An approach based
on the full concentration and the pseudo-saturation was introduced to exactly define the
separation of the one-phase domain from the two-phase domain. The general compositional
flow model of gas penetration in liquid through a porous medium was analysed numerically.
The influence of relative permeabilities on the flow regimes was analysed. Two principal
models were used: the mobile gas model and the low mobile gas model. Low mobile gas
creates a front which displaces with time, but evidently not as fast as in the case of mobile
gas flow. In both cases the influence of diffusion increases with time. However, diffusion
becomes very important for the low mobile gas case as the penetration accelerates for a large
range of saturation. In contrast to this, the direct mass exchange phenomenon has a small
influence on the flow type. Thus, the regimes of hydrogen penetration in liquid were shown
as very sensitive to the forms of relative permeability.

Part III.

Efficient model reduction methodology for uncertainty quantification

Many environmental systems, such as subsurface multiphase flow, ground water flow and carbon dioxide storage, are plagued in modeling by the ubiquitous presence of uncertainty. Lacking information about their system properties leads to model uncertainties up to a level where quantification of uncertainties may become the dominant question in modeling, simulation and application tasks. The prediction of environmental systems in large-scale vastly depends on our ability to quantify such uncertainties and related risks. Current numerical simulation models are often too expensive for advanced application tasks that involve accurate uncertainty quantification, risk assessment, robust design and model calibration. Even single deterministic simulations often require parallel high performance computing. Because the involved complex flow processes in porous media have a significantly non-linear character, the problem is too non-linear for quasi-linear and other simplified stochastic tools.

The current Part III deals with the development of approaches for these advanced challenges based on massive stochastic model reduction techniques. The reduction is achieved via projections on orthonormal polynomial bases which form a so-called response surface. This way, the model response to changes in uncertain parameters is represented by multivariate polynomials for each output quantity of interest, which allows for non-linear model-based propagation of parameter uncertainties onto the predicted quantities of interest. This technique is known as polynomial chaos expansion in the field of stochastic partial differential equations. The reduced model represented by the response surface is vastly faster than the original complex one, and thus provides a promising starting point for follow-up tasks: global sensitivity analysis, uncertainty quantification, robust design, probabilistic risk assessment as well as model calibration. For many cases, the fact that the response surface has known polynomial properties will help to further simplify these tasks. Thus, stochastic model reduction provides an opportunity to quantify uncertainty in complex non-linear and dynamic systems at acceptable costs.

Chapter 9 focuses on the arbitrary polynomial chaos which can adapt to arbitrary probability distribution shapes of input parameters. It provides improved convergence in comparison to classical chaos expansion techniques. Chapter 10 provides a flexible and efficient framework for global sensitivity analysis in order to quantify the effects of parameter variation or parameter uncertainty on the overall model uncertainty. It can convey the information of a global sensitivity analysis at computational costs that are hardly larger than those for local analysis. Chapter 11 proposes an advanced framework for model calibration and history matching based on the arbitrary polynomial chaos expansion and strict Bayesian principles. Through this combination, we obtain a statistical method for history matching that is accurate, yet has a computational speed that is more than sufficient for real-time application.

9. Uncertainty quantification using the arbitrary polynomial chaos

Bibliographic Note: The content of this chapter is based on the following original article: Oladyshkin S. and Nowak W. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliability Engineering & System Safety, Elsevier, V. 106, P. 179-190, 2012.

We discuss the arbitrary polynomial chaos (aPC), which has been subject of research in a few recent theoretical papers. Like all polynomial chaos expansion techniques, aPC approximates the dependence of simulation model output on model parameters by expansion in an orthogonal polynomial basis. The aPC generalizes chaos expansion techniques towards arbitrary distributions with arbitrary probability measures, which can be either discrete, continuous, or discretized continuous and can be specified either analytically (as probability density/cumulative distribution functions), numerically as histogram or as raw data sets. We show that the aPC at finite expansion order only demands the existence of a finite number of moments and does not require the complete knowledge or even existence of a probability density function. This avoids the necessity to assign parametric probability distributions that are not sufficiently supported by limited available data. Alternatively, it allows modellers to choose freely of technical constraints the shapes of their statistical assumptions. Our key idea is to align the complexity level and order of analysis with the reliability and detail level of statistical information on the input parameters. We provide conditions for existence and clarify the relation of the aPC to statistical moments of model parameters. We test the performance of the aPC with diverse statistical distributions and with raw data. In these exemplary test cases, we illustrate the convergence with increasing expansion order and, for the first time, with increasing reliability level of statistical input information. Our results indicate that the aPC shows an exponential convergence rate and converges faster than classical polynomial chaos expansion techniques.

9.1. Introduction

The lack of information about the properties of physical systems, such as material parameters or boundary values, can lead to model uncertainties up to a level where the quantification of prediction uncertainties may become the dominant question in application tasks. Most physical processes appearing in nature are non-linear and, as a consequence, the required mathematical models are non-linear. Traditional and very well-known approaches for stochastic simulation are brute-force Monte Carlo simulation (e.g. (Maltz and Hitzl, 1979)) and related approaches (e.g. latin hypercube sampling (Helton and Davis, 2003)). Unfortunately, for large and complex models, Monte Carlo techniques are inadequate. Even single deterministic simulations may require parallel high-performance computing. As a reasonably fast and attractive alternative, stochastic model reduction techniques based on the polynomial chaos expansion can be applied.

9.1.1. Polynomial chaos expansion

A large number of studies for diverse applications is based on the polynomial chaos expansion (PCE) introduced by Wiener (Wiener, 1938) in 1938. The chaos expansion offers an efficient high-order accurate way of including non-linear effects in stochastic analysis. PCE can be seen, intuitively, as a mathematically optimal way to construct and obtain a model response surface in the form of a high-dimensional polynomial in uncertain model parameters. The chances and limitations of polynomial chaos and related expansion techniques were discussed in (Augustin et al., 2008). The paper (Oladyshkin et al., 2011b) showed how to use PCE for robust design under uncertainty with controlled failure probability. Recently, the sensitivity analysis based on PCE decomposition (Buzzard, 2011), (Crestaux et al., 2009), (Haro Sandoval et al., 2012) has received increased attention. The papers (Sudret, 2008) and (Oladyshkin et al., 2011d) demonstrate correspondingly how classical PCE and its new aPC version can deliver the information required for global sensitivity analysis at low computational costs. Also, FORM and SORM methods (e.g. (Jang et al., 1994)) could be extended to higher order accuracy via PCE, however, this has not yet been achieved.

The PCE technique can mainly be sub-divided into intrusive and non-intrusive approaches for the involved projection integral. The intrusive approach requires manipulation of the governing equations and can sometimes provide semi-analytical solutions for stochastic analysis. The best-known method from this group is the stochastic Galerkin technique, which

originated from structural mechanics (Ghanem and Spanos, 1991) and has been applied in studies for modeling uncertainties in flow problems ((Ghanem and Spanos, 1993), (Matthies and Keese., 2005), (Xiu and Karniadakis, 2003)). However, because of the necessary symbolic manipulations, the procedure may become very complex and analytically cumbersome. For that reason, non-intrusive approaches like sparse quadrature (Keese and Matthies, 2003) and the probabilistic collocation method ((Isukapalli et al., 1998), (Li and Zhang, 2007)) have been receiving increasing attention.

9.1.2. Polynomial chaos expansion for non-Gaussian distributions

The original PCE is based on Hermite polynomials, which are optimal for normally distributed random variables. Unfortunately, natural phenomena and uncertainty in engineering are often not that simple, and the distribution of physical or model parameters often cannot be considered Gaussian. However, it is possible to put into conformity a physical variable with a normal variable by an adequate transformation called Gaussian anamorphosis or normal score transformation (e.g. (Wackernagel, 1998)) or approximate parametric transformations (Ditlevsen and Madsen, 1992). Using transformed variables for expansion cannot be considered an optimal choice because it leads to slow convergence of the expansion (e.g. (Xiu and Karniadakis, 2002a), (Xiu and Karniadakis, 2003)). In recent years, the PCE technique has been extended to the generalized polynomial chaos (gPC), based on the Askey scheme (Askey and Wilson, 1985) of orthogonal polynomials by (Xiu and Karniadakis, 2002a) and (Xiu and Karniadakis, 2003). The gPC extends PCE towards a counted number of parametric statistical distributions (Gamma, Beta, Uniform, etc.). However, application tasks demand further adaptation of the chaos expansion technique to a larger spectrum of distributions. In (Wan and Karniadakis, 2006) and (Prempraneerach et al., 2010), the authors presented a multi-element generalized polynomial chaos (ME-gPC) method. It is based on a decomposition of the random space into local elements, and subsequently implements gPC locally within the individual elements. An error control theory for the ME-gPC method was developed in (Wan and Karniadakis, 2009) for elliptic problems. The ME-gPC is the first adaptive piece-wise approach helping to deal with discontinuity of distributions or of model responses, and provides the desired adaptation to a wide spectrum of distributions. The ME-gPC conception offered in (Wan and Karniadakis, 2006) provides a flexible tool for stochastic modeling, but interprets these data as an exactly known probability distribution,

and considerably increases the computational effort for multidimensional stochastic problems.

9.1.3. Limited availability of data

The methods discussed above assume an exact knowledge of the involved probability density functions. Unfortunately, information about the distribution of data is very limited in realistic engineering applications, especially when environmental influences or natural phenomena are involved, or when predicting or engineering the environment (see also (Red-Horse and Benjamin, 2004)). Applied research on (partially) natural or complex realistic systems often faces the problem of limited information about the model parameters and even about their probability distributions. For example, material properties of underground reservoirs are insufficiently available to provide a full picture of their distribution. Moreover, the statistical distribution of model parameters can be nontrivial, e.g., bounded, skewed, multi-modal, discontinuous, etc. Also, the dependence between several uncertain input parameters might be unknown, compare (Der Kiureghian and Liu, 1986). Depending on the modeling task and circumstances, statistical information on model parameters may be available either discrete, continuous, or discretized continuous, they could exist analytically as PDF/CDF or numerically as histogram. The key shortcoming of current PCE approaches in this context is twofold. First, they are heavily restricted in handling most of these conditions, and second they assume that this information is complete and perfect.

Small samples or data sets do not contain perfect or complete information on the probability distribution of model input parameters. For example, the study (Red-Horse and Benjamin, 2004) demonstrated that limited information on input statistics introduces its own type of uncertainty in quantifying statistical model output distribution. Also, any attempt to construct probability density functions of any particular shape from samples of limited size or from sparse information introduces additional subjectivity into the analysis, which bears the severe risk of leading to biased results. In a related application study (Oladyshkin et al., 2011a), we illustrate that errors or additional (and mostly subjective) assumptions in data interpretation can severely bias uncertainty quantification and risk assessment, and hence could lead to failing designs. Methods of maximum entropy (Jaynes, 1982), closely related to known as the exponential polynomial method (Er, 1998) in reliability engineering, and minimum relative entropy (Woodbury and Ulrych, 1993) are often used in the engineering sciences to construct a probability distribution from sparse information (mostly in the form of a few statistical moments and bounds) that may be available from different instances of

the same object or from different objects with supposedly similar properties or conditions. Although these two methods are designed to minimize subjectivity and even though they can preserve the sample moments up to arbitrary order, they are heavily debated within the statistical community (e.g. (Sambucini, 2007)). In fact, they still introduce new assumptions and impose a specific assumption on distribution shape. The same is true for other typical methods to construct PDFs from moments in the field of reliability engineering, such as the Hermite polynomial transformation (Winterstein, 1988). Such methods, however, are more subjective than entropy-based methods, since they cannot keep the original sample moments up to higher orders unchanged. If one still desires to fit a PDF as a pragmatic tool to filter raw data against noise, one should have full freedom in the chosen distribution shapes, not restricted by the technical constraints of PCE or gPC.

9.1.4. Approach and novelties

To overcome the first part of the problem, we claim that it is not even necessary to cast the available statistical information into probability density functions. Instead, the available information can directly and most purely be used in stochastic analysis, when using our data-driven formulation of PCE, see section 9.2. We argue that applied tasks demand direct handling of arbitrary data distributions without additional assumptions for stochastic analysis. To overcome the second part of the problem, we suggest to perform a robustness analysis around the PCE (here: aPC) to assess the impact of incomplete statistical input information. Overall, we propose to align the complexity level and order of analysis with the reliability and detail level of statistical information on the input parameters.

The concept we propose in the current chapter is to approach the problem in a highly parsimonic and yet fully data-driven description of randomness. We draw attention to the arbitrary polynomial chaos (aPC) that has recently been touched upon in a few theoretical papers. Two studies focusing on proofs of existence were published in the mathematical stochastics community ((Ernst et al., 2012), (Soize and Ghanem, 2004)). Constructing the aPC polynomials by Gram-Schmidt orthogonalization was presented in the field of aerospace engineering ((Witteveen and Bijl, 2006), (Witteveen et al., 2007)). These studies did not discuss the aPC in the light of data availability, limited reliability of data and assumptions in data interpretation.

The aPC extends chaos expansion techniques by employing a global polynomial basis for arbitrary distributions of data. In a certain sense, it allows to return back to a global basis with

the new freedom of arbitrary polynomial chaos that the ME-gPC ((Wan and Karniadakis, 2006) and (Wan and Karniadakis, 2009)) uses only within piecewise local elements.

The most important property of the aPC that we will install and exploit in the current chapter is that the aPC can work with probability measures that may be (if necessary) implicitly and incompletely defined via their moments only, and that it requires no additional information. In fact, our equations will show explicitly (in closed form) that statistical moments are the only source of information that is propagated in all polynomial expansion-based stochastic approaches. Thus, exact probability density functions do not have to be known and do not even have to exist. For finite-order expansion, only a finite number of moments has to be known. This opens the path to data-driven applications, where data samples with limited size merely allow inference of a few statistical moments, but are not sufficient to support, without a substantial degree of subjectivity, the assumption of an exactly known probability measure (see discussion in section 9.4.1). Fully in line with the demands of application tasks, the statistical data of modeling parameters can be specified either analytically (as probability density/cumulative distribution functions), numerically as histogram as a raw data sets.

In section 9.2, we deliver the necessary mathematical material, and provide the necessary properties and proofs in section 9.3. The convergence rate of the aPC will be illustrated in the context of an example problem in section 9.4. How to address issues arising from the incomplete and inaccurate character of raw data sets used as statistical input information is discussed and illustrated in section 9.5.

9.2. The arbitrary polynomial chaos expansion

9.2.1. One-dimensional aPC

We will consider a stochastic process in the probability space (Ω, A, Γ) with space of events Ω, σ-algebra A and probability measure Γ, see e.g. (Grigoriu, 2002). Let us consider a stochastic model $Y = f(\xi)$ with model input $\xi \in \Omega$ and model output Y. For a stochastic analysis of Y, the model $f(\xi)$ may by expanded as following:

$$Y(\xi) \approx \sum_{i=1}^{d} c_i P^{(i)}(\xi), \tag{9.1}$$

where d is the order of expansion, c_i are the expansion coefficients that are determined by Galerkin projection, numerical integration or collocation, and $P^{(i)}(\xi)$ are the polynomials

forming the basis $\left\{ P^{(0)}, \ldots, P^{(d)} \right\}$ that is orthogonal (or even orthonormal) with respect to the measure Γ (see equation 9.6). The only difference between aPC and previous PCE methods is that the measure Γ can have an arbitrary form, and thus the basis $\left\{ P^{(0)}, \ldots, P^{(d)} \right\}$ has to be found specifically for the probability measure Γ appearing in the respective application.

This open the path to data-driven applications of aPC. If a function $Y(\xi)$ is expanded in the orthonormal polynomial basis $\left\{ P^{(0)}, \ldots, P^{(d)} \right\}$, then characteristic statistical quantities of $Y(\xi)$ can be evaluated directly from the expansion coefficients c_i. For example, the mean and variance of $Y(\xi)$ is given by the following simple analytical relations:

$$\mu_Y = c_1, \quad \sigma_Y^2 = \sum_{i=2}^{N} c_i^2. \tag{9.2}$$

Notice that, in the current chapter, we will focus on mono-dimensional stochastic input (i.e., only one uncertain parameter) for simplicity, but without loss of generality (see section 9.2.2).

9.2.2. Multi-dimensional aPC

Most realistic applications feature multi-dimensional model input ξ, i.e. $\xi = \{\xi_1, \xi_2, \ldots, \xi_N\}$. Here, the total number of input parameters is equal to N. The model parameters can be design or control parameters that can be chosen by the operator of a system, and uncertain parameters that describe our (incomplete) knowledge of the system properties. Hence, to investigate the influence of all input parameters $\xi_1, \xi_2, \ldots$ on the model output Y, the model output Y can be represented by a multivariate polynomial expansion as follows

$$Y(\xi_1, \xi_2, \ldots, \xi_N) \approx \sum_{i=1}^{M} c_i \Phi_i(\xi_1, \xi_2, \ldots, \xi_N) \tag{9.3}$$

Here, the coefficients c_i quantify the dependence of the model output Y on the input parameters $\xi_1, \xi_2, \ldots, \xi_N$. The number M of terms in the expansion (9.3) depends on the total number of input parameters N and on the order d of the expansion, according to the combinatory formula $M = (N+d)!/(N!d!)$. The function Φ_i is a simplified notation of the multi-variate orthogonal polynomial basis for $\xi_1, \xi_2, \ldots, \xi_N$. Assuming that the input parameters within $\xi_1, \xi_2, \ldots, \xi_N$, are independent (e.g. (Ghanem and Spanos, 1991)), the multi-dimensional

basis can be constructed as a simple product of the corresponding univariate polynomials:

$$\Phi_i(\xi_1, \xi_2, \ldots, \xi_N) = \prod_{j=1}^{N} P_j^{(\alpha_j^i)}(\xi_1, \xi_2, \ldots, \xi_N), \qquad (9.4)$$

$$\sum_{j=1}^{N} \alpha_j^i \leq M, \quad i = 1 \ldots N,$$

where α_j^i is a multivariate index that contains the combinatoric information how to enumerate all possible products of individual univariate basis functions. In other words, the index α can be seen as $M \times N$ matrix, which contains the corresponding degree (e.g. 0, 1, 2, etc.) for parameter number j in expansion term k.

Let us mention that, in the current state of science for polynomial chaos expansions, the random variables have to be statistically independent or may be correlated in a linear fashion only. Linear correlation can be removed by adequate linear transformation, such as the KL-expansion (Li and Zhang, 2007), also called proper orthogonal decomposition (Lumley, 1967) or principal component analysis (Pearson, 1901) in other disciplines. Construction of a joint polynomial basis for statistically dependent random variables beyond linear dependence is a very important issue for future research.

9.2.3. Stochastic analysis based on the aPC

Equations (9.1) and (9.3) can be interpreted as a model response surface for $Y = f(\xi_1, \xi_2, \ldots, \xi_N)$, and represent the basic key element for: (I) uncertainty quantification; (II) robust design and (III) global sensitivity analysis.

(I) The simplest way to quantify uncertainty is via the analytical relations, see equation (9.2). However, in order to evaluate more complex statistical quantities, Monte Carlo simulation can be performed directly and immensely fast on the obtained polynomial given by equation (9.3) (see e.g. (Oladyshkin et al., 2011a), (Walter et al., 2012)). For Monte Carlo simulation in absence of precise statistical information, (Oladyshkin et al., 2011a) discuss and suggest the maximum entropy method for PDF estimation.

(II) Including design and control parameters together with uncertain parameters in expansion (9.3) provides an effective basis for robust design. The paper (Oladyshkin et al., 2011b) showed how to use PCE for robust design under uncertainty with controlled failure probability.

(III) Expansion (9.3) also delivers the information required for global sensitivity analysis including simultaneous influences of different modeling parameters at low computational costs. For example, Sobol indices (Sobol, 1990, 2001) or Weighted indices (Oladyshkin et al., 2011d) can be computed directly from the coefficients α_i.

9.3. Moment-based analysis

Let us define the polynomial $P^{(k)}(\xi)$ of degree k in the random variable $\xi \in \Omega$:

$$P^{(k)}(\xi) = \sum_{i=0}^{k} p_i^{(k)} \xi^i, \quad k = \overline{0,d}, \qquad (9.5)$$

where $p_i^{(k)}$ are coefficients in $P^{(k)}(\xi)$.

Our goal is to construct the polynomials in equation (9.5) such that they form an orthonormal basis for arbitrary distributions. The arbitrary distributions for the framework presented in this chapter can be either discrete, continuous, discretized continuous, specified analytically, as histograms, raw data sets or by their moments. In this chapter, we exploit this freedom and show how to treat any given probability distribution solely defined by the statistical moments of ξ. For limited-order expansion, this allows to work with arbitrary probability measures that are implicitly and incompletely defined by a limited number of moments only.

The goals of this section are to: (1) Derive a constructive rule to obtain the orthonormal basis which clarifies that only moments of ξ are important, (2) show that finite-order expansion only requires a finite number of moments, (3) provide conditions for the existence of an aPC even for sampled data.

9.3.1. Constructing the aPC from moments

Orthonormality for polynomials $P^{(k)}$ of degree k and $P^{(l)}$ of degree l is defined as:

$$\int_{\xi \in \Omega} P^{(k)}(\xi) P^{(l)}(\xi) d\Gamma(\xi) = \delta_{kl}, \quad \forall k, l = \overline{0,d} \qquad (9.6)$$

where δ_{kl} is the Kronecker delta. For the further development, we will make use of only the orthogonality condition:

$$\int_{\xi \in \Omega} P^{(k)}(\xi) P^{(l)}(\xi) d\Gamma(\xi) = 0, \quad \forall k \neq l. \qquad (9.7)$$

Instead of the normality condition, we will at first introduce an intermediate auxiliary condition by demanding that the leading coefficients of all polynomials be equal to 1:

$$p_k^{(k)} = 1, \ \forall k. \tag{9.8}$$

The following conditions apply for the orthogonal polynomial basis $\left\{ P^{(k)}(\xi) \right\}$ $(k = \overline{0,d})$. First, for the zero-degree polynomial $P^{(0)}$, we obtain directly from (9.7) and (9.8) that $p_0^{(0)} = 1$, which also satisfies the normality condition (9.6). The orthogonality conditions for $P^{(1)}$ are as follows:

$$\int_{\xi \in \Omega} p_0^{(0)} \left[\sum_{i=0}^{1} p_i^{(1)} \xi^i \right] d\Gamma(\xi) = 0;$$

$$p_1^{(1)} = 1 \tag{9.9}$$

This procedure can be continued for the construction of all following polynomials to obtain an orthogonal basis. The generalized conditions of orthogonality for any polynomial $P^{(k)}$ of degree k with all lower-order polynomials can be written in the following form:

$$\int_{\xi \in \Omega} p_0^{(0)} \left[\sum_{i=0}^{k} p_i^{(k)} \xi^i \right] d\Gamma(\xi) = 0;$$

$$\int_{\xi \in \Omega} \left[\sum_{i=0}^{1} p_i^{(1)} \xi^i \right] \left[\sum_{i=0}^{k} p_i^{(k)} \xi^i \right] d\Gamma(\xi) = 0;$$

$$\cdots \tag{9.10}$$

$$\int_{\xi \in \Omega} \left[\sum_{i=0}^{k-1} p_i^{(k-1)} \xi^i \right] \left[\sum_{i=0}^{k} p_i^{(k)} \xi^i \right] d\Gamma(\xi) = 0;$$

$$p_k^{(k)} = 1.$$

The system of equations given by (9.10) is closed and defines the unknown polynomial coefficients $p_i^{(k)}$ $(i = \overline{0,k})$ of the required basis. Obviously, the above definition of the orthogonal polynomial of degree k uses the definition of all polynomials of lower degrees $0, \ldots, k-1$. We will use that particular property to simplify the system in equation (9.10) by substituting the first equation into the second, the first and the second into the third, and so on. In addition, we will apply condition (9.8). Hence, without loss of generality, the system in equation

(9.10) can be reduced to:

$$\int_{\xi \in \Omega} \sum_{i=0}^{k} p_i^{(k)} \xi^i d\Gamma(\xi) = 0;$$

$$\int_{\xi \in \Omega} \sum_{i=0}^{k} p_i^{(k)} \xi^{i+1} d\Gamma(\xi) = 0;$$

$$\dots \qquad\qquad (9.11)$$

$$\int_{\xi \in \Omega} \sum_{i=0}^{k} p_i^{(k)} \xi^{i+k-1} d\Gamma(\xi) = 0;$$

$$p_k^{(k)} = 1.$$

Note that this rearrangement defines the kth orthogonal polynomial independent of all other polynomials from the orthogonal basis. The kth raw moment of the random variable ξ is defined as:

$$\mu_k = \int_{\xi \in \Omega} \xi^k d\Gamma(\xi). \qquad\qquad (9.12)$$

This allows to re-write equation (9.11) based on only the raw moments of ξ:

$$\sum_{i=0}^{k} p_i^{(k)} \mu_i = 0;$$

$$\sum_{i=0}^{k} p_i^{(k)} \mu_{i+1} = 0;$$

$$\dots \qquad\qquad (9.13)$$

$$\sum_{i=0}^{k} p_i^{(k)} \mu_{i+k-1} = 0;$$

$$p_k^{(k)} = 1.$$

Alternatively, the system of linear equations (9.13) can be written in the more convenient matrix form:

$$
\begin{bmatrix}
\mu_0 & \mu_1 & \dots & \mu_k \\
\mu_1 & \mu_2 & \dots & \mu_{k+1} \\
\dots & \dots & \dots & \dots \\
\mu_{k-1} & \mu_k & \dots & \mu_{2k-1} \\
0 & 0 & \dots & 1
\end{bmatrix}
\begin{bmatrix}
p_0^{(k)} \\
p_1^{(k)} \\
\dots \\
p_{k-1}^{(k)} \\
p_k^{(k)}
\end{bmatrix}
=
\begin{bmatrix}
0 \\
0 \\
\dots \\
0 \\
1
\end{bmatrix}. \qquad\qquad (9.14)
$$

As a direct consequence, an orthogonal polynomial basis up to order d can be constructively defined for any arbitrary probability measure Γ under the following conditions: the coefficients $p_i^{(k)}$ can be constructed if and only if the square matrix of moments in the left-hand side of equation (9.14) is not singular. In the Section 9.3.2, we provide a proof for this under the condition that the number of support points in the distribution of ξ is greater than k and that all moments up to order $2k-1$ are finite. This holds for all continuous random variables, under the condition that $2k-1$ moments exist. If the moments of ξ are evaluated directly from a data set of limited size or from a discrete probability distribution featuring a finite number of possible outcomes, there need to be k or more distinct values in the data set or distribution. All moments are always finite if no element of the data set is infinite.

From the equations (9.7) to (9.14), it becomes evident that moments are the only required form of information on input distributions for constructing the basis and thus to operate the aPC. For finite-order expansion, a finite number of moments is sufficient. Hence, if a raw data set is the only form of available input information, computing its moments is sufficient, and estimating a full PDF from the data is not necessary. The same is true if only a limited number of moments are provided as input characterization. Also, arbitrary parametric distribution can be addressed, simply by working with their moments. This implies that any difference between distributions that becomes visible only in moments of order higher than $2d-1$ will be invisible to any order d polynomial expansion technique.

9.3.2. Non-singularity of the moments matrix

Let us write the square matrix of equation (9.14) in the following decomposed form:

$$\mathbf{M} = \begin{bmatrix} \mathbf{H} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \tag{9.15}$$

where

$$\mathbf{H} = \begin{bmatrix} \mu_0 & \dots & \mu_{k-1} \\ \dots & \dots & \dots \\ \mu_{k-1} & \dots & \mu_{2k-2} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mu_k \\ \dots \\ \mu_{2k-1} \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 0 & \dots & 0 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1 \end{bmatrix}$$

Evidently, $\mathbf{D}$ is always invertible, and hence the determinant of $\mathbf{M}$ is given by:

$$det(\mathbf{M}) = det(\mathbf{D})det(\mathbf{H} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}) \tag{9.16}$$

Because $det(\mathbf{D}) = 1$ and $\mathbf{C} = [0, \ldots, 0]$, we obtain:

$$det(\mathbf{M}) = det(\mathbf{H}) \qquad (9.17)$$

The matrix $\mathbf{H}$ is also known as the Hankel matrix of moments. The properties of its determinant were studied in the paper (Lindsay, 1989). Moreover, (Karlin, 1968) showed that $det(\mathbf{H})$ for $rank(\mathbf{H}) = k$ is zero *if and only if* the distribution of ξ has only k or fewer points of support. Thus, $\mathbf{M}$ is non-singular *if and only if* the number of support points in the distribution of ξ is greater than k and if all moments up to order $2k - 2$ are finite.

9.3.3. Explicit form of the data-driven polynomial chaos

In this section, we present an analytical explicit form of the coefficients for moderate degrees of polynomials, which can be easily used for diverse data-driven application tasks, such as uncertainty quantification, global sensitivity analysis and probabilistic risk assessment. The coefficients for the higher degrees of polynomials can be obtained using the implicit scheme presented above (9.14), via recursive relations (see Chapter 22 of book (Abramowitz and Stegun, 1965)), via Gram-Schmidt orthogonalization (see (Witteveen and Bijl, 2006), (Witteveen et al., 2007)) or via the Stieltjes procedure (Stieltjes, 1884).

To simplify the explicit form of coefficients, we will assume a normalized distribution of data with zero mean and unit variance after linear transformation:

$$\xi' = \frac{(\xi - \mu)}{\sigma}, \qquad (9.18)$$

which leads to a centralization and standardization of all moments.

Thus, the orthogonal polynomial basis $\left\{ P^{(k)}(\xi) \right\}$ $(k = \overline{0, d})$ can be presented as:

$$P^{(d)}(\xi) = \sum_{i=0}^{d} p_i^{(d)} \left(\frac{\xi - \mu}{\sigma} \right)^i, \qquad (9.19)$$

where $p_i^{(d)}$ are the coefficients of polynomial $P^{(d)}$ defined explicitly through the raw moments of ξ' from the relations below. Due to equation (9.18), the raw moments of ξ' are related to the central moments $\widehat{\mu}_k(\xi)$ of ξ via:

$$\mu_k(\xi') = \widehat{\mu}_k(\xi) \, \widehat{\mu}_2(\xi)^{-k/2}.$$

Coefficients for polynomial of 0 degree

$$p_0^{(0)} = 1; \tag{9.20}$$

Coefficients for polynomial of 1st degree

$$p_0^{(1)} = 0, \quad p_1^{(1)} = 1; \tag{9.21}$$

Coefficients for polynomial of 2nd degree

$$p_0^{(2)} = -1, \quad p_1^{(2)} = -\mu_3, \quad p_2^{(2)} = 1; \tag{9.22}$$

Coefficients for polynomial of 3rd degree

$$p_0^{(3)} = \mu_3^2 - \mu_3^3 + \mu_3\mu_4 - \mu_5;$$

$$p_1^{(3)} = -\mu_3\mu_5 + \mu_3^2 - \mu_4 + \mu_3\mu_4;$$

$$p_2^{(3)} = -\mu_3\mu_4 + \mu_5 - \mu_3;$$

$$p_3^{(3)} = 1 - \mu_3 + \mu_3^2; \tag{9.23}$$

Coefficients for polynomial of 4th degree

$$p_0^{(4)} = \mu_3^2\mu_5\mu_4 + \mu_3^3\mu_7 - \mu_3^2\mu_5^2 - 2\mu_3^2\mu_4\mu_6 + 2\mu_4^2\mu_5 - \mu_4^3;$$

$$p_1^{(4)} = -\mu_5^3 + \mu_4^2\mu_3^2 - \mu_3^2\mu_4\mu_5 + \mu_4\mu_5\mu_6 + \mu_3\mu_5\mu_6;$$

$$+ \mu_3\mu_5\mu_7 - \mu_3\mu_4\mu_7 - \mu_3^3\mu_5 + \mu_3^3\mu_6 - \mu_3\mu_6^2;$$

$$p_2^{(4)} = -\mu_4^2\mu_3^2 + \mu_3\mu_4\mu_7 + \mu_4\mu_5^2 - \mu_4^2\mu_6 + \mu_3^3\mu_5 - \mu_3\mu_5\mu_6;$$

$$p_3^{(4)} = \mu_3^3\mu_4 - \mu_5\mu_4 - \mu_3^2\mu_7 + \mu_5^2 + \mu_4\mu_6;$$

$$p_4^{(4)} = -\mu_3^4 + \mu_3^2\mu_6 - 2\mu_5\mu_4 + \mu_4^2; \tag{9.24}$$

9.3.4. Normalization

The above orthogonal polynomial basis can be used directly for analysis. However, an orthonormal basis has more useful properties (see equation (9.2) and section 9.3.5). Thus, the

next step is to normalize the orthogonal basis. We will use the norm for the polynomial P^k introduced in equation (9.6):

$$\left\| P^{(k)} \right\|^2 = \int_{\xi \in \Omega} \left[P^{(k)}(\xi) \right]^2 d\Gamma(\xi). \tag{9.25}$$

Hence, a valid orthonormal polynomial basis $\left\{ \Psi^{(k)}(\xi) \right\}$ $(k = \overline{0,d})$ is:

$$\Psi^{(k)}(\xi) = \frac{1}{\left\| P^{(k)} \right\|} \sum_{i=0}^{k} p_i^{(k)} \xi^i. \tag{9.26}$$

For normalization, the evaluation of $\left\| P^{(k)} \right\|$ for $k = d$ additionally requires finiteness and availability of the $2d$-th moment.

9.3.5. Summarized properties of the orthonormal basis

As consequence of the derivations in sections 9.3.1 and 9.3.4, the polynomial basis in equation (9.26) has the following properties:

Property I. The orthonormal basis can be constructed without any hierarchical conditions or recurrence relations that are used in Chapter 22 of (Abramowitz and Stegun, 1965) and in (Witteveen et al., 2007), (Witteveen and Bijl, 2006).

Property II. Existence of the moments $\mu_0, \ldots, \mu_{2d}$ is the necessary and sufficient condition for constructing an orthonormal basis $\left\{ \Psi^{(0)}, \ldots, \Psi^{(d)} \right\}$ up to degree d, together with the condition that the number of supports points of ξ is greater than d if ξ is a discrete variable or is represented by a data set.

Property III. The orthonormal polynomial basis for arbitrary probability measures is based on the corresponding moments only, and does not require the knowledge (or even existence) of a probability density function.

Property IV. All the zeros of the orthogonal polynomials are real, simple and located in the interior of the interval of orthogonality (Abramowitz and Stegun, 1965). This property is useful for numerical integration, especially for bounded distributions.

Property V. As particular cases, the Hermite, Laguerre, Jacoby polynomials, etc. from the Askey scheme and the polynomials for log-normal variables by Ernst at al. (Ernst et al., 2012) can be reconstructed within a multiplicative constant.

Property VI. All distributions that share the same moments up to order $2d$ will also share the same basis, and thus will lead to identical results in an expansion up to order d.

9.4. Data-driven modeling

The arbitrary polynomial chaos expansion presented in sections 9.2 and 9.3 provides a simple and efficient tool for analysing stochastic systems. We will consider a very simple model in order to focus all attention on our data-driven concept, which is based directly on the moments of sampled data without intermediate steps of data reinterpretation. This avoids the subjectivity usually introduced when choosing among a small limited number of theoretical distributions to represent a natural phenomenon, and so avoids the problems of subjectivity under limited data availability is discussed in section 9.1. These problems will be illustrated in section 9.4.1. An application to a problem with a realistic level of complexity and a detailed discussion of expert's subjectivity in uncertainty analysis is presented in (Oladyshkin et al., 2011a). That paper demonstrates how subjectivity of interpreting limited data sets can easily lead to substantial prediction bias, and that the subjective choice of distribution shapes has a similar relevance as uncertainties due to physical conceptualization, numerical codes and parameter uncertainty.

Here, for simplicity, we consider the exponential decay differential equation which was already used in (Xiu and Karniadakis, 2003) to illustrate the Askey scheme:

$$\frac{dY(t)}{dt} = -\xi Y, \quad Y(0) = 1 \tag{9.27}$$

Let Y_{PC} be the solution obtained using the polynomial chaos expansion (9.1) for the problem defined in equation (9.27). We use a Monte Carlo simulation as reference solution and define the time dependent relative error $\varepsilon(t)$ between the polynomial chaos expansion solution $Y_{PC}(t)$ and the Monte Carlo solution $Y_{MC}(t)$ as:

$$\varepsilon(t) = \frac{|Y_{PC}(t) - Y_{MC}(t)|}{|Y_{MC}(t)|}. \tag{9.28}$$

9.4.1. Fidelity of data-driven interpretation

To illustrate the fidelity of the data-driven chaos expansion, we will consider a synthetic example for an empirical data distribution, see Figure 9.1, and apply both aPC and the classical PCE for comparison. The illustrative data set (with sample size $N = 500$) presented in the left plot of Figure 9.1 was generated as superposition of normal and log-normal distributions and contains statistical noise introduced due to the small size of the data sample.
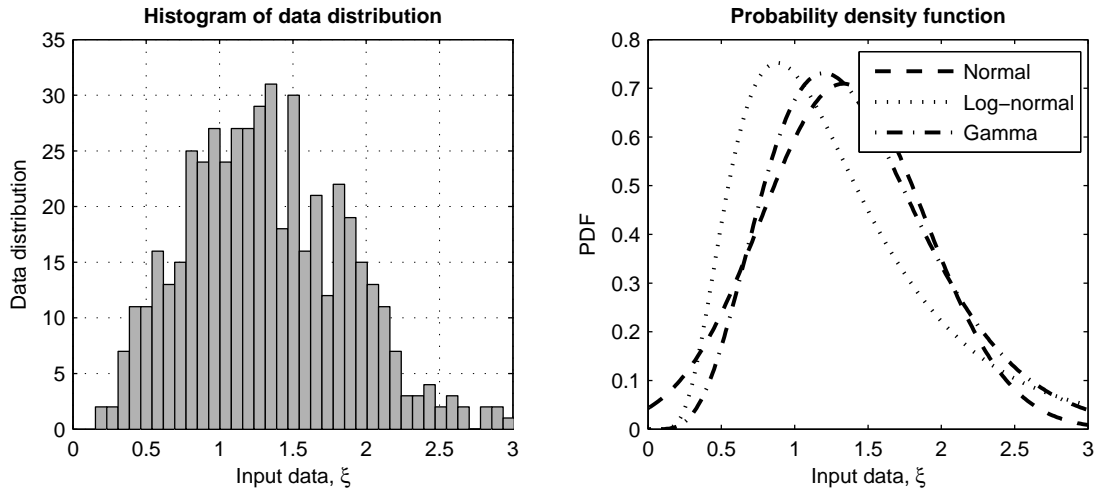
Figure 9.1.: Data distribution (left plot) and assumed stochastic distribution: Normal, Lognormal and Gamma (right plot)

The classical approach would be to introduce a parametric probability distribution, e.g., with fitted mean and variance or with maximum likelihood parameters. Here, for illustration, we select the Normal, Lognormal and Gamma distribution, see the right plot of Figure 9.1. Evidently, the list of possible candidate distributions for fitting to the considered data can be very long. Introducing a full probability density function (PDF) resembles a strong assumption on all higher moments up to infinite order, and claims to know the exact shape, e.g., also of the extreme value tails. Such assumptions on the alleged shape of the underlying probability density function, unfortunately, can lead to substantial errors in data interpretation. The data-driven approach strongly alleviates this situation, because it can directly handle a set of moments (e.g., the mean, variance, skewness, kurtosis, and so forth), without any further assumptions on higher-order moments and without having to introduce a PDF at all. This also provides the freedom to work with only a small number of moments obtained via expert elicitation, without asking for a full PDF.

We will apply different orders of the polynomial chaos expansion (9.1) to the test problem (9.27) using two sources of input information about the data distribution: (1) the three introduced assumptions on PDFs (right plot of Figure 9.1) and (2) the pure raw data sample (left plot of Figure 9.1). To observe the pure impact of data interpretation regardless of numerical techniques, we treat all four cases with the aPC, i.e., we construct an optimal orthonormal polynomial base (see section 9.2) for each case. This avoids the non-linear transformations that are usually necessary to map the assumed PDFs onto the normal PDF. Such transformations would introduce additional errors, as discussed in sections 9.4.2 and 9.4.3.
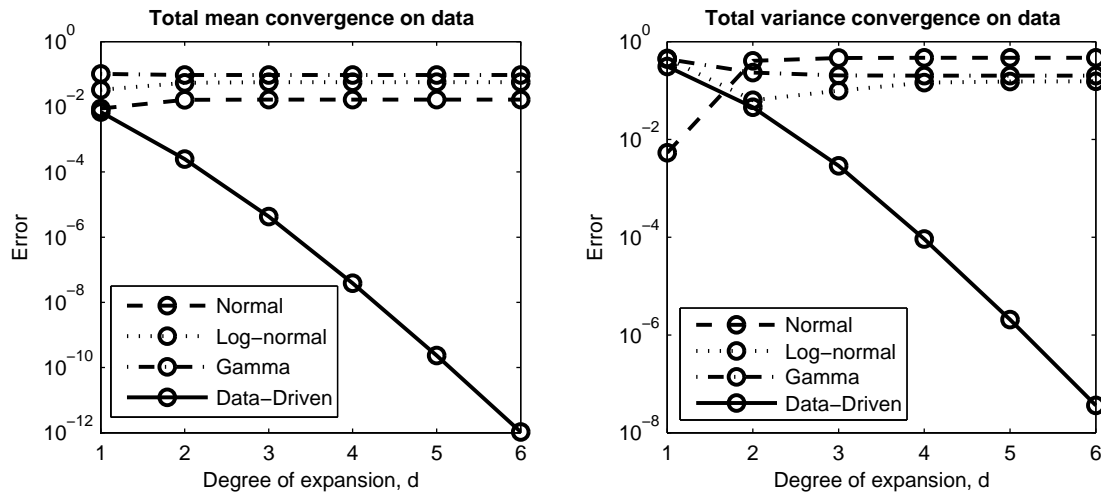
Figure 9.2.: Convergence on data: estimation of mean (left plot) and variance (right plot)

Technically, the coefficients c_i in the chaos expansion can be obtained, e.g., by Galerkin projection (e.g. (Ghanem and Spanos, 1993), (Matthies and Keese., 2005), (Xiu and Karniadakis, 2003)) or by the Collocation method (e.g. (Isukapalli et al., 1998), (Li and Zhang, 2007), (Shi et al., 2009)). Both methods lead to the same result when using the optimal polynomial basis in the case of univariate analysis. Figure 9.2 illustrates the convergence of the mean and variance (at time $t = 1$) for the assumed PDFs (left plot of Figure 9.1) and for the pure raw data (right plot of Figure 9.1).

All considered cases reproduce an acceptable approximation with the linear expansion. Increasing the expansion order shows strong convergence for the data-driven polynomial chaos expansion. However, increasing the order does not assure convergence for the expansions based on interpreted data. The problem does not lie in poor numerical properties when treating these distributions, but in accurate convergence to a wrong value. This error is introduced only by fitting parametric PDFs to the data instead of letting the data fully manifest themselves. In fact, the PDF-based analysis are good only up to first order, because only moments up to the second are represented accurately by the matched PDFs. This means, that all effort spent for higher order expansion (especially for complex problems) is invested to negligible improvement, if not matched with an adequate effort for accurate data interpretation. The key advantage of aPC is this respect is that (1) it allows full freedom in the used type of input information, and (2) our analysis in section 9.3 makes explicitly clear what amount of information (i.e., the moments up to a certain number) enters the analysis at what expansion order. This allows to align the complexity and order of analysis with the reliability and detail level of statistical information on the input parameters.

9.4.2. Evidence of improved convergence

In this section we will illustrate the efficiency of analysis within an optimal (data-driven) polynomial basis. We wish to show the improved convergence rate of the arbitrary polynomial chaos compared to the classical PCE technique. The classical PC requires non-linear transformations to map non-normal input data distributions onto the normal PDF. In this technical aspect, the classical PCE does not differ from the gPC, which requires transformation onto one of possible PDF from Askey scheme.

Equation (9.27) can be expanded (see equation 9.1) in the orthogonal polynomial basis $\Psi_i(\xi)$. The projection coefficients are defined as:

$$c_i(t) = \int_{\xi \in \Omega} Y(t)\Psi_i(\xi)d\Gamma(\xi), \quad i = \overline{0,d} \tag{9.29}$$

We will apply both the aPC and the classical PCE (including the mapping onto the normal distribution) to the example (9.27). For these two expansions techniques, we will apply both Galerkin projection (intrusive) and Gauss quadrature (non-intrusive) to evaluate the integral in equation (9.29). In both intrusive and non-intrusive approaches, the resulting values $c_i(t)$ from equation (9.29) will change with the distribution of the random variable ξ. For the optimal basis (used in aPC), however, the results from Galerkin projection and numerical integration coincide. This yields 3 distinguishable techniques. For all tree techniques, we analyse the performance for diverse exemplary distributions of the univariate input ξ (Rayleigh, Weibull, Log-normal). Detailed descriptions of these distributions can be found in (Evans et al., 200). For the classical PCE, the random variable ξ is not distributed in the same space as the polynomial basis $\Psi_i(\xi)$, and an additional conversion is required. Thus we map the model variable ξ onto a corresponding normal variable ξ_N by Gaussian anamorphosis or normal score transform (Wackernagel, 1998). Figure 9.3 illustrates the convergence of the mean and variance of Y (at time $t = 1$) for our three exemplary distributions of the model input ξ. As previously demonstrated for the gPC (Xiu and Karniadakis, 2002a), expansion in the optimal polynomial basis without transformation shows at least an exponential convergence. Convergence with a non-optimal basis (here: Hermite) after transformation strongly depends on the nonlinearity of the required transformation from ξ to ξ_N.
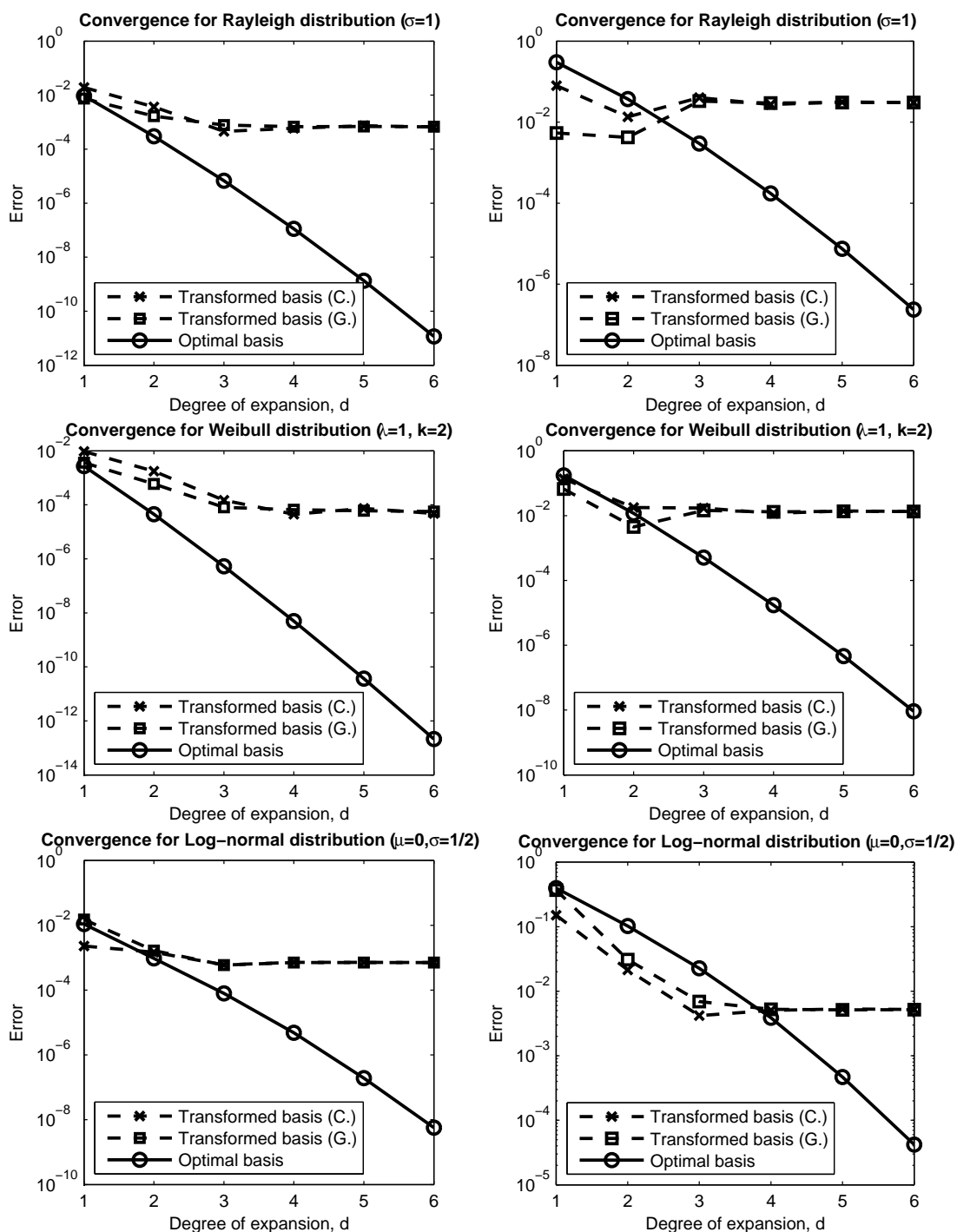
Figure 9.3.: Convergence of mean (left) and variance (right) estimation based on optimal basis and transformed basis using Galerkin projection (G.) and numerical integration (C.). For the optimal basis (used in aPC), the results from Galerkin projection and numerical integration coincide.

9.4.3. Clarification of error types

In section 9.4.1, we demonstrated the possible errors introduced by subjective data interpretation when fitting parametric PDFs to raw data. In that analysis, we deliberately excluded the error by the different numerical accuracies of aPC and classical PCE. Now, we will demonstrate the faster numerical convergence of aPC compared to classical PCE with transformation, but this time excluding the error of data interpretation. We classify the cause of error in two types: I - transformation expansion error and II - numerical integration error. Modeling within the non-optimal basis using Galerkin projection leads to errors by expanding and truncating the transformation, which we denote here as transformation expansion error (type I). Modeling within the non-optimal basis using Gauss quadrature entails numerical integration error (type II). Using the optimal basis provides identical results for both intrusive and non-intrusive methods, because numerical integration is exact when using the roots of the $d+1$ order polynomial from the optimal basis, and because no transformation from ξ to ξ_N is necessary.

9.4.4. Transformation expansion error (type I)

For intrusive manipulation, the anamorphosis transformation from ξ to ξ_N has to be expanded in ξ_N. The finite number d of terms in this expansion causes the first type of error. The difference between expansion in an optimal basis and expansion of $Y(\xi)$ after transformation to ξ_N is the so-called "aliasing error "(see (Xiu, 2007)). Figure 9.4 illustrates the nature of the this type of error. Expansion of the transformation (here $\xi = \exp(\xi_N)$) at different orders is shown in the left plot of Figure 9.4. The right plot of Figure 9.4 demonstrates the corresponding mapping of a normal probability density function (PDF) back to physical space using the expanded and truncated expansion. In these examples, the normal PDF should transform to a log-normal PDF. The strong nonlinearity of the logarithmic transformation leads to a poor approximation with a finite number of terms. Thus, the choice of a non-optimal polynomial basis for the model input ξ leads to a wrong representation of the probability measure Γ_ξ. This leads to the erroneous analysis of model output $Y(\xi)$ visible for Galerkin-based computations in Figure 9.3.
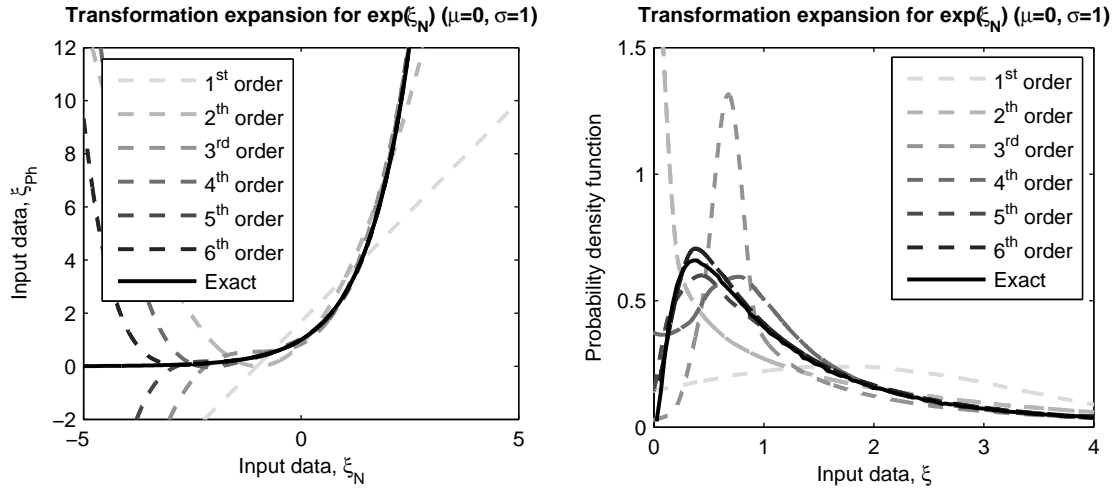
Figure 9.4.: Transformation expansion error: expansion series (left plot) and corresponding distribution(right plot)

9.4.5. Numerical integration error (type II)

The accuracy of numerical integration (especially sparse) strongly depends on the choice of integration points. For example, in Gauss-Hermite integration, the polynomial basis defines the positions ξ_i of integration points in the space of the input variable by the roots of the polynomial of degree $d + 1$. Thus, using a non-optimal polynomial basis provides a non-optimal choice of the integration points, which causes the second type of error. To illustrate this type of error, let us consider a stochastic model with a random variable ξ that follows a non-Gaussian distribution. The selected model is an extremely simple non-linear one:

$$Y(\xi_{Ph}) = \xi^6. \tag{9.30}$$

In our example, the input parameter ξ is distributed according to the Chi-square distribution. We will construct two expansions: one based on Hermite polynomials with adequate Gaussian anamorphosis, and one based on optimal polynomials for the Chi-square distribution of the input data. In both cases, we employ Gauss quadrature and compare the results to a reference solution. The supposedly optimal location of the integration points for Hermite polynomials correspond to the roots of the Hermite polynomial of order $d + 1$, back-transformed from ξ_N to ξ by anamorphosis. However, the truly optimal distribution of the integration points are the roots of the optimal polynomials that are orthogonal for ξ without further transformation. The obvious difference is shown in the left plot of Figure 9.5. The
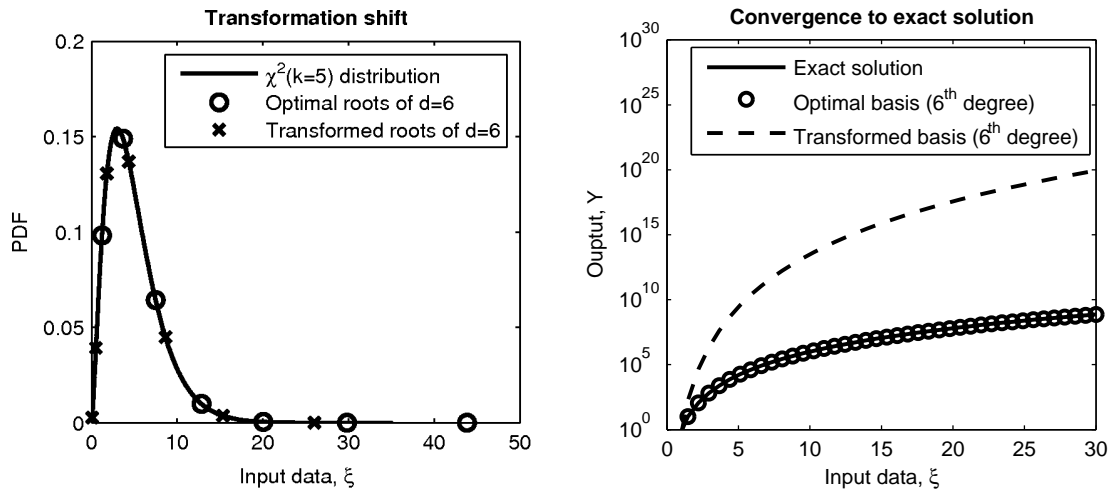
Figure 9.5.: Numerical integration error: transformation shift (left plot) and convergence to exact analytical solution (right plot)

transformed points are shifted against the optimal ones and thus cannot be considered as an optimal choice for numerical integration. Therefore, strong nonlinearity in the transformation leads to significant errors in PCE techniques that derive their numerical integration rules from the involved basis. Evidently, the example in equation (9.30) has an analytical solution, which can be reproduced by expansion of 6th order within the optimal basis (see the right plot in Figure 9.5). However, the transformed Hermite chaos combined with non-optimal Gauss quadrature does not converge to the known analytical solution even for the expansion degree $d = 6$ that should be, in theory, fully accurate construct $Y = \xi^6$.

9.5. Robustness analysis for inaccurate input data

The presented approach can handle different forms of input information. In particular, it can directly handle raw data, which can be useful for practical applications. However, when the input data set is small, the sample moments are only uncertain estimates of real moments. Hence, a direct application of the method presented becomes less robust. In that case, it would be useful to apply some standard methods to assess the robustness in the estimation of moments, such as Jackknife or Bootstrapping (e.g. (Efron, 1987)). In the field of reliability engineering, Bootstrap methods have been applied to construct upper confidence limits for unreliability in (Dargahi-Noubary and Razzaghi, 1992). Bootstrap-based confidence intervals caused by the uncertainty representing computationally demanding models

by meta-models has been investigated in the paper (Storlie et al., 2009) via regression based sensitivity analysis . The recent paper (Blatman and Sudret, 2010) explores sparse and partially random integration techniques for PCE, applied to sensitivity analysis, and provides Monte-Carlo based estimates for the error introduced by the random character of the used integration rules. In this section, we focus on the robustness of data-driven expansions with respect to the limited size of a raw data set, that represents the underlying probability distributions of model input only inaccurately.

For that, we repeatedly ($N = 1000$) generated raw data according to the assumed underling theoretical distribution. Each time, we constructed a new data-driven basis and performed a projection of the model output Y (equation (9.27)) to the corresponding data-driven polynomial basis and computed the mean μ_Y and variance σ_Y^2 of the model output $Y(t = 1)$ in each repetition. From this, we computed the variance of the mean $\sigma_{\mu_Y}^2$ and the variance of the variance $\sigma_{\sigma_Y^2}^2$. This entire nested Monte Carlo analysis was repeated for sizes of the raw data set ranging from $N = 20$ to 1000. Figure 9.6 shows the results for the distributions considered in Section 9.4.2. For this illustration, we used a 3rd degree of expansion. Other degrees of expansion ($1 - 6$) show similar results, all of them having error variances inversely proportional to the size of the sampled data set ($1/N$), i.e. having error standard deviations proportional to $1/\sqrt{N}$. This rate is visible as the slope of the scatter plots in Figure 9.6.

The scatter is caused by the finite number of Monte-Carlo repetitions used in the error estimation. It corresponds to the Monte-Carlo error of the error estimate. Only visually, the scatter increases with increasing size N of the data sets due to the logarithmic scale of the ordinate. The important aspect of the plots in the Figure 9.6, however, is not the degree of scatter (i.e. the uncertainty of the error estimation), but the average slope (i.e. the error estimate itself).

Apparently, the data-driven chaos expansion has a convergence rate proportional to $1/\sqrt{N}$ for the standard deviation and confidence intervals of computed model output statistics. This convergence rate is well-known for the variance of sample statics and from Monte-Carlo techniques in general (Caflisch, 1998). This means that the aPC does not modify the robustness and convergence properties with respect to insufficient sample size in comparison to moments from classical sample statistics or Monte Carlo simulation.

The analysis we performed here cannot be done with computationally expensive models, or if a single real data set is all that is available. In that case, one can perform a Jackknife or Bootstrapping method to estimate the sampling distribution of the used moments from that data set, and propogate the resulting randomized moments through the response surface
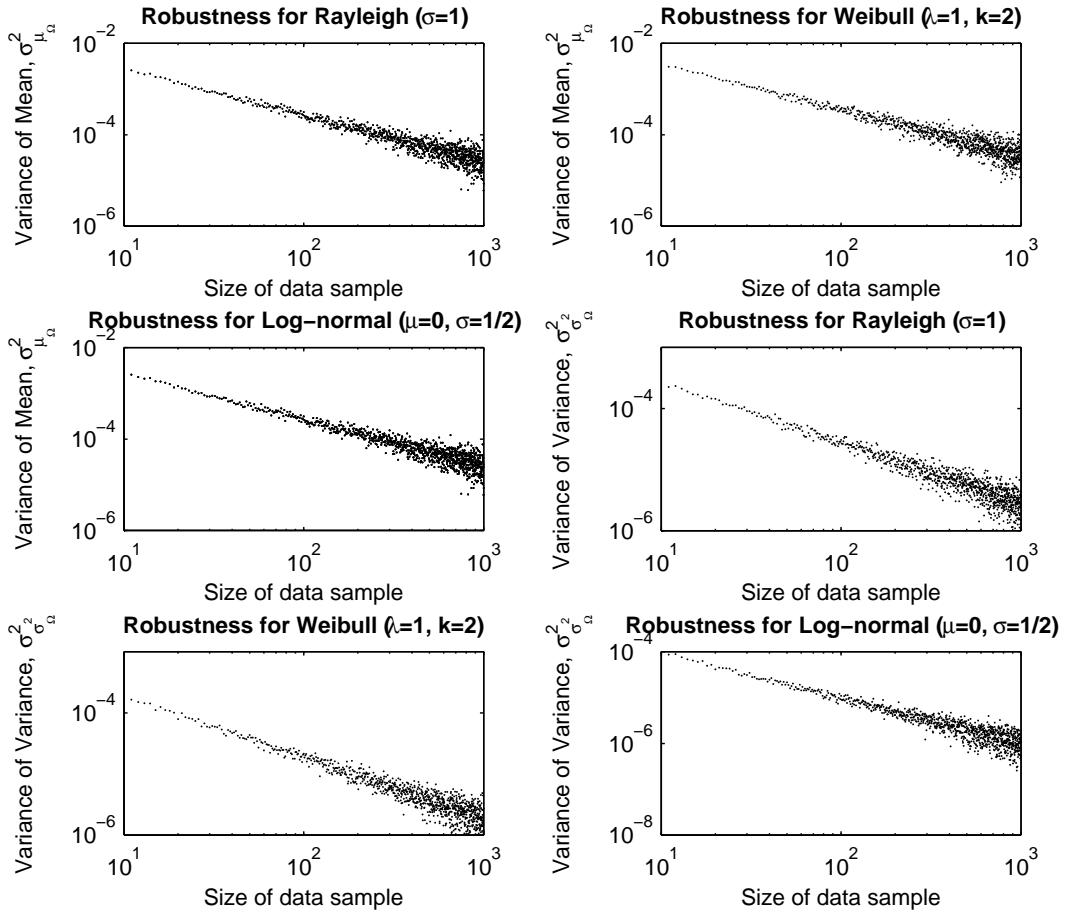
Figure 9.6.: Robustness of data-driven expansion with respect to the size of a raw data sample: variance of the mean (top) and variance of the variance (bottom)

obtained with a constant set of integration points. This will estimate the variances of the PCE solution due to the limited size of the data set, and corresponded to the value for a single given N in Figure 9.6.

Especially in such cases with very small data sets, expert opinion can be very useful to filter the data set, remove alleged outliers, fit a simple or complex PDF, and so forth. In our proposed approach, an expert will have total freedom of data interpretation (not restricted to the selection among standard PDFs) and can provide much more sophisticated information (e.g. lower and higher moments, complex and even non-parametric distributions, etc.). According to our approach, expert opinion (in a most general sense) will be incorporated directly without any additional transformation or additional subjectivity when translating it to the stochastic numerical framework. The presented methods allow experts to choose freely of technical constraints the shapes of their statistical assumptions.

9.6. Remaining issues for future research

The polynomial basis for continuous or discrete random variables can be constructed if and only if the number of support points (distinct values) in the distribution (within the available data set) is greater than the desired degree of the basis (see Property II in section 9.3.5). However, for discrete cases we cannot guarantee that the integration points (see section 9.4) will be distributed only within the space of a random variable. Still, it would be possible for each integration point to find a neighboring point that belongs to the discrete space, but the convergence is not guaranteed, and this remains an open question for future research.

Formal knowledge about the convergence of the polynomial basis for diverse random spaces can be very useful for practical needs, such as the convergence for the Hermite basis in normal space (Cameron and Martin, 1947). Karl Weierstrass established his approximation theorem in 1885, which states that every continuous function defined on an interval can be uniformly approximated as closely as desired by a polynomial function. A generalization of the Weierstrass theorem was proposed in the Stone-Weierstrass theorem where, instead of the interval, an arbitrary compact Hausdorff space is considered and, instead of the algebra of polynomial functions, approximation with elements from more general subalgebras were investigated (Stone, 1937). Thus, if the random space is an arbitrary compact Hausdorff space, uniform convergence is guaranteed and the polynomial space is dense in the arbitrary compact Hausdorf space. However, it is not assured that any polynomial expansion (including the optimal orthonormal basis) will uniformly converge in any random space and the definition of the random space will define convergence in that space.

The classical theorem of (Riesz, 1923) characterizes the problem of polynomial density by the unique solvability of a moment problem, which means that the distribution function is required to be uniquely defined by the sequence of its moments. Ernst et al. (Ernst et al., 2012) discussed such aspects for the generalized chaos expansion (gPC), and showed that the moment problem is not uniquely solvable for the lognormal distribution. However, the mentioned works demand existence and precise knowledge about the probability density function, which is neither required nor desired for the arbitrary polynomial chaos (aPC). We observed convergence of the aPC in section 9.4 for a counted number of useful cases, however research on a formal proof of convergence is a remaining question for future research.

As an outlook for future development, we point out the construction of a joint basis for parameters that have a complex statistical dependence beyond correlation. Following that direction, the aPC could be the first PCE family member that will allow to handle non-linear

statistical dependence between input variables.

9.7. Conclusions

In the current chapter, we presented the arbitrary polynomial chaos expansion (aPC). The aPC conception provides a constructive and simple tool for uncertainty quantification, global sensitivity analysis and robust design. It offers a new data-driven approach for stochastic analysis that avoids the subjectivity of assigning parametric probability distributions that are not sufficiently supported by available data. We show that a global orthonormal polynomial basis for finite-order expansion demands the existence of a finite number of moments only, and does not require exact knowledge or even existence of a probability density function. Thus, the aPC can be constructed for arbitrary parametric and non-parametric distributions of data, even if the statistical model output characterization of input data is incomplete.

Also, the orthonormal basis can be constructed without using any hierarchical conditions or recurrence relations with polynomials of lower-order. For discrete random variables, the aPC can be constructed if and only if the number of discrete values of the random variable is greater than the largest considered degree of the basis. In case of continuous random variables, the aPC can be constructed from a number of moments which equals to two times the degree of the basis. If desired, the method can work directly with raw sampled data sets to represent the uncertainty and possible variational ranges of input data. The presented methods allow experts to choose freely of technical constraints the shapes of their statistical assumptions and makes explicitly clear what amount of information (i.e., the moments up to a certain order) enters the analysis at what expansion order. Overall, this allows to align the complexity and order of analysis with the reliability and detail level of statistical information on the input parameters.

We also provided numerical studies for diverse exemplary distributions, where we illustrated convergence rates for optimal and non-optimal polynomial bases using intrusive and non-intrusive methods. This analysis strictly illustrated that using a non-optimal polynomial basis provides slow convergence of the chaos expansion and therefore causes additional errors in subsequent analysis. Modeling results within our new data-driven aPC show, at least, an exponential convergence for all examined cases. We defined, discussed and illustrated the difference between numerical integration error and transformation expansion error. Both errors lead to wrong estimates of statistical characteristics when using a non-optimal basis.

Thus, the aPC not only provides freedom for modeling physical systems with unknown probability distribution function, when only data sets of very limited size are available, but it also provides better convergence rates than conventional polynomial chaos techniques.

When the statistical information used as input is inaccurate or uncertain, i.e. when using small sets of raw data, a new form of uncertainty enters into the analysis and has to be considered in convergence assessments. We propose to apply Jackknife or Bootstrapping methods to asses the uncertainty of moments from small data sets, and then propagate that uncertainty through the aPC onto the output statistics to asses robustness. In an illustrative test case, we observe the classical Monte-Carlo convergence rate for the results of the PCE analysis with respect to the size of the raw data set.

10. A flexible and efficient framework for global sensitivity analysis

When investigating, modeling or operating uncertain systems, sensitivity analysis with respect to uncertain model parameters yields valuable information. It helps to quantify the relevance of parameters, to estimate their individual contributions to prediction uncertainty, and to better direct further data acquisition. In this work, we propose a response surface method for global sensitivity analysis (based on the arbitrary Polynomial Chaos Expansion, aPC). The key advantages of our proposed technique are: (1) aPC alleviates the computational burden associated with conventional global sensitivity analysis methods that require many evaluations of a simulation model; (2) the proposed method incorporates arbitrary independent probability densities or weighting functions for the investigated model parameters, thus generalizing several existing methods to reflect the expected relevance of parameters values within their allowable ranges; and (3) our framework allows to incorporate this information while requiring only a finite number of statistical moments for the investigated parameters. We generalize the polynomial-based computation of Sobol indices to arbitrary distributions, and suggest an associated complementary new sensitivity measure based on polynomial representation, which allows both univariate and multivariate global analysis. Compared to Sobol indices, our new weighted measure is absolute rather than relative, and converges faster with increasing order of expansion. We use analytical and hybrid analytical-numerical formulations that further improve computational efficiency. Altogether, we can conduct global sensitivity analysis at computational costs that are almost as low as those of local sensitivity analyses. We illustrate our approach for a 3D groundwater quality and human health risk problem in heterogeneous porous media.

10.1. Introduction

Understanding the general role of parameters in models and the impact of varying model parameters on the response of prediction models is a relevant subject in various fields of science and engineering. Characterizing the impact of parameter variations is known as sensitivity analysis and can be subdivided into local and global analysis (Sobol, 1990), (Saltelli et al., 2008), (Sudret, 2008). In many cases of practical interest, we wish to perform a Global Sensitivity Analysis (GSA) in order to analyze a model as such or to investigate, quantify and rank the effects of parameter variation or parameter uncertainty on the overall model uncertainty. GSA can also be used to: (1) quantify the relative importance of each individual input parameter in the final prediction (Sobol, 1990; Anderson and Burt, 1985; Winter et al., 2006); (2) aid engineers to produce more robust designs; and finally (3) help decision makers to allocate financial resources towards better uncertainty reduction, e.g., in problems related to risk assessment (de Barros and Rubin, 2008; de Barros et al., 2009). An important recommendation to keep in mind is that GSA should be global not only in the sense of looking at the entire range of possible parameter variations. It should also be used to assess the importance of parameters on a global, final model output or post-processing result that is relevant to generate new insight, or relevant for final decisions. GSA should not merely be applied to model-internal quantities that are of secondary importance for the scientific or application task at hand (Saltelli et al., 2008).

For example, the field of subsurface contaminant hydrology requires uncertainty estimates due to the ubiquitous lack of parameter knowledge caused by spatial heterogeneity of hydraulic properties in combination with incomplete characterization (Dagan, 1989; Rubin, 2003). For such reasons, we need to rely on probabilistic tools to predict contaminant levels and their overall health effects, and to quantify the corresponding uncertainties. Having efficient computational approaches to estimate uncertainty and to perform GSA in hydrogeological applications (and many other fields of science and engineering that feature uncertain dynamic or distributed systems) is desirable. It can inform modelers about the relevance of processes or parameters in the models they compile, and can inform engineers and decision makers about which parameters require most attention and where and how characterization efforts should be allocated such that prediction uncertainty can be minimized. Hence, there is an ever-increasing demand for having a GSA method that efficiently quantifies uncertainty and parameter relevance in complex and non-linear systems.

Several methods for estimating uncertainty are available such as numerical Monte Carlo (MC) simulations and perturbation techniques (for a detailed review in subsurface hydrol-

ogy, see Zhang (Zhang, 2002) or Rubin (Rubin, 2003)). Although conceptually straightforward, the MC approach is computationally demanding since the statistical accuracy of its predictions depend on the number of realizations used. Furthermore, perturbation techniques require some closure approximation and are limited to certain classes of problems. An alternative approach for handling uncertainty is through the polynomial chaos expansion (PCE) (Wiener, 1938). Generally, all PCE techniques can be viewed as an efficient approximation to full-blown stochastic modeling (e.g., exhaustive MC).

PCE has already been used for sensitivity analysis in different fields of applications (Sudret, 2008; Crestaux et al., 2009; Plischke, 2010). Applications in hydrogeological sciences can also be found in the literature (Zhang and Lu, 2004; Foo and Karniadakis, 2010; Fajraoui et al., 2011). The basic idea is to represent the response of a model to change in parameters variables through a response surface that is defined with the help of an orthonormal polynomial basis in the parameter space. One of the attractive features of PCE is the high-order approximation of error propagation (Ghanem and Spanos, 1990, 1991; Oladyshkin et al., 2011b) as well as its computational speed (when compared to MC). PCE techniques can mainly be sub-divided into intrusive (Ghanem and Spanos, 1993; Matthies and Keese., 2005; Xiu and Karniadakis, 2003) and non-intrusive (Keese and Matthies, 2003; Isukapalli et al., 1998; Li and Zhang, 2007; Oladyshkin et al., 2011b) approaches, i.e., methods that require or do not require modifications in the system of governing equations and corresponding changes in simulations codes. Reviews of the mathematical and conceptual background behind intrusive and non-intrusive PCE approaches are provided in (Jakeman and Roberts, 2008). The original PCE concept roots back to the work of Wiener in 1938 (Wiener, 1938) and can be used only for Gaussian distributed input parameters. In recent years, the classical PCE technique was extended to the generalized polynomial chaos (gPC) (Wan and Karniadakis, 2006) which accommodates for the use of an increased, yet limited number of statistical distributions (Askey and Wilson, 1985).

The PCE methods discussed above assume an exact knowledge of the probability density function and they are optimal only when applied to a finite number of certain parametric probability distributions. Unfortunately, information about the distribution of data or input parameters is very limited in many realistic applications, especially in environmental engineering and sciences. Data that characterize model parameters often indicate a variety of statistical distribution shapes (e.g., bounded, skewed, multi-modal, discontinuous, etc). Also, imperial parameter distributions derived from raw data sets do in general not follow analytically known distribution shapes. For such reasons, application tasks demand further adaptation of the chaos expansion technique to a larger spectrum of distributions.

To accommodate for a wide range of data distributions, a recent generalization of PCE denoted as the arbitrary polynomial chaos (aPC) was developed (Oladyshkin et al., 2011a,c; Soize and Ghanem, 2004; Ghanem and Doostan, 2006; Witteveen and Bijl, 2006). Compared to earlier PCE techniques, the aPC adapts to arbitrary probability distribution shapes of input parameters and, in addition, can even work with unknown distribution shapes when only a few statistical moments can be inferred from limited data or from expert elicitation. The arbitrary distributions for the framework here can be either discrete, continuous, or discretized continuous. They can be specified either analytically (as probability density/cumulative distribution functions), numerically as histogram or as a raw data set. The aPC approach provides improved convergence in comparison to classical PCE techniques, when applied to input distributions that fall outside the range of classical PCE (Oladyshkin et al., 2011c).

In this chapter, we tackle GSA based on the aPC technique. Because our framework accounts for arbitrary bounds or weighting functions for input parameters, it provides a weighted global sensitivity. In some sense, our work can be perceived as a generalization of the Morris method (Morris, 1991) to weighted analysis. Up to presence, the Morris method considers a uniform importance of input parameters within pre-defined intervals. We also generalize the Sobol indices (Sobol, 1990) for GSA to the aPC context, and provide a novel GSA measure which resembles a weighted square norm of sensitivities. Compared to Sobol indices, our new measure is absolute rather than relative. The advantage of an absolute index over a relative one is that it is a quantitative expression for the (averaged) derivative (slope), and hence is (a) useful even for a single parameter without comparison to other parameters, and (b) keeps the original meaning of a sensitivity as known from linear, local analysis.

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

Foglia et al. (Foglia et al., 2007) pointed out that, according to their experience, one can get 70% of the insight from 2% of the model runs when using local sensitivity analysis methods versus global methods. The big advantage of aPC-based GSA (or more generally: GSA based on any PCE technique), is that one can obtain global sensitivity information at computational costs that are hardly larger than those for local analysis. The reason is the following: Local methods use infinitesimally small spacing between parameter sets for model evaluation to get numerical derivatives evaluated at a single point. Our aPC based-method places

the parameter sets for model evaluation at an optimized spacing in the parameter space. This can be interpreted as fitting secants (or polynomials for non-linear analysis) to the model response. These secants (polynomials) approximate the model over the entire parameter space in a weighted least-square sense (compare with the best unbiased ensemble linearization approach described in (Nowak, 2009b)). This is more beneficial to computing a tangent or local second derivatives (compare FORM, SORM methods, e.g., (Jang et al., 1994)) that approximate the model well just around one point in the parameter space.

The main contributions of this work are to provide an alternative procedure to perform GSA that is computationally efficient and highly flexible. In particular, due to aPC, the proposed GSA can be interpreted as exploiting a smart (mathematically optimal) interpolation rule of model output between optimally chosen sets of input parameters, where the number of model evaluations is minimal. Also, we make full use of a hybrid analytical-numerical formulation that contributes even further to the computational efficiency of the approach. Compared to earlier works that related GSA to classical PCE, we (1) emphasize a more engineering-like language as compared to otherwise intense mathematical derivations and provide a clear 3-step procedure to perform our analysis, (2) develop easy-to-use analytical and semi-analytical expressions for frequent use in applications, (3) generalize PCE-based GSA to arbitrary probability distributions of the investigated parameters, and moreover, (4) the presented methods allows to align the complexity and order of analysis with the reliability and detail level of statistical information on the input parameters.

The methodology proposed here is designed for GSA as an investigative tool about model behavior, relevance of model input parameters towards prediction uncertainty, and hence, also for decision which parameters deserve most attention during further data collection or model calibration. It is not designed as a tool for conditional (post-calibration) analysis, unless the calibration outcome does not infer non-linear statistical dependence among parameters. For other situations, it would be adequate to develop similar methods based on approaches such as followed by (Bliznyuk et al., 2008).

The ideas presented in this work (to perform GSA within the aPC framework) can be summarized in a 3-step procedure, see Fig. 10.1. These steps are:

1. Data characterization: Assess or compute the statistical moments of the input parameters;

2. Approximate the model response surface by a polynomial;

3. Evaluate sensitivity metrics efficiently from the polynomial.
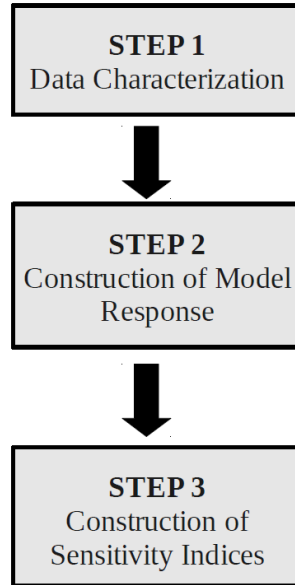
Figure 10.1.: Illustration of the 3-step procedure for the methodology used in this work.

In Section 10.2, we describe the methodological framework for multi-parameter GSA. Section 10.3 provides the simple 3-step algorithm shown in Fig. 10.1 and explains the computational details. In Section 10.4, we derive an explicit analytical form for single-parameter sensitivity analysis based on a novel weighted sensitivity measure. Section 10.5 demonstrates how to obtain Sobol indices and our novel weighted sensitivity indices for multi-parameters analysis. We illustrate the ideas in a synthetic application of assessing human health risk due to a 3D contaminant transport problem in a heterogeneous aquifer in Section 10.6. Conclusions are provided in Section 10.7.

10.2. Non-intrusive chaos expansion for arbitrary parameter distributions

10.2.1. Definitions and polynomial chaos expansion

In this Section, we will introduce the PCE-based framework used in this chapter for multi-parameter global sensitivity analysis (GSA). Let $\omega = \{\omega_1,...,\omega_N\}$ represent the vector of N input parameters for some model $\Omega = f(\omega)$. The model $\Omega(\omega)$ may be an explicit or implicit expression (e.g., a partial or ordinary differential equation or a coupled system). To

perform GSA, we wish to investigate the influence of all parameters ω on the model output Ω. Within hydrogeological tasks, Ω can represent, e.g., hydraulic head or concentration values at a specified location $\mathbf{x} = (x_1, x_2, x_3)$ and time t, or the resulting human health risk for an exposed population. The vector ω may contain input parameters such as hydraulic conductivity, velocity, porosity, reaction rate coefficients, geostatistical structural parameters, toxicity or exposure durations, etc. In the following, we will approximate the model response by a truncated polynomial expansion for each point in space $\mathbf{x}$ and time t. According to polynomial chaos theory (Wiener, 1938), the model output Ω can be approximated by polynomials $\Psi_j(\omega)$ as follows:

$$\Omega(\mathbf{x}, t; \omega) \approx \sum_{j=0}^{M} c_j(\mathbf{x}, t) \Psi_j(\omega), \tag{10.1}$$

where the number M of polynomials depends on the total number of analyzed input parameters (N) and the order d of the polynomial representation, according to the combinatoric formula $M = (N+d)!/(N!d!) - 1$. The coefficients c_j in equation (10.1) quantify the dependence of the model output Ω on the input parameters ω for each desired point in space $\mathbf{x}$ and time t. The symbol Ψ_j is a simplified notation of the multi-variate orthonormal polynomial basis for ω including all cross-terms between different parameters, as explained below. Let us mention that, in the current state of science for polynomial chaos expansion, the random variables have to be statistically independent or may be linearly correlated, which can be removed by adequate linear transformation. Construction of a polynomial basis for statistically dependent random variables beyond linear dependence is a very important issue for future research.

10.2.2. Data-driven orthonormal basis

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

linear combination of the different powers i of ω_j:

$$P_j^{(k)}(\omega_j) = \sum_{i=0}^{k} p_{i,j}^{(k)} \omega_j^i, \quad k = 0 \dots d, \quad j = 0 \dots N. \tag{10.2}$$

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

$$\Psi_k(\omega) = \prod_{j=1}^{N} P_j^{(\alpha_j^k)}(\omega_j), \quad \sum_{j=1}^{N} \alpha_j^k \leq M, \quad k = 1 \dots N, \tag{10.3}$$

where α_j^k is a multivariate index that contains the combinatoric information how to enumerate all possible products of individual univariate basis functions. In other words, the index α can be seen as $M \times N$ matrix, which contains the corresponding degree (e.g. 0, 1, 2, etc.) for parameter number j in expansion term k. The multivariate basis allows to represent the reaction of a model Ω to several (N) parameters ω_j ($j = 0 \dots N$), as an N-dimensional polynomial response surface, defined by the expansion in equation (10.1).

We will show now, how to construct the data-driven orthogonal polynomial basis for each individual component ω_j from the vector ω. The main idea of the data-driven approach, see (Oladyshkin et al., 2011a,c), consists in constructing the coefficients $p_{i,j}^{(k)}$ for equation 10.2 in such a way that the polynomials in equation (10.2) form a basis that is orthonormal in precisely the given input distributions of model parameters. It does so without posing any restrictions to the statistical distribution shapes or weighting functions that available data, expert opinion or modeler experience may assume.

According to (Oladyshkin et al., 2011a,c), an orthogonal polynomial basis up to order d can be constructively defined for any arbitrary probability measure, given that ω_j has finite statistical moments (e.g., mean, variance, skewness, etc) up to order $2d - 1$. The unknown polynomial coefficients $p_{i,j}^{(k)}$ can be defined (among other available options for construction (Abramowitz and Stegun, 1965), (Witteveen and Bijl, 2006), (Stieltjes, 1884)) from the following matrix equation:

$$\begin{bmatrix} \mu_{0,j} & \mu_{1,j} & \dots & \mu_{k,j} \\ \mu_{1,j} & \mu_{2,j} & \dots & \mu_{k+1,j} \\ \dots & \dots & \dots & \dots \\ \mu_{k-1,j} & \mu_{k,j} & \dots & \mu_{2k-1,j} \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} p_{0,j}^{(k)} \\ p_{1,j}^{(k)} \\ \dots \\ p_{k-1,j}^{(k)} \\ p_{k,j}^{(k)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \\ 1 \end{bmatrix}. \tag{10.4}$$

Here, $\mu_{i,j}$ are the ith non-central (raw) statistical moments for random variable ω_j. It becomes evident from equation (10.4) that statistical moments are the only required form of information on the input distributions. This property carries over to all types of Taylor or polynomial chaos expansions. An analytical explicit form for $p_{i,j}^{(k)}$ at moderate degrees of expansion can be obtained (see Section 10.2.3). More detailed properties of the basis $\{P_j^{(0)},\ \ldots,P_j^{(d)}\}$ and the conditions for non-singularity of the matrix in equation (10.4) are presented in (Oladyshkin et al., 2011c).

The above orthogonal polynomial basis can be used directly for analysis. However, a normalized basis has further useful properties. For example, the mean and variance of $\Omega(\omega)$ according to the expansion (10.1) is given by simple analytical relations (see (Oladyshkin et al., 2011a)), by virtue of the orthonormality property. This follows from the general properties of Fourier expansions, which encompass all expansions in orthonormal bases. The squared coefficients of such expansions are called the spectrum (compare with the Fourier transformation). The sum of squared coefficients is the total energy or variance. Due to orthonormality, most terms cancel out in subsequent steps. For example, if we consider a stochastic process in the probability space (Λ, A, Γ) with space of events Λ, σ-algebra A and probability measure Γ (see e.g. (Grigoriu, 2002)) and if $\widehat{P}_j^{(k)}(\omega_j)$ is an orthonormal basis, then by definition of orthogonality:

$$\int_{\omega_j \in \Lambda} \widehat{P}_j^{(k)}(\omega_j) \widehat{P}_j^{(l)}(\omega_j) d\Gamma(\xi) = \delta_{kl}. \tag{10.5}$$

The orthonormal polynomial basis can be obtain as:

$$\widehat{P}_j^{(k)}(\omega_j) = \frac{P_j^{(k)}}{\left\| P_j^{(k)} \right\|}, \tag{10.6}$$

where the normalizing constant of the polynomial $\|P_j^k\|$ for space of events Λ (where $\omega_j \in \Lambda$) with probability measure Γ is defined as:

$$\left\| P_j^{(k)} \right\|^2 = \int_{\omega_j \in \Lambda} \left[P_j^{(k)}(\omega) \right]^2 d\Gamma(\omega_j). \tag{10.7}$$

or, if a probability density function (PDF) or weighting functions $f(\omega_j)$ is given, we have:

$$\left\| P_j^{(k)} \right\|^2 = \int_{-\infty}^{+\infty} \left[P_j^{(k)}(\omega) \right]^2 f(\omega_j) d\omega_j. \tag{10.8}$$

Because the square of a polynomial of order k yields a polynomial of order $2k$, normalization according to equation (10.7) or (10.8) requires the statistical moments of ω up to order $2d$.

10.2.3. Explicit analytical form of orthonormal data-driven polynomial basis

In this section, we present an explicit analytical form for linear and non-linear GSA. The definition in equation (10.15) is invariant to the transformation (10.23). Thus, the coefficients of the orthonormal data-driven polynomial basis $p_i^{(d)}$ $(i = 0 \ldots d)$ for parameter ω_j can be defined explicitly through the raw moments of ξ. For moderate order polynomials, they can be written as:

$$P_j^{(0)}(\xi) = 1; \tag{10.9}$$

$$P_j^{(1)}(\xi) = \xi; \tag{10.10}$$

$$P_j^{(2)}(\xi) = -1 - \mu_3 \xi + \xi^2; \tag{10.11}$$

$$P_j^{(3)}(\xi) = \mu_3^2 - \mu_3^3 + \mu_3 \mu_4 - \mu_5 + \xi(\mu_3^2 - \mu_3 \mu_5 - \mu_4 + \mu_3 \mu_4) \tag{10.12}$$

$$+ \xi^2(-\mu_3 \mu_4 + \mu_5 - \mu_3) + \xi^3(1 - \mu_3 + \mu_3^2)$$

10.2.4. Non-intrusive determination of the expansion coefficients

The remaining task is to evaluate the coefficients c_j in equation (10.1). In this chapter, we use the non-intrusive probabilistic collocation method (PCM) (Li and Zhang, 2007). The collocation formulation does not require any knowledge of the initial model structure, i.e., of Ω. It only requires knowledge on how to obtain the model output for a given set of input parameters, which allows to treat the model Ω like a "black-box". The idea of PCM is to evaluate the model exactly M times, which allows to directly fit the polynomial representation of Ω, see equation (10.1), with its M unknown coefficients c_j to the obtained M model results. The M model evaluations are performed with M different parameter sets $\{\omega_1^{(i)}, ..., \omega_N^{(i)}\}$, $i = 1, .., M$, called collocation points.

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

$$\mathbf{M}_\Psi(\omega)\mathbf{V}_c(\mathbf{x}, t) = \mathbf{V}_\Omega(\mathbf{x}, t; \omega) \tag{10.13}$$

where $\mathbf{V}_c$ is the $M \times 1$ vector of unknown coefficients c_j in expansion (10.1), the $M \times 1$ vector $\mathbf{V}_\Omega$ contains the model output for each collocation point, and the $M \times M$ matrix $\mathbf{M}_\Psi$

contains the polynomials evaluated at the collocation points:

$$\mathbf{M}_{\Psi} = \left\{ \Psi_i(\omega_1^{(i)}, ..., \omega_N^{(i)}) \right\}, \quad i = 1...M, j = 1...M;$$

$$\mathbf{V}_{\Omega} = \left\{ \Omega_i(x, y, z, t, \omega_1^{(i)}, ..., \omega_N^{(i)}) \right\}, \quad i = 1...M; \quad \quad (10.14)$$

$$\mathbf{V}_c = \left\{ c_i(x, y, z, t) \right\}, \quad i = 1...M.$$

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

The solution $\mathbf{V}_c$ of the system (10.13) depends on the selection of collocation points. According to Villadsen and Michelsen (Villadsen and Michelsen, 1978), the optimal choice of collocation points corresponds to the roots of the polynomial of one degree higher $(d+1)$ than the order used in the chaos expansion (d). Once the orthonormal polynomial basis is constructed using data (or assumptions on data), the collocation points become as well data-driven and optimally distributed in the space of input parameters. This strategy is based on the theory of Gaussian integration (e.g., (Abramowitz and Stegun, 1965)), and allows exact numerical integrations of order d given $d+1$ values of the function to be integrated.

The data-driven polynomial basis (see Section 10.2.2) defines the positions of the collocation points specific to the distribution of input parameters at hand and, thus, indicates what are the optimal parameter sets for model evaluation using all available information about the input parameters. For multi-parameter analysis, the number of available points from the original optimal integration rule is $(d+1)^N$, which is larger than the necessary number M of collocation points. The full tensor grid can be used for low-order (1st, 2nd) analysis of few parameters. However, for higher-order analysis of many parameters, the tensor grid suffers from the curse of dimensionality ($(d+1)^N$ points). In that case, a smart choice of a sparse subset of the tensor grid becomes necessary. For this reason, the collocation approach became more popular in the last years. Probabilistic collocation chooses the collocation points according to their probability weight, i.e. their importance as specified by the available probability distribution of ω. This simply means to select the collocation points from the most probable regions of the input parameters' distribution (see (Oladyshkin et al., 2011b)) and the modeler can extract a lot of information in the main range of the parameter distribution.

10.3. Three-step algorithm for data-adaptive global sensitivity analysis

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

Step 1. Characterize the model parameters to be investigated: Compute the raw moments $\mu_{k,j}$ ($k = 1\ldots2d$) of the input data for each input parameter ω_j ($j = 1\ldots N$). If a PDF is provided, then we can evaluate the actual theoretical moments *in lieu* of the raw data moments of a given data set. Alternatively, expert elicitation may serve to provide opinions on these moments, e.g., by guessing a distribution shape and then using its moments.

Step 2. Approximate the model response surface by polynomials: For the specific moments $\mu_{k,j}$, compute the coefficients of the optimal polynomial basis in equation (10.3) using the system of linear equations equation (10.4) and the normalization in equation (10.6). In addition, compute the coefficients of the expansion using the system of linear equations, equation (10.13), to represent the model response.

Step 3. Compute the desired sensitivity information from the polynomials: With the original model reduced to a multi-variable polynomial, the sensitivity indices can be obtained analytically without any heavy additional computational efforts. This is achieved by using the relations provided in Eqs. (10.18) and (10.39) or (10.40) provided in the two upcoming Sections.

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

1. It constructs an optimal orthonormal polynomial basis for any desired distribution of data, whereas previous PCE techniques were limited to a small number of statistical distributions. Since only statistical moments are relevant, the input distributions may be either discrete, continuous, or discretized continuous and can be specified either through some statistical moments, analytically as PDF, numerically as histograms, or theoretically through the even more general format of probability measures.

2. The algorithm performs an optimal projection of the physical model onto a polynomial basis with minimum computational effort. This reduces the original model to a set of polynomials with many useful properties that allow immensely fast evaluation and offer a list of analytical relations.

3. The presented algorithm can easily be implemented without any deep knowledge related to the theory of chaos expansion and projection techniques.

4. The model Ω does not have to be modified and no specific properties are required. Hence, it may by given in any arbitrary form for simple analytical solution up to a multi-scale multi-physical simulation software framework.

10.4. Single-parameter global sensitivity analysis

10.4.1. Weighted sensitivity measure

Previous studies that exploited polynomial representations for sensitivity analysis looked at Sobol sensitivity indices (e.g., (Sudret, 2008), (Crestaux et al., 2009)). Unfortunately, Sobol sensitivity indices (see Section 10.5) cannot be applied for single-parameter sensitivity analysis, because their aim is to compare the relative impact between several parameters or combinations of parameters among each other. First, we will consider the influence of an individual parameter ω_j from the vector ω on the model response Ω in an alternative fashion. Later, we will move on to multi-parameter analysis.

Classical single-parameter sensitivity analysis (e.g., (Anderson and Burt, 1985)) investigates the influence of an input parameter on the model output Ω by changing its value by some defined small amount Δ (e.g., $\pm 10\%$) around some reference value ω_j^o or within some defined interval $[\omega_j^l, \omega_j^r]$. Such approaches fail to reflect the expected magnitude of possible changes in ω_j as well as expectations concerning their chances to occur. Instead, let us introduce a weighted global index for GSA using an arbitrary probability measure $d\Gamma$:

$$S_{\omega_j}^2 = \int_{\omega_j \in \Lambda} \left[\frac{\partial \Omega}{\partial \omega_j}\right]^2 d\Gamma(\omega_j) \tag{10.15}$$

This global sensitivity index reflects the squared slope $\partial \Omega / \partial \omega_j$ averaged over the statistical distribution or weighting function of ω_j. Compared to Sobol indices (see Section 10.5) which analyze the spectral energy (variance) contribution of individual parameters to the model Ω, our new measure looks at the spectral energy within the derivatives $\partial \Omega / \partial \omega_j$.

When using a uniform weight within some interval $[\omega_j^l, \omega_j^r]$, our approach yields a computationally optimal evaluation method for sensitivity as defined by the Morris method (Morris, 1991). For all other weighting functions, probability distributions or measures, it is a generalization of the Morris method.

Substituting the expansion in equation (10.1) into equation (10.15) and considering just one parameter ω_j, we obtain:

$$S^2_{\omega_j} = \int_{\omega_j \in \Lambda} \left[\sum_{k=0}^{d} c_k \frac{\partial P_j^{(k)}(\omega_j)}{\partial \omega_j} \right]^2 d\Gamma(\omega_j), \tag{10.16}$$

Expressing the derivative of $P_j^k(\omega_j)$ via a polynomial expansion of order $k-1$, equation (10.16) can be re-written as:

$$S^2_{\omega_j} = \int_{\omega_j \in \Lambda} \left[\sum_{k=0}^{d} c_k \sum_{i=0}^{k-1} b_i^{(k-1)} P_j^{(i)}(\omega_j) \right]^2 d\Gamma(\omega_j)$$

$$= \int_{\omega_j \in \Lambda} \left[\sum_{i=0}^{d-1} P_j^{(i)}(\omega_j) \sum_{k=1}^{d} c_k b_i^{(k-1)} \right]^2 d\Gamma(\omega_j), \tag{10.17}$$

where the coefficients $b_i^{(k-1)}$ are a simple re-collection of the coefficients c_k (see section 10.4.2).

Due to the orthonormality of the polynomials $P_j^{(i)}(\omega_j)$, equation (10.17) immediately simplifies to:

$$S^2_{\omega_j} = \sum_{i=0}^{d-1} \left[\sum_{k=1}^{d} c_k b_i^{(k-1)} \right]^2. \tag{10.18}$$

In the following, we obtain explicit forms for the expansion (10.1) and for global sensitivity according to equation (10.18) in a linear and non-linear analysis.

10.4.2. Exact expansion of partial derivatives of polynomials

Every partial derivative of polynomial $P_j^{(\alpha_j^k)}(\omega_j)$ can be represented through orthogonal polynomial basis of order $\alpha_j^k - 1$:

$$\frac{\partial P_j^{(\alpha_j^k)}(\omega_j)}{\partial \omega_j} = \sum_{i=0}^{\alpha_j^k - 1} b_i^{(\alpha_j^k - 1)} P_j^{(i)}(\omega_j), \tag{10.19}$$

where the $b_i^{(\alpha_j^k - 1)}$ result from a re-collections of polynomial coefficients.

Taking the derivative on the left-hand side, equation(10.19) can be written as:

$$
\begin{bmatrix} p_{1,j}^{(\alpha_j^k)} \\ 2p_{2,j}^{(\alpha_j^k)} \\ \cdots \\ \alpha_j^k p_{\alpha_j^k,j}^{(\alpha_j^k)} \end{bmatrix}^T \begin{bmatrix} 1 \\ \omega_j \\ \cdots \\ \omega_j^{\alpha_j^k-1} \end{bmatrix} = \begin{bmatrix} b_0^{(\alpha_j^k-1)} \\ b_1^{(\alpha_j^k-1)} \\ \cdots \\ b_{\alpha_j^k-1}^{(\alpha_j^k-1)} \end{bmatrix}^T \begin{bmatrix} P_j^{(0)}(\omega_j) \\ P_j^{(1)}(\omega_j) \\ \cdots \\ P_j^{(\alpha_j^k-1)}(\omega_j) \end{bmatrix}. \tag{10.20}
$$

At the same time, the polynomials of ω_j in equation (10.2) can be re-written in the following matrix form:

$$
\begin{bmatrix} P_j^{(0)}(\omega_j) \\ P_j^{(1)}(\omega_j) \\ \cdots \\ P_j^{(\alpha_j^k-1)}(\omega_j) \end{bmatrix} = \begin{bmatrix} p_{0,j}^{(0)} & 0 & 0 & 0 \\ p_{0,j}^{(1)} & p_{1,j}^{(1)} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ p_{0,j}^{(\alpha_j^k-1)} & p_{1,j}^{(\alpha_j^k-1)} & \cdots & p_{\alpha_j^k-1,j}^{(\alpha_j^k-1)} \end{bmatrix} \begin{bmatrix} 1 \\ \omega_j \\ \cdots \\ \omega_j^{\alpha_j^k-1} \end{bmatrix}, \tag{10.21}
$$

From equation (10.20) and equation (10.21), we obtain all re-collected coefficients b_i as solution of the following linear system:

$$
\begin{bmatrix} b_0^{(\alpha_j^k-1)} \\ b_1^{(\alpha_j^k-1)} \\ \cdots \\ b_{\alpha_j^k-1}^{(\alpha_j^k-1)} \end{bmatrix} \begin{bmatrix} p_{0,j}^{(0)} & 0 & 0 & 0 \\ p_{0,j}^{(1)} & p_{1,j}^{(1)} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ p_{0,j}^{(\alpha_j^k-1)} & p_{1,j}^{(\alpha_j^k-1)} & \cdots & p_{\alpha_j^k-1,j}^{(\alpha_j^k-1)} \end{bmatrix}^T = \begin{bmatrix} p_{1,j}^{(\alpha_j^k)} \\ 2p_{2,j}^{(\alpha_j^k)} \\ \cdots \\ \alpha_j^k p_{\alpha_j^k,j}^{(\alpha_j^k)} \end{bmatrix}. \tag{10.22}
$$

10.4.3. Linear sensitivity analysis

First, we will present an explicit form for linear data-driven global sensitivity analysis, which demands only knowledge about the first three moments μ_1, μ_2 and μ_3 for the statistical distribution or weighting function of ω_j. To simplify the explicit form of the coefficients, we will assume a normalized distribution of the parameter with zero mean and unit variance after the linear transformation:

$$
\xi = \frac{(\omega_j - \mu)}{\sigma}, \tag{10.23}
$$

where $\mu = \mu_1$ is the mean and $\sigma = \sqrt{\mu_2 - \mu_1^2}$ is standard deviation of the parameter ω_j. This may be done without loss of generality.

According to the system of equations equation (10.4) and the standardizing transformation (10.23), the linear polynomial expansion can be expressed as:

$$\Omega(\xi) = c_0 + c_1 \xi \tag{10.24}$$

Using the expansion (10.24) and the explicit form of the polynomials (10.9) and (10.10) presented in Section 10.2.3, the sensitivity index S_{ω_j} according to equation (10.18) for linear analysis can be simply written as follows:

$$S_{\omega_j}^2 = c_1^2 \tag{10.25}$$

According to the optimal integration rule (Villadsen and Michelsen, 1978) (see details in Section 10.4.4), the collocation points can be defined explicitly by equation (10.31). Finally, substituting the analytical solution for the collocation points (10.31) into equation (10.25) and incorporating the transformation (10.23), we obtain the following analytical form of the linear sensitivity index S_{ω_j} for parameter ω_j:

$$S_{\omega_j} = \frac{\left| \Omega\left(\omega_j^*\right) - \Omega\left(\omega_j^{**}\right) \right|}{\sqrt{\mu_2}\sqrt{\mu_3^2 + 4}}, \tag{10.26}$$

$$\omega_j^* = \mu_1 + \frac{\sqrt{\mu_2}\mu_3}{2} + \frac{\sqrt{\mu_2}\sqrt{\mu_3^2 + 4}}{2},$$

$$\omega_j^{**} = \mu_1 + \frac{\sqrt{\mu_2}\mu_3}{2} - \frac{\sqrt{\mu_2}\sqrt{\mu_3^2 + 4}}{2}.$$

For a symmetric weighting function with $\mu_3 = 0$ distributed around zero with standard deviation σ, this simplifies to:

$$S_{\omega_j} = [\Omega(\sigma) - \Omega(-\sigma)]/(2\sigma), \tag{10.27}$$

which reflects a central finite difference approximation to the slope, based on a finite parameter variation of $\Delta\omega_j = \pm\sigma$.

10.4.4. Linear analysis: explicit form of expansion

For a linear expansion, according to the optimal integration rule (Villadsen and Michelsen, 1978), the collocation points (see section 10.2.4) are defined as roots of the second order polynomial:

$$-1 - \mu_3 \xi + \xi^2 = 0 \tag{10.28}$$

thus, the collocation points are given by the following quadratic formula:

$$\xi_{1,2} = \frac{\mu_3}{2} \pm \frac{\sqrt{\mu_3^2 + 4}}{2}. \tag{10.29}$$

Hence, when approximating the projection of $\Omega(\omega)$ onto a polynomial basis (see equation 10.13) via collocation, we can write the following linear system:

$$\begin{bmatrix} 1 & \xi_1 \\ 1 & \xi_2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} \Omega(\xi_1) \\ \Omega(\xi_2) \end{bmatrix}, \tag{10.30}$$

which has an explicit solution

$$c_0 = \frac{\Omega(\xi_2)\xi_1 - \Omega(\xi_1)\xi_2}{\xi_1 - \xi_2}, \tag{10.31}$$

$$c_1 = \frac{\Omega(\xi_1) - \Omega(\xi_2)}{\xi_1 - \xi_2}$$

10.4.5. Non-linear sensitivity analysis

We now address a non-linear data-driven GSA based on a second order expansion, which includes the first five moments of the input parameters. According to the system of equations in equation (10.4) and the transformation provided in equation (10.23), the second-order polynomial expansion can be written as:

$$\Omega(\xi) = c_0 + c_1 \xi + c_2(-1 - \mu_3 \xi + \xi^2) \tag{10.32}$$

Using the explicit form of the polynomials (10.9)-(10.11) and the expansion (10.32), the sensitivity index S_{ω_j} for second-order single-parameter sensitivity analysis has the following form:

$$S_{\omega_j}^2 = (c_1 - \mu_3 c_2)^2 \tag{10.33}$$

For simplicity of notation, the expansion (10.32) can be rewritten as follows:

$$\Omega(\xi) = b_0 + b_1 \xi + b_2 \xi^2 \tag{10.34}$$

where $b_0 = c_0 - c_2$, $b_1 = c_1 - \mu_3 c_2$ and $b_2 = c_2$. Again, the coefficients in expansion (10.34) can be defined using the collocation method (see equation (10.13)) as solution of the following linear system:

$$\begin{bmatrix} 1 & \xi_1 & \xi_1^2 \\ 1 & \xi_2 & \xi_2^2 \\ 1 & \xi_3 & \xi_3^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \Omega(\xi_1) \\ \Omega(\xi_2) \\ \Omega(\xi_3) \end{bmatrix}, \tag{10.35}$$

Notice that, by definition, the coefficient b_1 is exactly equal to S_{ω_j} and, hence, the sensitivity index S_{ω_j} can be found as solution of the linear system (10.35), which forms a Vandermonde matrix:

$$S_{\omega_j} = \frac{\Omega_1(\xi_3{}^2 - \xi_2{}^2) + \Omega_2(\xi_1{}^2 - \xi_3{}^2) + \Omega_3(\xi_2{}^2 - \xi_1{}^2)}{\xi_2\,\xi_1{}^2 + \xi_3\,\xi_2{}^2 - \xi_3\,\xi_1{}^2 - \xi_2\,\xi_3{}^2 - \xi_1\,\xi_2{}^2 + \xi_1\,\xi_3{}^2}, \tag{10.36}$$

where $\Omega_1 = \Omega(\xi_1)$, $\Omega_2 = \Omega(\xi_2)$, $\Omega_1 = \Omega(\xi_3)$ and ξ_1, ξ_2, ξ_3 are the points where model the $\Omega(\xi)$ should be evaluated. Similar to the previous Section 10.4.3, the optimal location of integration points (collocation points) is defined via the roots of the third-order polynomial (10.12), and therefore can be obtained as the solution of the following cubic equation:

$$\mu_3^2 - \mu_3^3 + \mu_3\mu_4 - \mu_5 + \xi(\mu_3^2 - \mu_3\mu_5 - \mu_4 + \mu_3\mu_4) \tag{10.37}$$

$$+ \xi^2(-\mu_3\mu_4 + \mu_5 - \mu_3) + \xi^3(1 - \mu_3 + \mu_3^2) = 0$$

To obtain the roots for given moments μ_3, μ_4 and μ_5, standard analytical or numerical methods can be applied, provided as built-in functions in commonly used mathematical software such as MATLAB, MAPLE and MATHEMATICA.

10.5. Multi-parameter global sensitivity analysis

Within the data-driven PCE framework used in this work, the statistics of the model output are based directly on the model formulation and the specified moments of the input data. If a model output Ω is expanded in the normalized polynomial basis, equation (10.6), then many statistics of Ω can be evaluated directly via closed-form analytical relations. For example, the mean μ_Ω and variance σ_Ω^2 are given by the following simple analytical relations:

$$\mu_\Omega = c_0, \quad \sigma_\Omega^2 = \sum_{j=1}^{M} c_j^2, \tag{10.38}$$

where the mean depends on the zero-order constant only, and the variance is based on Parseval's expression (e.g., (Siebert, 1986)).

The so-called Sobol indices for sensitivity estimation (Sobol, 1990) can be computed analytically based on the PCE (see (Sudret, 2008) and (Crestaux et al., 2009)) using equa-

tion (10.39):

$$S_{i_1,...,i_s} = \frac{\sum_{j=1}^{M} \chi_j c_j^2}{\sum_{j=1}^{M} c_j^2},$$

(10.39)

$$\chi_j = \begin{cases} 1, & if \quad \alpha_j^k > 0, \quad \forall j \in (i_1,...,i_s) \\ 0, & if \quad \alpha_j^k = 0, \quad \exists j \in (i_1,...,i_s) \end{cases},$$

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

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

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

(10.40)

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

10.5.1. Multi-parameter weighted sensitivity measure

The weighted sensitivity measure introduced in Section 10.4.1 can be generalized for multi-parameter GSA by averaging the univariate index defined in equation (10.15) over all possible values of all other considered parameters:

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

(10.41)

As done for the individual parameter analysis, this global sensitivity index reflects the absolute value of the slope $\partial\Omega/\partial\omega_j$, but averaged over the statistical distributions or weighing

functions of $\omega_1, ..., \omega_N$. Hence, substituting the expansion (10.1) and the definition of the multi-dimensional basis (10.3) into equation (10.41), we obtain:

$$S_{\omega_j}^2 = \int_{\omega_1 \in \Lambda} ... \int_{\omega_N \in \Lambda} \left[\frac{\partial}{\partial \omega_j} \left(\sum_{k=0}^{M} c_k \prod_{l=1}^{N} P_l^{(\alpha_l^k)}(\omega_l) \right) \right]^2 d\Gamma(\omega_1)...d\Gamma(\omega_N), \qquad (10.42)$$

$$\sum_{l=1}^{N} \alpha_l^k \leq M, \quad k = 1,..N.$$

The expression (10.42) can be re-written as:

$$S_{\omega_j}^2 = \int_{\omega_1 \in \Lambda} ... \int_{\omega_N \in \Lambda} \left[\sum_{k=0}^{M} c_k \frac{\partial P_j^{(\alpha_j^k)}(\omega_j)}{\partial \omega_j} \prod_{l=1, l \neq j}^{N} P_l^{(\alpha_l^k)}(\omega_l) \right]^2 d\Gamma(\omega_1)...d\Gamma(\omega_N). \quad (10.43)$$

When exploiting the orthonormality of polynomials $P_l^{(\alpha_l^k)}(\omega_l)$ and integrating over all ω_l ($l \neq j$), equation (10.43) can be simplified to:

$$S_{\omega_j}^2 = \sum_{k=0}^{M} c_k^2 \int_{\omega_j \in \Lambda} \left[\frac{\partial P_j^{(\alpha_j^k)}(\omega_j)}{\partial \omega_j} \right]^2 d\Gamma(\omega_j). \qquad (10.44)$$

Again, introducing the coefficients $b_i^{(\alpha_j^k - 1)}$ as a simple re-collection (see equation (10.19) and other details in section 10.4.2) of the coefficients c_j, from equation (10.44) we obtain:

$$S_{\omega_j}^2 = \sum_{k=0}^{M} c_k^2 \int_{\omega_j \in \Lambda} \left[\sum_{i=0}^{\alpha_j^k - 1} b_i^{(\alpha_j^k - 1)} P_j^{(i)}(\omega_j) \right]^2 d\Gamma(\omega_j). \qquad (10.45)$$

Then, considering the orthonormality of the polynomials $P_j^{(i)}(\omega_j)$ in equation (10.45), the averaged sensitivity index S_{ω_j} can be explicitly expressed as:

$$S_{\omega_j}^2 = \sum_{k=0}^{M} c_k^2 \sum_{i=0}^{\alpha_j^k - 1} \left[b_i^{(\alpha_j^k - 1)} \right]^2 P_j^{(i)}(\omega_j). \qquad (10.46)$$

where the re-collection coefficients $b_i^{(\alpha_j^k - 1)}$ are defined as solution of the coresponding linear system equation (10.22).

This index reflects the influence of a parameter ω_j onto the model output Ω in a similar fashion to the total index defined in equation (10.40). However, our new multivariate index in equation (10.41) does not rely on comparison among different parameters, i.e., it is an absolute measure. This is an advantage over the existing Sobol-based total index which is only a comparative and relative measure. Also, because it work on the derivative with respect to investigated parameter, it maintains the original meaning of sensitivity.

10.6. Illustration

10.6.1. Physical scenario

We will demonstrate our methodology for a contaminant transport problem in a 3D heterogeneous aquifer and the resulting human health risk for an exposed population. We chose an example with simple conditions in order to use analytical solutions and not to burden the presentation of our method with too many details of complex setups, numerical solutions and so forth. The aquifer has a hydraulic conductivity tensor $\mathbf{K}(\mathbf{x})$ and constant effective porosity n_e with $\mathbf{x} = (x_1, x_2, x_3)$. For illustration purposes, we consider flow to be incompressible, single-phased, at steady-state, free of boundary effects and with velocity $\mathbf{u}(\mathbf{x})$ satisfying Darcy law:

$$\mathbf{u}(\mathbf{x}) = -\frac{\mathbf{K}(\mathbf{x})}{n_e} \nabla h, \tag{10.47}$$

where the hydraulic head h is determined from the continuity equation:

$$\nabla \cdot [\mathbf{K}(\mathbf{x}) \nabla h(\mathbf{x})] = 0. \tag{10.48}$$

A tracer with initial concentration C_0 is instantaneously released from a rectangular source volume V_0 under purely advective transport conditions. Under these conditions, the governing equation for contaminant transport is:

$$\frac{\partial C}{\partial t} + u_i \frac{\partial C}{\partial x_i} = 0 \tag{10.49}$$

With the aid of the Lagrangian framework (Dagan, 1987), the solution to equation (10.49) is:

$$C(\mathbf{x}, t) = \int_{V_0} C_0(\mathbf{a}) \, \delta[\mathbf{x} - \mathbf{X}(t; \mathbf{a})] \, d\mathbf{a}, \tag{10.50}$$

where δ is the Dirac operator, $\mathbf{a}$ denotes the initial position of a solute particle within V_0, and $\mathbf{X}$ is the particle trajectory of the solute that follows the local velocities $\mathbf{u}(\mathbf{x})$. In our illustration, we consider $C_0 = 1$ constant within V_0. More details concerning equation (10.50) can be found in Ch. 9 and 10 of Rubin (Rubin, 2003).

10.6.2. Concentration moments

Since we are unable to fully characterize the hydraulic properties of the porous medium, we must treat $\mathbf{K}(\mathbf{x})$, $\mathbf{u}(\mathbf{x})$ and $C(\mathbf{x},t)$ as Random Space Functions (RSF) (Rubin, 2003). For this reason, we will resort to probabilistic concepts in order to estimate the statistical moments of resident concentration. The RSF for the log-conductivity $Y = \ln K$ is statistically characterized by its mean $\langle Y \rangle$, variance σ_Y^2, covariance model C_Y (assumed here exponential) and its integral scale $I_{Y,i}$, where $i = 1$, 2 and 3 for $\mathbf{x} = (x_1, x_2, x_3)$. For our case, we will consider $I_Y \equiv I_{Y,1} = I_{Y,2}$ and $I_{Y,v} \equiv I_{Y,3}$ with anisotropy ratio $f = I_3/I_Y$. Statistical stationarity for Y is assumed. Flow is uniform-in-the-average with mean velocity $\langle \mathbf{u}(\mathbf{x}) \rangle \equiv (U, 0, 0)$.

In this work we will limit ourselves to the first two statistical moments of C for illustration purposes. Under the above mentioned conditions and assumptions, the concentration mean and variance can be expressed as (Rubin et al., 1994):

$$\langle C(\mathbf{x},t) \rangle = C_0 \prod_{i=1}^{3} \psi_i(\mathbf{x},t) \tag{10.51}$$

$$\psi_i(\mathbf{x},t) = \frac{1}{2}\operatorname{erf}\left[\frac{x_i - U_i t + L_i/2}{\sqrt{2X_{ii}(t;\mathbf{a})}} \right] - \frac{1}{2}\operatorname{erf}\left[\frac{x_i - U_i t - L_i/2}{\sqrt{2X_{ii}(t;\mathbf{a})}} \right]$$

$$\sigma_C^2 = \langle C(\mathbf{x},t) \rangle \left[C_0 - \langle C(\mathbf{x},t) \rangle \right] \tag{10.52}$$

where $X_{ii}(t;\mathbf{a})$ is the one-particle displacement covariance function. Note that this expression also assumes that the statistical distribution of particle displacements is Gaussian. In our test case, we will make use of the semi-closed expressions for $X_{ii}(t;\mathbf{a})$ from Dagan (Dagan, 1988) (reproduced in section 10.6.3). Additional details concerning $X_{ii}(t;\mathbf{a})$ are given in Ch. 10 of Rubin (Rubin, 2003). The concentration statistical moments in Eqs. (10.51)-(10.52) are valid for low to mildly heterogeneous aquifers, $\sigma_Y^2 \lesssim 1$ (Bellin et al., 1992).

Note that in the current scenario, the parameter vector ω is represented by the structural parameters in the RSF. These parameters are, e.g., inferred from site characterization data or from expert knowledge gained at geologically similar sites. Because of data scarcity, uncertainty in the estimates of structural parameters needs to be accounted for in the prediction (Kitanidis, 1986; Rubin and Dagan, 1992; Riva and Willmann, 2009; Nowak et al., 2010). For our case, we will consider $\omega = \{I_Y, \sigma_Y^2, U\}$ (see Section 10.2.1). In the following, we

will analyze the sensitivity of concentration moments and a risk-based performance metric, see Eqs. (10.51)-(10.52), with respect to these three and several other parameters. To the best of our knowledge, our illustration is the first application of the PCE technique to the structural parameters of an RSF.

10.6.3. Semi-analytical expressions of particle displacement covariances

We present here the semi-analytical expressions of particle displacement covariances $X_{ii}(t; \mathbf{a})$ for statistically anisotropic conductivity fields. The details concerning the derivation of the following expressions can be found in Dagan (Dagan, 1988) and in Ch. 10 of Rubin (Rubin, 2003). The results are reproduced in this Appendix for completeness. For the longitudinal component ($i = 1$), we have:

$$\frac{X_{11}}{I_Y^2 \sigma_Y^2} = 2\tau + 2\left(\exp[-\tau] - 1\right) \tag{10.53}$$

$$+8f \int_0^\infty \left[J_o(k\tau) - 1\right] \widetilde{\omega}_1(k; f) \, dk$$

$$- \dots 2f \int_0^\infty \left[J_o(k\tau) - \frac{J_1(k\tau)}{k\tau} - \frac{1}{2}\right] \widetilde{\omega}_2(k; f) \, dk;$$

with $\tau = Ut/I_Y$ and:

$$\widetilde{\omega}_1(k; f) = \frac{1}{\left(1 + k^2 - f^2 k^2\right)^2} \tag{10.54}$$

$$-\frac{fk}{\left(1 + k^2 - f^2 k^2\right)^2 \sqrt{\left(1 + k^2\right)}} -$$

$$\dots \frac{fk}{2\left(1 + k^2 - f^2 k^2\right)\left(1 + k^2\right)^{3/2}};$$

$$\widetilde{\omega}_2(k; f) = \frac{f^3 k^3 \left(f^2 k^2 + 5 + 5k^5\right)}{\left(f^2 k^2 - 1 - k^2\right)^3 \left(1 + k^2\right)^{3/2}} \tag{10.55}$$

$$+\frac{1 + k^2 - 5f^2 k^2}{1 + k^2 - f^2 k^2}.$$

In equation (10.53), J_o and J_1 denote the Bessel functions of zeroth and first order. The transverse particle displacement covariances are ($i = 2 = 3$):

$$\frac{X_{22}}{I_Y^2 \sigma_Y^2} = -2f \int_0^\infty \left[\frac{J_1(k\tau)}{\tau} - \frac{k}{2} \right] \tag{10.56}$$

$$\cdots \left[\frac{f^3 k^2 \left(f^2 k^2 - 5k^2 - 5 \right)}{\left(f^2 k^2 - 1 - k^2 \right)^3 \left(1 + k^2 \right)^{3/2}} + \frac{1 + k^2 - 5 f^2 k^2}{k \left(1 + k^2 - f^2 k^2 \right)^3} \right] dk;$$

$$\frac{X_{33}}{I_Y^2 \sigma_Y^2} = -4f \int_0^\infty \frac{\left[J_o(k\tau) - 1 \right]}{\left(f^2 k^2 - 1 - k^2 \right)^2} \tag{10.57}$$

$$\cdots \left[\frac{1}{2} + \frac{2 f^2 k^2}{\left(1 + k^2 - f^2 k^2 \right)} + \frac{fk \left(f^2 k^2 + 3 + 3k^2 \right)}{2 \left(f^2 k^2 - 1 - k^2 \right) \left(1 + k^2 \right)^{1/2}} \right] dk.$$

10.6.4. Human health risk

In most cases, decision makers are interested in quantifying adverse effects in human health due to contaminated groundwater exposure. For this case, the environmental performance metric of interest is human health risk, denoted by r. In contrast to the concentration C, where the input parameters are related to the hydrogeological aspects of the problem, r depends not only on the hydrogeological parameters but also on the physiological and behavioural parameters of the exposed individual.

For illustration purposes, we will consider the increased lifetime cancer risk model from the EPA (USEPA, December 1989,D), although many other risk models exist as discussed in the literature (de Barros et al., 2011; Maxwell and Kastenberg, 1999; Siirila et al., 2012). The increased lifetime cancer risk formulation for the groundwater ingestion pathway is given by:

$$r = a\overline{C} \tag{10.58}$$

where $\overline{C}$ is the maximum running average over the exposure duration ED (years) defined in (Maxwell and Kastenberg, 1999)

$$\overline{C} = \max \left[\frac{1}{ED} \sum_{t}^{t+ED} C(t) \right]_{t=0}^{\infty} \tag{10.59}$$

and a is given by

$$a = CPF \times \frac{IR}{BW} \times \frac{ED \times EF}{AT} \tag{10.60}$$

where CPF (mg kg^{-1} day^{-1}) is the cancer potency factor, IR (L/day) is the ingestion rate of tap water, BW (kg) is the body weight, EF (days/year) denotes exposure frequency and AT (days) is the averaging time. It is out of the scope of this chapter to perform a detailed analysis of each of these parameters present in equation (10.60) and, for simplicity, we will only work with a as a single health parameter. Although each of the parameters given in a are uncertain and vary from individual to individual (see discussion in (Maxwell and Kastenberg, 1999)), we will only concentrate on the bulk uncertainty in a.

10.6.5. Results and discussion

In this Section, we will apply our framework introduced in Section 10.3 for the contaminant transport problem and human health risk analysis described in Sections 10.6.1 to 10.6.4. We will illustrate our approach for single-parameter GSA (Section 10.4) and multi-parameter GSA (Section 10.5). Let us at first consider two predictions: the concentration mean and variance, as mentioned in Section 10.6.2. We will quantify the influence of the model parameters $\omega = \{I_Y, \sigma_Y^2, U\}$ (i.e. $N = 3$) on the concentration mean and variance at a dimensionless fixed location $\mathbf{x} = (10, 0, 0)$. Corresponding to the presented framework, the parameters I_Y, σ_Y^2, U are denoted as ω_1, ω_2 and ω_3, respectively.

According to Step 1 of the algorithm presented in Section 10.3, we first characterize all input parameters. Here, we assume for a simple example that the PDFs of the parameters within ω are as follows:

- For ω_1: $\omega_1 = 1 + 2\omega_1'$, with ω_1' following a beta distribution with $\alpha = 2$ and $\beta = 2$;

- For ω_2: Uniformly distributed within the interval $[0.1, 0.7]$;

- For ω_3: Log-normally distributed with $\mu = 3.6$ and $\sigma = 0.3$.
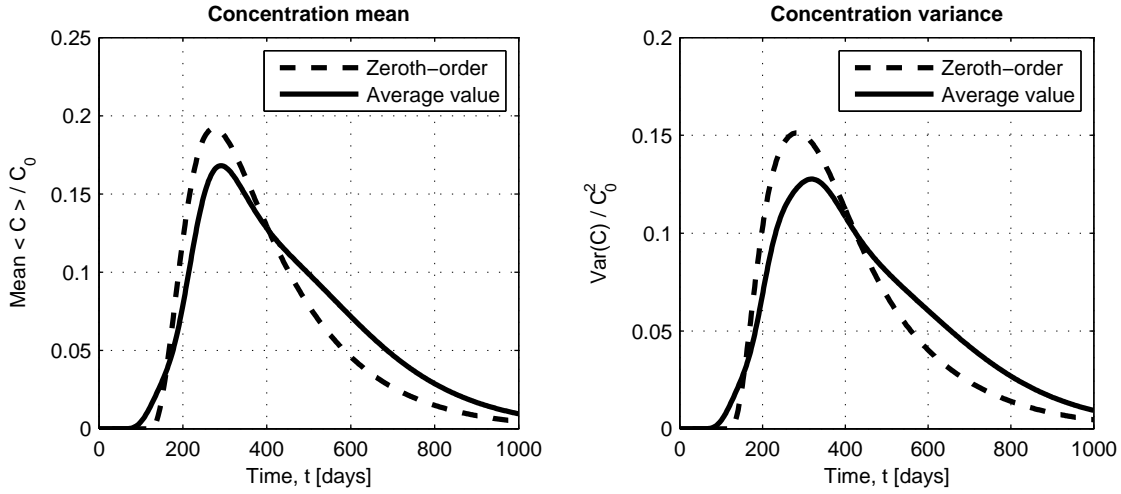
Figure 10.2.: Normalized concentration mean (left plot) and variance (right plot): Curves obtained for the mean values of parameters (dashed line) and for the averaged behavior of the system over all possible values of ω (solid lines).

These PDFs are used in the Step 1 to calculate the corresponding raw moments of input data.

According to Step 2 of our algorithm (Section 10.3), we compute the coefficients of the orthonormal polynomial basis using equation (10.4) and equation (10.6) for given raw moments and for the desired degree of expansion d. Also, we consider the response surface of the model (Eqs. (10.51) and (10.52)) according to equation (10.1) and equation (10.13). Immediately, statistical quantities can be extracted. For example, Figure 10.2 shows the behavior of the physical problem obtained with the second degree expansion ($d = 2$). The dashed lines illustrate the concentration mean and variance, which were computed for the average value of all parameters in ω (i.e., a zeroth-order approximation to the expected value). The solid lines represent the averaged behavior of the system at 2nd order, averaged over all possible values of ω using equation (10.38).

Qualitative sensitive analysis can be done directly using the model expansion (response surface) obtained from Step 2. Figures 10.3 and 10.4 show global sensitivities of the time and level of the peak in $\langle C(t) \rangle$ with respect to the relevant parameters. In all figures, dashed lines represent linear single-parameter GSA, solid lines are for non-linear single-parameter GSA, and the scattered points visualize realizations of the multi-parameter model response. Figure 10.3 illustrates the sensitivity of the peak time of $\langle C(t) \rangle$ (see equation 10.51), denoted as t_{peak}, to each individual parameter in $\omega = \{I_Y, \sigma_Y^2, U\}$. Analyzing Figure 10.3 closely, we see that t_{peak} is much more sensitive to the mean longitudinal velocity U than to the other
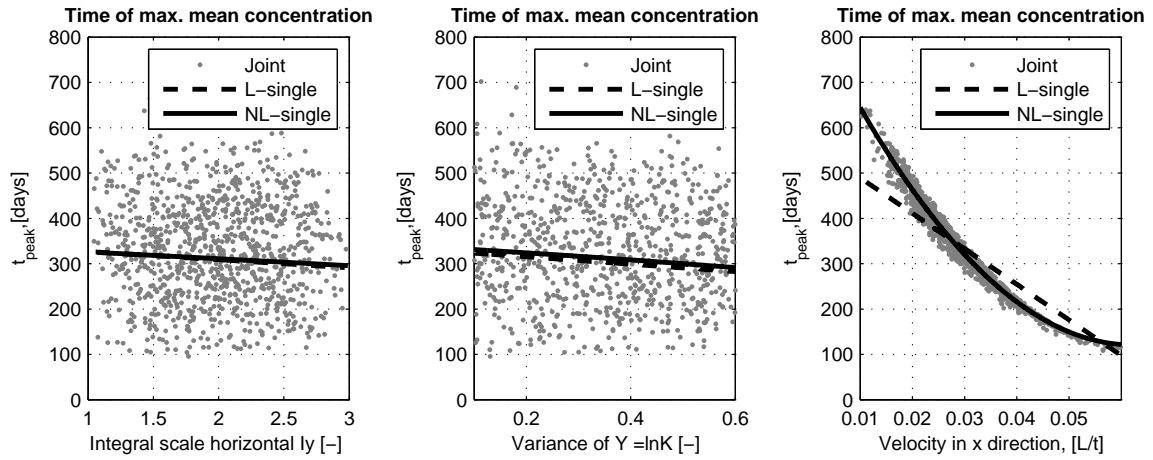
Figure 10.3.: Sensitivity of time of maximal mean concentration: dashed line is linear single parameter GSA, solid line is non-linear single parameter GSA (here: second order) and points represent realizations of the multi-parameter model response

two parameters (I_Y and σ_Y^2). This is an outcome of the physical scenario under investigation, where the uniform-in-the-average flow almost completely determines the motion of the plume's centroid, while I_Y and σ_Y^2 can merely modify the shape of $\langle C(t) \rangle$, and only slightly shift the time of peak arrived relative to the time of bulk arrival.
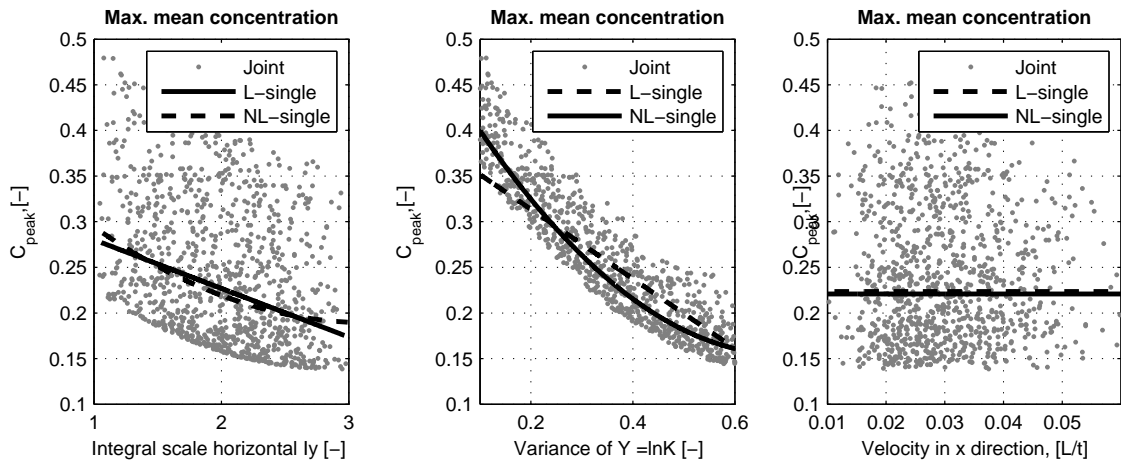
Figure 10.4.: Sensitivity of maximal mean concentration: dashed line is linear single parameter GSA, solid line is non-linear single parameter GSA and points represents multi-parameter model response

On the other hand, if we perform the sensitivity analysis for the peak level of the mean concentration, denoted by C_{peak} (see Figure 10.4), we observe a completely different picture

than in Figure 10.4. In this case, having knowledge of the level of heterogeneity (σ_Y^2) and its correlation length scale (I_Y) helps to determine the peak level of the mean concentration. This is because σ_Y^2 and I_Y determine the magnitude of the plume dispersion in the system. This is a clear example of how the response of parameters and corresponding required characterization efforts vary according to the quantity to be predicted, see (de Barros et al., 2009; Nowak et al., 2010).

At the last phase of analysis, Step 3 of our algorithm (Section 10.3), the desired quantitative sensitivity information is extracted from the polynomial response surface. We denote the parameters I_Y, σ_Y^2 and U with subscripts 1, 2 and 3 of the Sobol (and sensitivity) indices, respectively. Table 10.1 represents the ranked Sobol indices for the time of maximal mean concentration (t_{peak}) and peak level of mean concentration (C_{peak}). Total sensitivity

Sobol index	S_{I_Y}	$S_{\sigma_Y^2}$	S_U	$S_{I_Y \sigma_Y^2}$	$S_{I_Y U}$	$S_{\sigma_Y^2 U}$
Value for t_{peak}	0.004	0.016	0.975	8×10^{-4}	6×10^{-4}	0.003
Rank for t_{peak}	3	2	1	6	5	4
Value for C_{peak}	0.111	0.887	6×10^{-9}	0.002	9×10^{-10}	1×10^{-10}
Rank for C_{peak}	2	1	4	3	5	6

Table 10.1.: Sobol indices for time of maximal mean concentration (t_{peak}) and maximal mean concentration (C_{peak}).

Total sensitivity index	$S_{I_Y}^T$	$S_{\sigma_Y^2}^T$	S_U^T
Value for t_{peak}	0.005	0.021	0.980
Rank for t_{peak}	3	2	1
Value for C_{peak}	0.2113	0.889	8×10^{-9}
Rank for C_{peak}	2	1	3

Table 10.2.: Total sensitivity indices for time of maximal mean concentration (t_{peak}) and maximal mean concentration (C_{peak}).

Weighted sensitivity index	S_{I_Y}	$S_{\sigma_Y^2}$	S_U
Value for t_{peak}	17.857	81.6645	8×10^3
Rank for t_{peak}	3	2	1
Value for C_{peak}	0.053	0.381	2×10^{-5}
Rank for C_{peak}	2	1	3

Table 10.3.: Weighted first-order sensitivity indices for time of maximal mean concentration (t_{peak}) and maximal mean concentration (C_{peak}).

Weighted sensitivity index	S_{I_Y}	$S_{\sigma_Y^2}$	S_U
Value for t_{peak}	14.246	66.163	2×10^4
Rank for t_{peak}	3	2	1
Value for C_{peak}	0.143	0.958	7×10^{-4}
Rank for C_{peak}	2	1	3

Table 10.4.: Weighted second-order sensitivity indices for time of maximal mean concentration (t_{peak}) and maximal mean concentration (C_{peak}).

indices are presented in Table 10.2. The weighted sensitivity indices for the single parameter analysis introduced in Eqs. (10.26) and (10.36) are given in Tables 10.3 and 10.4 (first and second-order, respectively). These results quantify to what extent the time of peak mean concentration is sensitive with respect to U. On the other hand, the magnitude of the mean concentration is very sensitive to σ_Y^2.

These sensitivity indices are useful to quantify the simultaneous influence of model parameters, especially when the number of parameters becomes large and visualization of a multivariate model response like in Figure 10.4 becomes unfeasible and confusing. To illustrate this, we will additionally consider an uncertain contaminant spill location and uncertain parameters in the model for health risk. Let us include the position x_s, y_s, z_s of the source with the source volume V_0 and the slope of the risk model a as additional parameters ω_4, ω_5, ω_6 and ω_7 into our analysis.

- For ω_4: $\omega_4 = 9 + 2\omega_4'$, where ω_4' is beta distributed with $\alpha = 2$ and $\beta = 2$;

- For ω_5 and ω_6: $\omega_{5,6} = -1 + 2\omega_{5,6}'$, where $\omega_{5,6}'$ is beta distributed with $\alpha = 2$ and $\beta = 2$.

- For ω_7: ω_7 is lognormal distributed with $\mu = 1.7$ and $\sigma = 0.14$.

Now, we look at the sensitivity of the probability of a critical heath effect with respect to these seven parameters using equation(10.58). Again, all sensitivity indices can be constructed as it was shown above. For brevity, we will focus on the total and weighted indices only. Figure 10.5 illustrates the total (left plot) and weighted (right plot) sensitivity indices at different degrees of expansion. This figure shows the overall influence of all model parameters on the total human health effect (final prediction), which implicitly includes a time integral over the concentration history (see equation 10.58 and 10.59). For this particular case, we observe that risk is more sensitive towards hydrogeological parameters than to the health parameters. However, results indicate that the health component is still as important
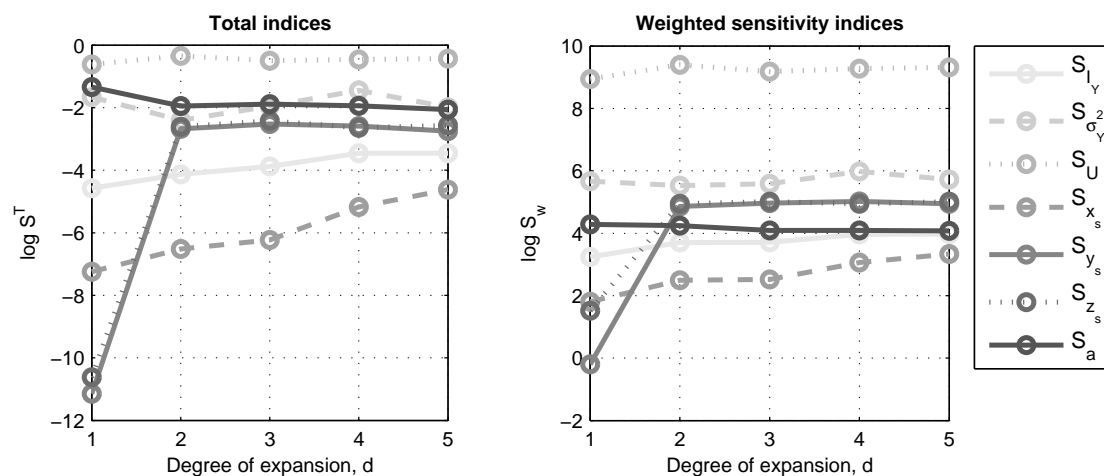
Figure 10.5.: Convergence of total (left plot) and weighted (right plot) sensitivity indices for the health risk prediction.

when compared to the uncertainty of the source location. As shown in Figure 10.5, the flow distance and resulting macrodispersive effects are so large that effects related to the source offsetting matter less. The results in Figure 10.5 illustrate how our framework can be used in problems with multiple sources of uncertainty. As discussed in (de Barros and Rubin, 2008; de Barros et al., 2009), health risk related problems are challenging because uncertainty is present in both the physical problem (flow and transport) and in the human physiological response and methods are needed to help decision-makers allocate resources towards uncertainty reduction. Figure 5 also re-emphasizes how GSA results change according to the goal of prediction (concentration versus risk, see discussion in (de Barros et al., 2011)).

In the case study related to health risk (Figure 10.5), our new weighted measure converges faster in comparison to Total indices with increasing order of the expansion. Total indices show the main reaction of the model to the analyzed parameters, but as relative quantities do not provide faster stabilization in the ranking of analyses parameters in comparison to absolute weighted indices. The presented approach for GSA provides a relatively accurate approximation even for moderate orders (see also the convergence study for aPC in (Oladyshkin et al., 2011c)) and work already with a small set of model evaluations. Thus, global and weighted sensitivity analysis can already be performed at computational costs that are only slightly larger than those of local analysis.

10.7. Conclusions

In this work, we have presented an alternative method for global sensitivity analysis (GSA). The methodology shown is based on the arbitrary polynomial chaos expansion (aPC) . The aPC approach provides improved convergence in comparison to classical PCE techniques, when applied to input distributions that fall outside the range of classical PCE (Oladyshkin et al., 2011a,c). Compared to existing polynomial-based GSA methods, it can accommodate for all types of statistical distributions or weighting functions of the input parameters. We denote this approach as *data-adaptive* because the aPC method can be applied even in situations where precise statistical information (e.g., known parametric distributions) for the input parameters is not available. If desired, the method can work directly with raw sampled data sets to represent the uncertainty and possible variational ranges of input data. The presented methods allow experts to choose freely of technical constraints the shapes of their statistical assumptions, and they allow to align the complexity and order of analysis with the reliability and detail level of statistical information on the input parameters.

The proposed method incorporates the full range of possible simulation outcomes for the investigated model parameters as it approximates the model's full response surface by multivariate polynomials. We have extended an existing polynomial-based method to compute Sobol indices to the much more general aPC framework. While Sobol indices only allow a comparison between several model parameters, we also derive a new sensitivity measure in the context of polynomial model approximation, which allows to quantify sensitivity even for individual parameters. While Sobol indices look at the contributions from individual parameters to the energy (variance) of the model, we look at the energy norm of derivatives of the model w.r.t individual parameters. The resulting sensitivity measure shows better convergence properties than Sobol analysis and maintains he original derivative based character of sensitivity analysis.

We made use of an analytical formulation of GSA to investigate the sensitivity of a single parameter on a model prediction. As for multi-parameter analysis, a hybrid analytical-numerical 3-step algorithm is presented. These features contribute to the computational efficiency of our approach. Compared to earlier works that related GSA to classical PCE, we emphasize a more engineering-like language as compared to otherwise intense mathematical derivations. The framework presented in this chapter is limited for statistically independent random variables, which reflect the current state-of-art with PCE techniques. Construction of a polynomial basis for statistically dependent random variables beyond linear dependence is an important topic for future research. In the current work, we suggest the non-intrusive

collocation approach for construction of the aPC-based expansion, however other projection techniques can be used as well, if desired.

In summary, we proposed a rational and computationally efficient framework that allows modelers and decision makers to better judge the relative and absolute importance of model parameters. It allows modelers and decision makers to better allocate resources towards uncertainty reduction in model predictions or model-based decisions by goal-oriented data acquisition. For instance, our method could be applied in human health risk assessment where both health-related and hydrological parameters are sources of uncertainty (Maxwell and Kastenberg, 1999; de Barros and Rubin, 2008). In such applications, decision makers are interested in knowing the relative impact of uncertainty of each input parameter on risk in order to better invest (and allocate) characterization efforts between hydrogeology and behavioral parameters to obtain more reliable estimates in human health risk (Benekos et al., 2007; de Barros and Rubin, 2008; de Barros et al., 2009).

The methodology was illustrated for a 3D groundwater contamination problem in a heterogeneous aquifer and the resulting human health risk. Although our illustrative example relied on a relatively simple physical-mathematical formulation, more complicated problems (e.g., highly heterogeneous aquifers, reactive transport, uncertain boundary conditions, etc.) can be tackled through the same procedure.

11. Bayesian updating based on data-driven polynomial chaos

Model calibration and history matching are important techniques to adapt simulation tools to real-world systems. When prediction uncertainty needs to be quantified, one has to use the respective statistical counterparts, e.g. Bayesian updating of model parameters and data assimilation. For complex and large-scale systems, however, even single forward deterministic simulations may require parallel high-performance computing. This often makes accurate brute-force and non-linear statistical approaches infeasible. We propose an advanced framework for parameter inference or history matching based on the arbitrary polynomial chaos expansion (aPC) and strict Bayesian principles. Our framework consists of two main steps. In step one, the original model is projected onto a mathematically optimal response surface via the aPC technique. The resulting response surface can be viewed as a reduced (surrogate) model. It captures the model's dependence on all parameters relevant for history matching at high-order accuracy. Step two consists of matching the reduced model from step one to observation data via Bootstrap filtering. Bootstrap filtering is a fully nonlinear and Bayesian statistical approach to the inverse problem in history matching. It allows to quantify post-calibration parameter and prediction uncertainty, and is more accurate than Ensemble Kalman filtering or linearized methods. Through this combination, we obtain a statistical method for history matching that is accurate, yet has a computational speed that is more than sufficient to be developed towards real-time application. We motivate and demonstrate our method on the problem of CO_2 storage in geological formations, using a low-parametric homogeneous 3D benchmark problem. In a synthetic case study, we update the parameters

of a CO_2/brine multiphase model on monitored pressure data during CO_2 injection.

11.1. Introduction

11.1.1. Modeling carbon dioxide storage

We would like to motivate our work on the example of modeling CO_2 storage in geological formations. CO_2 storage is currently being discussed intensively as an interim technology with a high potential for mitigating CO_2 emissions (e.g. (IPCC, 2005)). In recent years, great research efforts have been directed towards understanding the processes in CO_2 storage. The multiphase flow and transport processes involved are strongly non-linear. They include phase changes in the region of the critical point and effects such as gravity-induced fingering and convective mixing as well as geo-chemical and geo-mechanical processes, etc.

In order to describe the space-time evolution of injected CO_2 plumes, to analyze the influence of potentially leaky abandoned wells and to investigate possible geo-mechanical failure of reservoirs, (semi-) analytical solutions have been derived by Nordbotten et al. and others (Nordbotten et al., 2005a, 2009). A comparison study of various simplifying semi-analytical models with complex numerical simulation tools was performed by Ebigbo et al. (Ebigbo et al., 2007). The analysis in (Birkholzer et al., 2009) focused on the effects of large-scale CO_2 leakage through low-permeability layers. Various optimization strategies for monitoring surface leakage using near-surface measurement approaches were developed in (Cortis et al., 2008). These studies are cited here merely to provide a few examples. More detailed reviews are provided in, e.g., (IPCC, 2005), (Class et al., 2009), (Ebigbo et al., 2007).

Modeling underground CO_2 storage involves many conceptual and quantitative uncertainties (Hansson and Bryngelsson, 2009a). Class et al. (Class et al., 2009) published a benchmark study that compares a number of mathematical and numerical models of varying complexity. However, the lack of information on subsurface properties (porosity, permeability, etc.) may lead, depending on the specific question at hand, to parameter uncertainties up to a level where parameter uncertainties dominate or even override the influence of secondary physical processes. In (Oladyshkin et al., 2011a; Oladyshkin and Nowak, 2012a), the authors of the current study pointed out, that efforts invested in improved physical conceptualization, numerical codes and stochastic modeling can easily be overwhelmed by error through human subjectivity in data interpretation at a very early stage of modeling, and proposed a purely data-driven approach to overcome this problem.

In the development of CO_2 injection as a large-scale interim solution, our ability to quantify its uncertainties and risks will play a key role. Sensitivity analyses (e.g. (Birkholzer et al., 2009), (Kopp et al., 2009)) for CCS have been applied up to the present. Fault-tree analyses have been used to identify risks through different factors (Wildenborg et al., 2005), but have not yielded quantitative risk information. A significant part of the applied and scientific community still refrains from explicitly considering uncertainty in modeling, although the corresponding arguments are discussed and rejected one by one in (Pappenberger and Beven, 2006). A key hindrance to quantitative risk assessment is that current numerical simulation models are often inadequate for stochastic simulation techniques based on brute-force Monte Carlo simulation and related approaches (e.g. (Maltz and Hitzl, 1979), (Robert and Casella, 2004)), because even single deterministic simulations may require parallel high-performance computing.

This triggered an urgent need to develop reasonably fast stochastic approaches for probabilistic risk assessment of CO_2 sequestration. In (Oladyshkin et al., 2011b), the authors of the current study have pioneered the application of massive stochastic model reductions based on the polynomial chaos expansion (PCE) to CO_2 storage and developed a novel approach to join robust design ideas with the PCE technique. Statistical uncertainty has been combined with scenario uncertainty in the recent work (Walter et al., 2012) to estimate the risk related to the brine migration resulting from CO_2 injection into saline aquifers. In the current study, we suggest a history matching (or model calibration, parameter inference) framework based on Bayesian updating, which is both accurate and sufficiently fast to be applied to complex large-scale models. Putting history matching into the Bayesian updating context allows to combine available data with prior expert judgment and to quantify parametric uncertainty after calibration. In this context, our approach will help to override the subjectivity of prior assumptions on parameter distribution by incorporating observed data.

11.1.2. Inverse modeling and history matching

Inverse modeling and history matching to past production data is an extremely important issue in order to improve the quality of prediction. When cast into a stochastic/statistical framework, the results also provide information on the remaining parameter and prediction uncertainty (e.g., (Lia et al., 1997)). Also, the statistical framework allows to assess the worth of monitoring data and then to optimize future monitoring strategies (e.g. (Leube et al., 2012a)). This would help to raise studies on CO_2 plume monitoring (e.g. (Cortis et al., 2008)) to the level of formal optimization.

The accuracy of inversion or history matching depends on the quality of the established physical model (including, e.g. seismic, geological and hydrodynamic characteristics, fluid properties etc.), and on the accuracy of the involved parameter calibration, stochastic inversion or data assimilation techniques. The quality will also depend on the computational efficiency of all involved methods (Oliver and Chen, 2011), because too large computational time would have to be mitigated by compromises in numerical or inversion accuracy.

The principal challenge of history matching is that the full reservoir behavior has to be reproduced from local measurements (Gavalas et al., 1976). The reliability of history matching is increasing with the number of available observations. Moreover, besides the proper definition of the involved geological and physical models, the list of uncertain parameters has to be defined adequately for further investigation (Kravaris and Seinfeld, 1985). Compromises in this task are indispensable. On the one hand, a too short list of parameters leads to a very parsimonious model that allows robust parameter inference, but simply can not accurately reproduce the observation data. On the other hand, a too long list of parameters leads to a very complex task which cannot be afforded, computationally, and one can run into under-determined or otherwise ill-posed problems (Ewing et al., 1994). In this stage, an expert opinion on relevant parameters that are useful for calibration, probably combined with sensitivity analyses (see e.g. (Saltelli et al., 2008; Oladyshkin et al., 2011d; Cominelli et al., 2007)) or regularization choices is very important.

History matching is very well known in the field of reservoir engineering (Makhlouf et al., 1993). Traditionally, an iterative manual process of trial and error (see e.g. (Williams et al., 1998)) is applied to adjust the reservoir geological model in order to reproduce past observations of oil or gas production. Such a manual approach is very popular among experts in reservoir engineering and demands a very strong understanding of geology and processes. However, the non-trivial and non-linear interaction of the matched parameters can complicate the history matching procedure a lot (Oliver et al., 2001). Instead of the manual technique, formal optimization methods such as gradient search or the adjoint method (see (Gao and Reynolds, 2006; Li et al., 2003; Rodrigues, 2006)) can be applied. Unfortunately, the mentioned optimization approaches often lead to high computational costs and cannot be easily applied for complex real-world tasks. The state-of-the-art and its recent progress for history matching is presented in detail in the review paper (Oliver and Chen, 2011).

Another important point about history matching techniques is that they can produce non-unique solutions (Oliver et al., 2008), which means that several virtual models and parameter sets can match the observation data equally well. In fact, this problem is common to most in-

verse problems (Sun, 1999; Tarantola, 2005). Stochastic approaches can handle such type of uncertainty occurring during the matching procedure without the need to introduce regularization or to artificially restrict the parameter space. Their result is a probability distribution of possible parameters sets instead of a single best estimation. As an improvement of classical optimization, several successful stochastic approximation methods have been adapted for the history matching problem (Bangerth et al., 2006; Gao and Reynolds, 2007; Li et al., 2011).

However, stochastic approaches are more expensive than classical optimization-based (deterministic) calibration techniques, because they need to explore the full range of possible model outcomes with many model runs. In particular, this requires to draw samples from the conditional distribution of the parameters as equally likely calibrated parameter sets. This can be done, e.g., via Markov chain Monte Carlo (Gilks et al., 1996), Bootstrap filtering (Smith and Gefland, 1992), GLUE (Leube et al., 2012a) or rejection sampling (Smith and Gefland, 1992) . A particular reason for the high computational costs is that, when exploring the full range of possible model outcomes, many model runs are rejected upon comparison with the data. Therefore, the high computational costs of forward modeling get multiplied by large factors in stochastic calibration methods.

The overall efficiency challenge of history matching or inverse modeling can be subdivided into to two principal parts. The first part of the challenge consists in the acceleration of forward modeling itself, using reduced forward models (e.g. response surfaces, surrogate models, low-parametric representation, etc.). The second part of the challenge consists in developing efficient searching algorithms or integration rules for inverse modeling.

To accelerate forward modeling, fast streamline forward simulations have been used in (Daoud, 2004). An optimal representation of uncertain heterogeneous parameter fields in terms of a Karhunen-Loeve expansion is presented in (Sarma et al., 2006), leading to a reduced number of parameters to be treated. Later, a kernel-based principal component analysis was presented in (Sarma et al., 2008) as an alternative to the standard Karhunen-Loeve expansion. That method is capable to parametrize large-scale non-Gaussian, non-stationary random fields and to reproduce complex geological structures. In low-parametric inverse problems, surrogate models can be constructed with response surface techniques. For example, experimental design techniques were applied in the paper (Zabalza-Mezghani et al., 2004) to build an accurate proxy (surrogate) model using polynomials of second order. So-called proxy functions based on polynomials combined with multi-dimensional kriging have been used to approximate the output of a flow simulator in (P.A. and Smørgrav, 2008). Para-

metric response surfaces have been used in (Feraille and Marrel, 2012) to approximate the reservoir production forecasts and obtain their probabilistic distribution by propagating the remaining posterior uncertainty of input parameters. Recently, a distance-based stochastic technique has been presented in (Scheidt et al., 2011).

The Ensemble Kalman filter (EnKF) method is one of the simplest yet most successful ways to transfer Bayesian theory (see e.g. (Smith and Gefland, 1992)) to practice for model updating and forecasting. The EnKF (Evensen, 2006) is derived from a first-order second-moment approximation of error propagation for Bayesian updating. The Special Issue in Computational Geosciences (Naevdal et al., 2011) was fully devoted to (Ensemble) Kalman filtering for model updating. As practically successful attempt to accelerate inverse modeling, the EnKF recently received a lot of attention for history matching (e.g. (Aanonsen et al., 2009; Naevdal et al., 2005)). It is a comparatively cheap method that can generate reasonable history-matched models for real fields. Due to its foundation on first-order second-moment analysis, it is optimal only if all involved model parameters, model states and data follow a joint multi-Gaussian distribution, and if the used ensemble is sufficiently large for accurate covariance estimation. For example, the paper (Liu and Oliver, 2005) address the problem of matching production data by the EnKF for updating facies delineations in reservoir models. The recent article (Pajonk et al., 2012) offers an alternative way of computing a linear Bayesian estimator, which allows updating of non-Gaussian quantities. It has been shown that, although no formal linearization is involved, the EnKF contains an implicit form of linearization (Nowak, 2009a). Thus, the EnKF has a theoretical limitation which does not allow it to deal with strongly non-linear problems. For example, the papers (Wang et al., 2010) and (Zafari and Reynolds, 2007) pointed out that the EnKF suffers from non-linearity and inconsistency in matching. The univariate deviation from Gaussianity can be removed by empirical data transformation (e.g. (Hendricks Franssen and Kinzelbach, 2009; Schoeniger et al., 2012)), but the multivariate structure still poses a problem.

Therefore, we believe that further advancements can be achieved by using more accurate non-linear approaches to Bayesian updating, when combined with sufficiently accurate (and more expansive) model reduction techniques. In fact, techniques based on response surfaces or other surrogate models have lately been combined with more accurate versions of Bayesian updating. In (Jin, 2008) and (Marzouk et al., 2007), the polynomial chaos expansion has been applied to accelerate Bayesian updating via Markov chain Monte Carlo (MCMC) methods. Even the EnKF has been combined with polynomial chaos expansion for accelerated and accurate computation of the required covariance matrices, see (Saad and Ghanem, 2009), (He et al., 2011) and also related work (Pajonk et al., 2012).

To summarize, we would like to point out that the methodology for history matching has progressed from manual procedures to automatic optimization approaches. During the last years, we see a transition from classical optimization to statistical inference based on Bayesian principles for uncertainty quantification. Also we can observe that, due to the high computational demands of reservoir models especially in the stochastic framework, the use of surrogate models in history matching received a quickly increasing attention in the very recent years.

The most promising novel approach seems to be the combination of PCE-based response surfaces with EnKF filtering or with more accurate but more expensive implementations of Bayesian updating, such as MCMC. However, more accurate Bayesian updating approaches incorporate higher order stochastic information on the input parameters (e.g., in form of high-order statistical moments) (Ghanem and Spanos, 1991; Le Maitre and Knio, 2010; Wiener, 1938) which goes beyond the scope of the Gaussian assumption. This leaves the optimality range of the EnKF. Also, the Gaussian assumption in stochastic inversion typically goes along with (possibly implicit) linearization of model dependencies.

Our key motivation is that more accurate Bayesian updating also requires to work with high-order approximations of the involved models than linear ones. Using a more accurate updating rule for higher moments would be inadequate when the involved surrogate model is too inaccurate (e.g. by linearizing strongly non-linear model dependencies). As a consequence, model reduction techniques also need to retain the non-linearity of models at sufficiently high orders.

11.1.3. Approach and contributions

The goal of this work is to further advance statistical (Bayesian) model calibration, here working on the example of history matching. We wish to find an even better compromise between computational efficiency and accuracy of the inversion. The resulting framework will take into consideration the non-linearity of the forward model and of inversion, and will provide a cheap but highly accurate tool for reducing prediction uncertainty. We will also aim at a consistent use and processing of high-order statistical moments for the considered uncertain model parameters. Our approach is to combine fully accurate Bayesian updating via Bootstrap filtering (or, alternatively, via MCMC) with a model reduction based on the arbitrary polynomial chaos expansion (aPC) technique (Oladyshkin and Nowak, 2012a; Oladyshkin et al., 2011a; Oladyshkin and Nowak, 2012b), see details in Section 11.2.

The aPC is a so-called data-driven generalization of the PCE technique (see details in Section 11.2.1). It provides the necessary flexibility to handle arbitrary distributions of model parameters, including high-order statistical moments. Due to this flexibility, it can also handle the conditional parameter distributions that occur after Bayesian updating (see Section 11.2.2), if one would desire to renew the PCE-based model reduction after Bayesian updating. A second advantage of the aPC approach is that it provides improved convergence with respect to increasing order of expansion (Oladyshkin and Nowak, 2012a) of the surrogate to the original model in comparison to more classical PCE techniques, when applied to input distributions that fall outside the range of classical PCE. Thus, the aPC allows efficient allocation of computational resources for constructing of the surrogate model.

Bootstrap filtering (BF) is the most direct yet simple numerical implementation of Bayes theorem, based on brute-force Monte-Carlo. It approximates the conditional PDF by a sufficiently large ensemble of realizations, and it is exact at the limit of infinite ensemble size. It can account for arbitrary non-linear model equations and for arbitrary distribution shapes in comparison to (Ensemble) Kalman Filters. Hence, BF is a perfect match for combination with the aPC technique. This combination is superior to the previous combinations of the classical PCE with (Ensemle) Kalman Filtering, because the latter two components are both optimal only for (multi-) Gaussian distributions.

Technically, the conditional ensemble is obtained by simple rejection sampling (Smith and Gefland, 1992) applied to an ensemble of parameter vectors drawn from the prior PDF, i.e., to the parameter distributions assumed before incorporating the calibration or history matching data. However, an accurate representation of conditional statistics for model parameters and responses demands a sufficiently large size of the simulated conditional sample. This can be a problem if the used data set is large and accurate, because then the acceptance probability in the rejection sampling approaches to zero. For that reason, BFs can be feasible for very complex models, only when extremely fast evaluation techniques for the response surface are available. For the polynomial response surface resulting from the aPC, this is fulfilled in cases where small acceptance probability still pose a problem, BF can simply be exchanged for more efficient MCMC methods, e.g., (Vrugt et al., 2009).

Overall we expect that (1) thanks to the computational efficiency caused by the improved convergence of the aPC and (2) due to the accuracy of Bootstrap filtering, Bayesian updating for history matching or more general data assimilation can be developed further towards real time even for complex or large-scale simulation models. For example, we demonstrate in Section 11.4 that, four our computational example, it takes only about 20 seconds of CPU

time to perform the actual Bayesian updating step for history matching.

In Section 11.4, we demonstrate a straightforward implementation of aPC-based BF on the benchmark problem of CO_2 leakage taken form (Class et al., 2009) and described in Section 11.3. We apply our method to match the model repose to synthetic pressure data monitored during CO_2 injection. However, the surrogate model may be very inaccurate and lead to wrong results, when the prior information is strongly offset against reality. This is caused by the fundamental property of all PCE techniques that the error of approximation is lowest where the (prior) probability density is high, i.e., large errors may occur in low-probability regions. For that case, we introduce an advanced iterative approach for aPC-based BF in Section 11.5. It allows to perform Bayesian updating even in the case where the prior assumptions on model parameters are far from reality, such that the response surface has to be re-iterated in order to be accurate in the relevant regions of high posterior probability. The iteration can be seen as an extension of statistical inversion based on successive linearization to successive high-order expansion. Statistical inversion via successive linearization can be found, e.g., in Extended Kalman Filtering, in iterative Ensemble Kalman Filters, or in the quasi-linear geostatistical approach (Evensen, 2006; Kitanidis, 1995; Gu and Oliver, 2007; Nowak, 2009a).

11.2. Bootstrap filtering on the polynomial chaos expansion

As outlined in Section 11.1.3, we propose an advanced stochastic inversion framework for history matching based on a recent generalization of the polynomial chaos expansion (PCE) and Bootstrap filtering. Hence, our framework consists of two main steps: a massive but high-order accurate stochastic model reduction via the arbitrary polynomial chaos expansion (aPC) (Section 11.2.1) and accurate numerical Bayesian updating of the reduced model via Bootstrap filtering (BF) (Section 11.2.2).

11.2.1. Response surface based on the arbitrary polynomial chaos expansion

PCE techniques can be seen as an efficient and mathematically optimal approach to construct the response surface of a model with uncertain parameters. A response surface can be con-

structed in different ways. The most straightforward way is to construct it directly on a dense Cartesian grid of input parameters together with an adequate interpolation rule. Due to the curse of dimension (Babuska et al., 2007) for tensor grids, this approach has an extremely high computational effort if more than a single parameter is of interest. Alternatively, conceptually straightforward numerical Monte Carlo (MC) simulation techniques can be used to screen the parameter space. Their convergence independent of the number of uncertain parameters. However, they are also computationally demanding since the statistical accuracy of their predictions increases only with the square root of the number of realizations used.

The PCE introduced by Wiener (Wiener, 1938), generally, can be viewed as an efficient approximation to full-blown stochastic modeling (e.g., exhaustive MC). The basic idea is to approximate the response surface of a model with the help of an orthonormal polynomial basis in the parameter space (Ghanem and Spanos, 1991; Le Maitre and Knio, 2010). In simple words, the dependence of model output on all relevant input parameters is approximated by projection onto a high-dimensional polynomial. This projection can be interpreted as an advanced approach to statistical regression. Alternatively, PCE can be interpreted as a smart and polynomial-based interpolation and extrapolation rule of model output between and beyond different parameter sets.

The key attractive features of all PCE techniques are the high-order approximation of the model (Fajraoui et al., 2011; Foo and Karniadakis, 2010; Ghanem and Spanos, 1990, 1991; Zhang and Lu, 2004) combined with its computational speed when compared to alternatives such as MC (Oladyshkin et al., 2011b). This advantage holds mostly for low-parametric problems. The exact break-even point against Monte-Carlo, however, depends on model non-linearity, the dimension of the parameter space and on the chaos expansion order, so it varies from problem to problem.

Formally, let $\omega = \{\omega_1,...,\omega_N\}$ represent the vector of N input parameters for some model $\Omega = f(\omega)$. The model $\Omega(\omega)$ may be an explicit or implicit expression (e.g., a partial or ordinary differential equation or a coupled system). We wish to investigate the influence of all parameters ω on the model output Ω. That the model output may be time and space dependent, $\Omega = f(\omega, t, \mathbf{x})$, where $\mathbf{x} = (x_1, x_2, x_3)$. According to polynomial chaos theory (Wiener, 1938), the model output Ω can be approximated by polynomials $\Psi_j(\omega)$:

$$\Omega(\mathbf{x}, t; \omega) \approx \sum_{j=0}^{M} c_j(\mathbf{x}, t) \Psi_j(\omega) = \widetilde{\Omega}(\mathbf{x}, t; \omega). \qquad (11.1)$$

The number M of polynomials Ψ_j and corresponding coefficients c_j depends on the total number of analyzed input parameters (N) and on the order d of the polynomial representation

as discussed below. The coefficients $c_j(\mathbf{x}, t)$ in Eq. (11.1) quantify the dependence of the model output Ω on the input parameters ω for each desired point in space $\mathbf{x}$ and time t, resulting in a surrogate model $\widetilde{\Omega}$.

Assuming that the input parameters within ω are statistically independent, the multi-dimensional basis Ψ_k can be constructed as a simple product of the corresponding univariate polynomials $\P_j$ (e.g. (Ghanem and Spanos, 1991)):

$$\Psi_k(\omega) = \prod_{j=1}^{N} P_j^{(\alpha_j^k)}(\omega_j), \quad \sum_{j=1}^{N} \alpha_j^k \leq M, \quad k = 1\ldots N. \tag{11.2}$$

Here α_j^k is a multivariate index that contains the combinatoric information how to enumerate all possible products of individual univariate basis functions. In other words, the index α can be seen as $M \times N$ matrix, which contains the corresponding degree (e.g. 0, 1, 2, etc.) for parameter number j in expansion term k. The set of polynomials $\{P_j^{(0)}, \ldots, P_j^{(d)}\}$ forms an an orthogonal basis of degree d in the space of parameter ω_j. For example, the polynomials $P_j^{(k)}(\omega_j)$ are Hermite polynomials, if the parameters ω_j are Gaussian distributed (e.g., (Wiener, 1938)). Generally, the polynomial $P_j^{(k)}(\omega_j)$ of degree k in an individual parameter ω_j can be written as a simple linear combination of the different powers i of ω_j:

$$P_j^{(k)}(\omega_j) = \sum_{i=0}^{k} p_{i,j}^{(k)} \omega_j^i, \quad k = 0\ldots d, \quad j = 0\ldots N, \tag{11.3}$$

where $p_{i,j}^{(k)}$ are the coefficients for the power $i = 0\ldots k$ within the polynomial $P_j^{(k)}(\omega_j)$.

Up to the present, all implementations of PCE require the random variables in which one expands to be statistically independent. However, PCE can also be applied to correlated variables, if correlation can be removed (or minimized) by adequate linear or non-linear transformation. Only in a few specific cases (including linear correlation), adequate linear (or non-linear) transformations allow expanding statistical by dependent variables in independent random variables. The advantages of using an expansion basis that is orthonormal directly in the physical random variables without further need for transformation are manifold. They include direct accessibility of the moments of Ω from the expansion coefficients c_j in equation (11.1), better performance of integration rules, and the fact that additional transformation would include new non-linear terms in the expansion (Oladyshkin and Nowak, 2012a). In particular, the technique presented can be extended to many classes of heterogeneous systems, where spatially correlated heterogeneous parameter fields can be decomposed into their uncorrelated principal components using a truncated KL-expansion (e.g. (Li and Zhang, 2007)), if heterogeneity does not span over too many scales.

In the current work, we will apply a most recent generalization of the PCE technique known as the arbitrary polynomial chaos (aPC) (see (Oladyshkin et al., 2011a; Oladyshkin and Nowak, 2012a,b)). Highly similar ideas can be found in (Ghanem and Doostan, 2006; Soize and Ghanem, 2004; Witteveen and Bijl, 2006; Witteveen et al., 2007). Compared to earlier PCE techniques, the aPC adapts to arbitrary probability distribution shapes of input parameters and, in addition, can even work with unknown distribution shapes when only a few statistical moments can be inferred from limited data or from expert elicitation. The arbitrary distributions for the framework can be either discrete, continuous, or discretized continuous. They can be specified either analytically (as probability density/cumulative distribution functions), numerically as histogram or as raw data sets. Thus, exact probability density functions do not have to be known and do not even have to exist. The available information can directly and most purely be used in stochastic analysis, when using our data-driven formulation of aPC.

The aPC approach provides improved convergence with respect to increasing order of expansion (e.g. (Oladyshkin and Nowak, 2012a)) in comparison to classical PCE techniques, when applied to input distributions that fall outside the range of classical (Wiener, 1938) or generalized version (Xiu and Karniadakis, 2003) of the PCE. aPC deals with a highly parsimonic and yet purely data-driven description of uncertainty. The new freedom opens the path to accessing with the chaos expansion technique even those applications where data samples of limited size merely allow the inference of a few moments, and one would not be able to construct a probability density function without introducing subjective assumptions and hence dangerous sources of bias. The necessity to adapt to arbitrary distributions in practical tasks is discussed in more detail in (Oladyshkin et al., 2011a). A future incentive to work with the aPC is that, during sequential Bayesian updating for non-linear problems, parameter distributions generally change their shapes from updating step to updating step. Thus, they will almost surely fall out of the optimality range of many previous PCE techniques.

According to (Oladyshkin and Nowak, 2012a), an orthogonal polynomial basis up to order d can be constructively defined for any arbitrary probability measure, given that ω_j has finite statistical moments (e.g., mean, variance, skewness, etc) up to order $2d-1$. The coefficients $p_{i,j}^{(k)}$ within the basis polynomial in Eq. (11.3) can be defined from the following matrix

equation:

$$
\begin{bmatrix}
\mu_{0,j} & \mu_{1,j} & \cdots & \mu_{k,j} \\
\mu_{1,j} & \mu_{2,j} & \cdots & \mu_{k+1,j} \\
\cdots & \cdots & \cdots & \cdots \\
\mu_{k-1,j} & \mu_{k,j} & \cdots & \mu_{2k-1,j} \\
0 & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
p_{0,j}^{(k)} \\
p_{1,j}^{(k)} \\
\cdots \\
p_{k-1,j}^{(k)} \\
p_{k,j}^{(k)}
\end{bmatrix}
=
\begin{bmatrix}
0 \\
0 \\
\cdots \\
0 \\
1
\end{bmatrix}.
\tag{11.4}
$$

Here, $\mu_{i,j}$ are the ith non-central (raw) statistical moments of the random variable ω_j. It becomes evident from Eq. (11.4) that statistical moments are the only form of information required on the input distributions. Other possible contribution methods are discussed in (Ghanem and Doostan, 2006; Soize and Ghanem, 2004; Witteveen and Bijl, 2006; Witteveen et al., 2007). If the matrix in equation (11.4) has a poor condition for high-order expansions, adequate problem scaling will be helpful.

The above orthogonal polynomial basis can be used directly for analysis. However, a normalized basis has further useful properties and can be obtained as:

$$
\widehat{P}_j^{(k)} = \frac{P_j^{(k)}}{\left\| P_j^{(k)} \right\|}, \quad \left\| P_j^{(k)} \right\|^2 = \int_{\omega_j \in \Lambda} \left[P_j^{(k)}(\omega) \right]^2 d\Gamma(\omega_j),
\tag{11.5}
$$

where $\| P_j^k \|$ is the normalizing constant of the polynomial P_j^k for space of events Λ (where $\omega_j \in \Lambda$) with probability measure Γ. In practical applications of the aPC, the integration in equation (11.5) is performed analytically based in the moments of ω up to order $2d$, which avoids (like equation 11.4) the necessity to know exactly the probability measure (Oladyshkin and Nowak, 2012a).

To determine the unknown coefficients $c_j(\mathbf{x}, t)$ of the expansion in equation (11.1), many different methods exist. We choose to apply the non-intrusive probabilistic collocation method (PCM for short, see e.g. (Li and Zhang, 2007; Oladyshkin et al., 2011b)). Non-intrusive methods do not require modifications in the system of governing equations. Thus, they require no corresponding changes in simulation codes, which is very valuable for modeling complex flow or transport processes with commercially available software. The PCM is based on a minimal and optimally chosen set of model evaluations, each with a defined set of model parameters (called collocation points) that is related to the roots of the polynomial basis via the optimal integration theory (Villadsen and Michelsen, 1978).

From the practical point of view, the computational costs of our framework are dominated by the model calls required in construction of the surrogate model, i.e., by aPC combined with

PCM. In the PCM technique, the number M of model evaluations is equal to the number M of coefficients, which can be obtained from the following equation:

$$M = (N + d)!/(N!d!) - 1. \tag{11.6}$$

The order d of expansion is typically found as a compromise between the accuracy required by the application, the number of parameters N included in the analysis, the computational costs of individual model evaluations and the available computer power.

The challenge here is to find a compromise between computational effort and a reasonable approximation of the physical processes. Sometimes, a much smaller number M can be found by a sparse selection (Blatman and Sudret, 2010) of the relevant terms in equation (11.1). Also, an initial sensitivity analysis can be applied to keep the number of analyzed parameters and their respective degree of expansion at a low and efficient level (see, e.g., (Saltelli et al., 2008; Oladyshkin et al., 2011d)). Because the number M quickly increases with the degree d of expansion when the number of parameters N is large, the framework is most efficient for low-parametric systems. In order to translate these advantages to spatially heterogeneous systems, a combination with the KL-expansion (e.g. (Li and Zhang, 2007)) to truncate the representation of spatial heterogeneity can be very useful.

11.2.2. Bootstrap filter

The aPC method in Section 11.2.1 reduces the original simulation model $\Omega(\omega)$ to a response surface (surrogate model) $\widetilde{\Omega}(\omega)$ that approximates the original model. Because the reduced model is merely a polynomial and has exploitable properties due to the orthogonal basis, it is vastly faster than the original one and offers a large playground for stochastic analysis (Ghanem and Spanos, 1993; Wan and Karniadakis, 2006; Oladyshkin et al., 2011a), risk assessment (Oladyshkin et al., 2011b; Walter et al., 2012) and global sensitivity analysis (Sudret, 2008; Crestaux et al., 2009; Oladyshkin et al., 2011a). In this Section, we will apply Bayesian updating (see e.g. (Smith and Gefland, 1992)) to match the surrogate model $\widetilde{\Omega}(\omega)$ to available measurements in a data vector $\mathbf{y}$ of state variables or to other past or real-time observations of system behavior. In this context, Bayes' theorem is:

$$f(\omega|\mathbf{y}) = \frac{f(\mathbf{y}|\omega)f(\omega)}{f(\mathbf{y})}, \tag{11.7}$$

where $f(\omega)$ is the joint prior probability density function (PDF) for the vector of model parameters ω, $f(\mathbf{y})$ is the prior probability of $\mathbf{y}$ used as normalization constant, $f(\mathbf{y}|\omega)$ is

the conditional PDF of **y** for given ω, i.e. the likelihood of the parameters, and $f(\omega|\mathbf{y})$ is the conditional PDF of ω for given **y**, that we seek to approximate swiftly and accurately.

Under the common assumption that the measurement errors are independent and Gaussian, i.e. $\mathbf{y} = \Omega(\omega) + \varepsilon, \varepsilon \sim N(0, \mathbf{R})$, the likelihood function becomes:

$$f(\mathbf{y}|\omega) \propto \exp\left[-0.5\left(\mathbf{y} - \Omega(\omega)\right)^T \mathbf{R}^{-1}\left(\mathbf{y} - \Omega(\omega)\right)\right], \tag{11.8}$$

where **R** is the diagonal (co)variance matrix of measurement errors. We will adapt this assumption for the remainder of this study, although other arbitrary error models could be considered as well.

The most direct numerical implementation of Bayes theorem (Eq. 11.7) is known as Bootstrap filtering (e.g. (Smith and Gefland, 1992)). Formally, we draw a number N_p of realizations of parameter vectors ω_i from the prior PDF $f(\omega)$:

$$\omega \sim f(\omega), \quad i = \overline{1, N_p} \tag{11.9}$$

where N_p is a sufficiently large number. The correction form $f(\omega)$ to $f(\omega|\mathbf{y})$ in equation (11.7) is represented by assigning importance weights w_i to each realization ω_i:

$$w_i = \frac{f(\mathbf{y}|\omega_i))}{max(f(\mathbf{y}|\omega_i))}. \tag{11.10}$$

Then, realization ω_i is accepted as a legitimate ensemble member of the posterior distribution, if $w_i \geq u_i$ is fulfilled for a random number u_i drawn from the uniform distribution $u(0,1)$. $max(f(\mathbf{y}|\omega_i))$ is the largest individual values $f(\mathbf{y}|\omega_i)$ appearing across all realizations $i = \overline{1, N_p}$. The convergence for a given value N_p can be assessed, e.g, via weighted Jackknife of Bootstraping techniques (Efron and Tibshirani, 2010; Leube et al., 2012b).

11.3. Scenario definition for history matching: the problem of CO$_2$ leakage

We will consider a relatively simple benchmark model to illustrate and thoroughly test the concept presented in the current study. We use the benchmark leakage problem of injected CO$_2$ into overlying formations through a leaky well defined by Class et al. (Class et al., 2009) as described in Section 11.3.1. Within our approach, the physical complexity of the problem can be increased to an arbitrary extent because of the non-intrusive black-box conception

of the proposed framework. How the approach scales with larger numbers of uncertain parameters has already been discussed in Section 11.2.1. Statistical assumptions, generation of the synthetic data sets and the test cases are defined in Section 11.3.2

11.3.1. Physical and numerical set-up of the CO$_2$ leakage problem

Figure 5.1 illustrates a 2D section of the 3D domain under consideration. The top of the figure is not the ground surface, but the transition to yet another aquitard. CO$_2$ is injected into a deep aquifer, spreads within the aquifer and, upon reaching a leaky abandoned well, rises to a shallower aquifer. CO$_2$ is injected at a constant rate of 8.87 kg/s, which corresponds to $1600 m^3/d$ at reservoir conditions. The leaky well is at the center of the domain and the injection well is 100m away. Both aquifers are 30m thick and the separating aquitard has a thickness of 100m. The initial conditions in the fully saturated domain include a hydrostatic pressure distribution which is dependent on the brine density. The aquifers are initially filled with brine. The initial pressure at the bottom of the domain (at 3000m depth) is 3.086×10^7 Pa.

The benchmark problem assumes that fluid properties such as density and viscosity are constant, all processes are isothermal, CO$_2$ and brine are two separate and immiscible phases, mutual dissolution is neglected, the pressure conditions at the lateral boundaries are constant over time, the formation is isotropic, rigid and chemically inert, and capillary pressure is negligible. More details and modeling parameters can be found in (Class et al., 2009).

Mass balances of the two phases and the multiphase version of Darcy's Law give the following system of differential equations:

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

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

(11.11)

which is constrained by the equation

$$S_w + S_g = 1. \tag{11.12}$$

The subscripts w and g stand for the brine (water) phase, and the CO$_2$-rich (gas) phase, respectively. The primary variables in equation (11.11) are the gas-phase saturation S_g and

pressure p. S_w is the brine-phase saturation. The relative permeabilities k_{rw} and k_{rg} are secondary variables and linear functions of S_w and S_g, ($k_{rw} = S_w = 1 - S_g$; $k_{rg} = S_g$), **g** is the gravity vector, **K** is the absolute permeability tensor, ϕ is porosity and q_w, q_g are sources/sinks.

The CO$_2$ leakage rate is defined in the benchmark study as the total CO$_2$ mass flux integrated over a horizontal control plane midway between the top and bottom aquifer, divided by the injection rate, in percent. In the current study, the benchmark problem is simulated using DuMuX, a multi-scale multi-physics toolbox for the simulation of flow and transport processes in porous media (Flemisch et al., 2007). The goal of the simulation is to quantify the leakage rate which depends on the pressure build-up in the aquifer due to injection, on the proprieties of formation and the leakage well and on the plume evolution.

11.3.2. Set-up for the history matching task

In our study we will consider three uncertain parameters: reservoir absolute permeability, reservoir porosity and permeability of the well. For simplicity, we feature here only a scalar permeability, homogeneous within the entire reservoir. The prior assumptions on the parameters (see Section below) will be updated using our framework, such that model predictions match a synthetic time series of pressure monitored daily at the top of the well (Figure 5.1) during the 100 days of injection.

We assume prior distributions of the uncertain model parameters as illustrated in Figure 11.1. For reservoir permeability and porosity, these distributions are represented by raw data sets taken from the U.S. National Petroleum Council Public Database (which includes
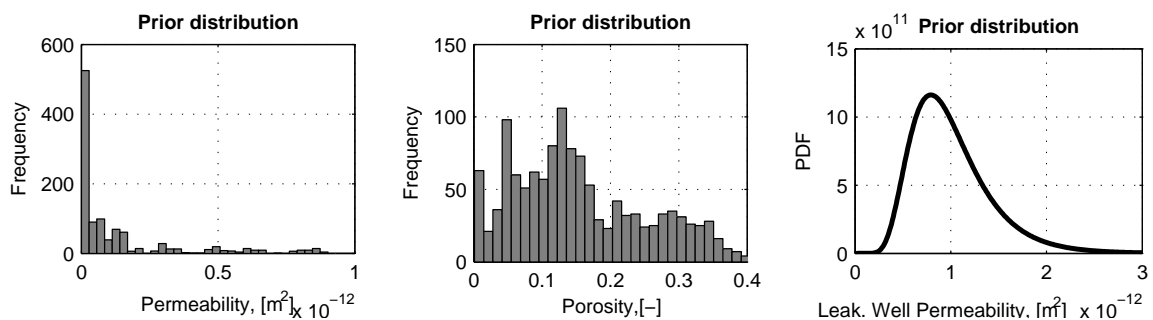
Figure 11.1.: Prior distribution of model parameters: absolute permeability, porosity and leakage well permeability.

1270 reservoirs), see also (Kopp et al., 2009). For the well, we assume as expert knowledge a log-normal PDF with the location parameter $\mu = -27.7$ and the scale parameter $\sigma = 0.4$. The presented aPC-based framework does not have any restrictions on the shapes of prior distributions, thus it can handle the statistical information provided here without further modification.

As a specific example, the positivity of permeability and the (0,1) -boundedness of porosity are directly enforced through their assigned probability distributions, without any additional necessary action. In our test case, we use a third-order aPC expansion, because it demonstrates the advantages of non-linear information and expert opinion processing over linear methods (e.g., the entire family of Kalman filters). Typically, a third-order expansion has the freedom to handle non-trivial and possibly non-monotonic behavior of model output. However, a second-order analysis has been shown to be a reasonable approximation for aPC applied to this benchmark problem in (Oladyshkin et al., 2011a). In general, the degree of expansion can be chosen according to the specific needs and available computer resources in any specific application, and it can also be increased adaptively during the matching procedure (see details in Section 11.5). We prefer here a third-order expansion, because most parameter-state relations in the featured model are monotonic, and second-order polynomials do not offer the possibility to approximate monotonic behavior over large parameter ranges.

To illustrate the proposed method, we will generate two reference simulations (Case1 and Case 2), which provide synthetic reference time series of pressure data at the monitoring well. In Case 1, the parameter values (Table 11.1) used in the reference simulation are reasonably probable within the assumed prior distributions (see Figure 11.1). In contrast to Case 1, Case 2 will use parameter reference values (see porosity in Table 11.1) in a pressure response which is extremely improbable according to the prior. The resulting reference pressure time series and how they compare to the prior distribution of pressure can be seen in Figure 11.2 (Case 1) and Figure 11.8 (Case 2). Test Case 1 is treated in Section 11.4 and Test Case 2 is treated in Section 11.5.

	Reservoir absolute permeability	Reservoir porosity	Leaky well permeability
Case 1	$1.43 \cdot 10^{-13} \mathrm{m}^2$	0.023 [-]	$1.09 \cdot 10^{-12} \mathrm{m}^2$
Case 2	$1.43 \cdot 10^{-13} \mathrm{m}^2$	0.259 [-]	$1.09 \cdot 10^{-12} \mathrm{m}^2$

Table 11.1.: Reference values: Case1 and Case 2

We will consider a random synthetic measurement error (i.e. standard deviation) of 1 bar. Thus, in our case studies, the original model can never reproduce exactly the data, even when the "true" parameter set used in data generation is applied, which is typical in real-world applications. Please note that, due to the overdetermined and stochastic problem character, the individual data points will not be matched precisely, but the solution will be a best compromise between prior information and a best fit of the physical model to the cloud of data points.

11.4. Straightforward aPC-based bootstrap filtering

We will now apply the straightforward aPC-based BF described in Section 11.2 to Case 1 of the benchmark problem of CO_2 leakage described in Section 11.3. That means, we will perform history matching to the synthetic pressure data marked as squares in the right plot of Figure 11.2. First, we construct a surrogate model by projecting the original model (Eq. 11.11) onto a polynomial response surface via the aPC at third order (see details in Section 11.2.1). Then, we apply the BF as described in Section 11.2.2. Figure 11.2 illustrates the unmatched pressure histogram after 100 days of injection (left plot) and the correspondence of predicted values (mean $\pm$ two standard deviations) to the synthetic observed values during the first 100 days (right plot).

We draw a number of $N_p = 100.000$ particles from the prior distributions. This number was assured to be sufficiently large to yield stable posterior statistics upon filtering. Because the data base we use for defining the prior distributions in not as large as 100.000 data sets, we use a PDF estimation technique based on Kernel smoothing (see e.g. (Wand and Jones, 1994)) with minimal Kernel width to obtain $N_p = 100.000$ particles. Then, the obtained surrogate model is evaluated for each particle, and each particle is reweighed according to its correspondence to observation values (see equations (11.7) and (11.8)).

At first, we are interested in the statistical distribution of the uncertain parameters and of the pressure at the monitoring well before and after matching, in order to assess the accuracy of the inversion. The resulting posterior distributions of model parameters are presented in Figure 11.3 and Figure 11.4 against the reference values. The quality of matching the pressure data is shown as PDF in Figure 11.5. Figure 11.5 also shows how the quality of matching can be improved when increasing the number of available observation values. Overall, we observe very satisfactory and consistent results: increasing the degree of expansion helps to better catch non-linearity of the original model and shows convergence of distribution peaks

Figure 11.2.: Prior pressure at monitoring well: pressure histogram after 100 days of injection (upper plot) and correspondence to observation values during 100 days (bottom plot), for Case 1

and mean values to reference values in Figure 11.3, as much as allowed by the data considered for history matching. The peaks of the posterior parameter PDFs in Figure 11.5 (for 3rd order) and for the posterior pressure values in Figure 11.4 (also 3rd order) are close to the reference values, with improved accuracy for an increasing number of considered data values. The posterior pressure PDF at the monitoring well converges to the noise-free reference value. In Figure 11.5, the time-dependent posterior mean of pressure at the monitoring well converges, already with just few data values, to a best-fit curve through the noisy data, almost coinciding completely with the nose-free reference curve. Finally, the noisy data points fall into the 95% confidence interval of the posterior pressure distribution (see Figure 11.5), indicating reasonable error bounds.
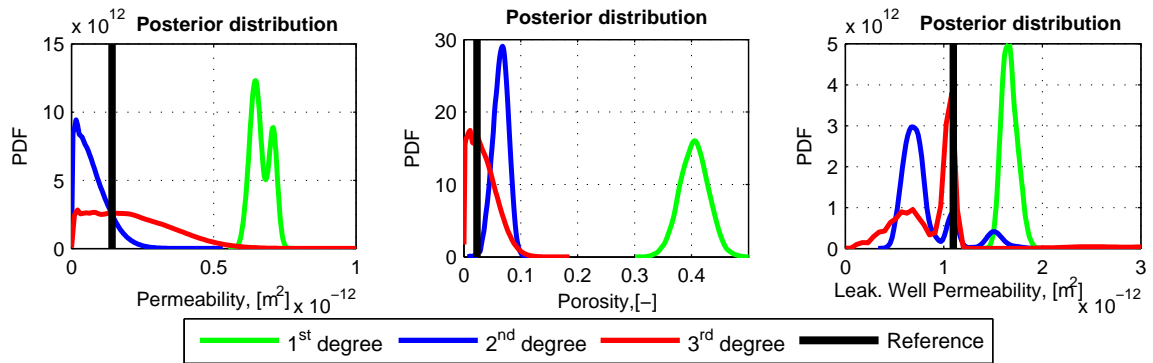
Figure 11.3.: Posterior distribution of model parameters obtained by 1st, 2nd and 3rd degree of expansion using 100 measurement values: absolute permeability (left), porosity (center) and leakage well permeability (right) in Case 1
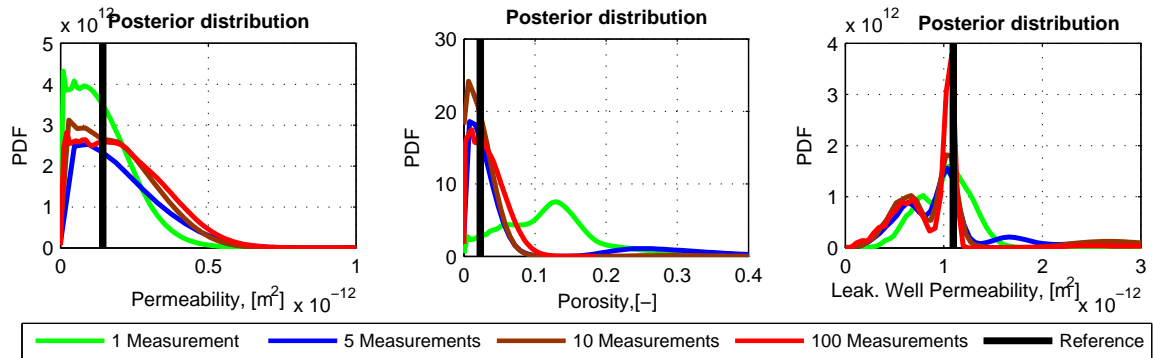
Figure 11.4.: Posterior distribution of modeling data using 1, 5, 10 and 100 measurements at 3rd order of expansion: absolute permeability, porosity and leakage well permeability in Case 1

In addition, the multivariate posterior probability density function can be constructed from the posterior ensemble for the model parameters, see the left plot in Figure 11.6. Here we used the multivariate kernel density estimation based on FIGTree algorithm (Morariu et al., 2008). The black point in the left plot of Figure 11.6 indicates the peak of the posterior multivariate PDF. As a final convergence check, the original model has been simulated within the corresponding parameter values. The root square error of the obtained pressure from the reference is presented in the right plot of Figure 11.6. It shows an acceptable approximation error that is small compared to the 95% confidence interval of the data noise (± 2 bars for 95% confidence interval). It can be improved even further by adjustment of the response surface in the domain of high posterior probability density, as it will be discussed in the next
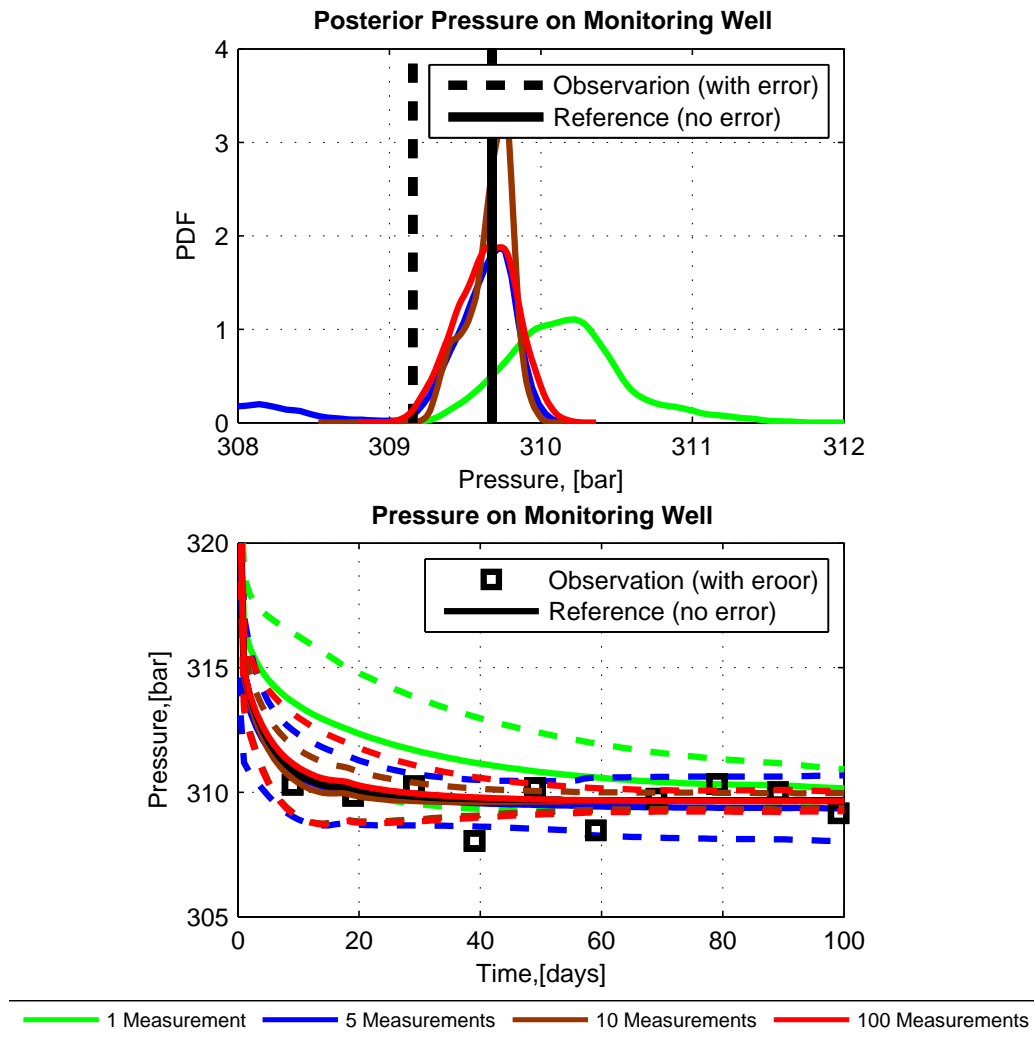
Figure 11.5.: Posterior pressure at monitoring well: pressure distribution after 100 days of injection using 1, 5, 10 and 100 measurements (upper plot), and matching of pressure (mean $\pm 2 \cdot$ std) to all observation values during all 100 days (bottom plot) for Case 1. All results obtained at 3rd order.

Section.

Once the virtual model is matched to observation values, the forecast of storage behavior can be investigated. Diverse physical quantities, such as saturation, pressure distribution within the overall domain and the CO_2 leakage rate can be predicted. We will illustrate the reservoir behavior after 1000 days by the cumulative probability function (CDF) of the CO_2 leakage rate, which is an integrative and very important characteristic of the overall benchmark problem for probabilistic risk assessment. The CDF represents the probability
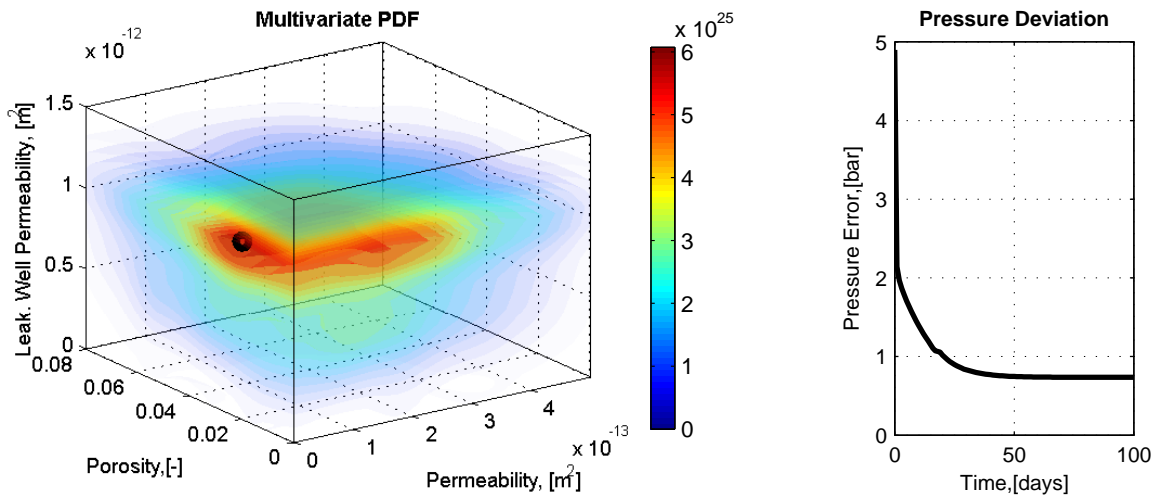
Figure 11.6.: Multivariate posterior probability density function based on 100 measurements (left plot) and deviation of posterior peak PDF pressure from the reference (right plot) for Case 1. The large values for probability density on the color scale in the left plot are caused by the small magnitudes of permeability

that the CO_2 leakage after 1000 days is less than or equal to a particular value. Figure 11.7 shows the corresponding prior and posterior CDFs. It also illustrates how the matching with more and more data improves the prediction. This illustrates how the proposed framework could be used within worth-of-data and information analysis (e.g., (Leube et al., 2012a; Nowak et al., 2012))
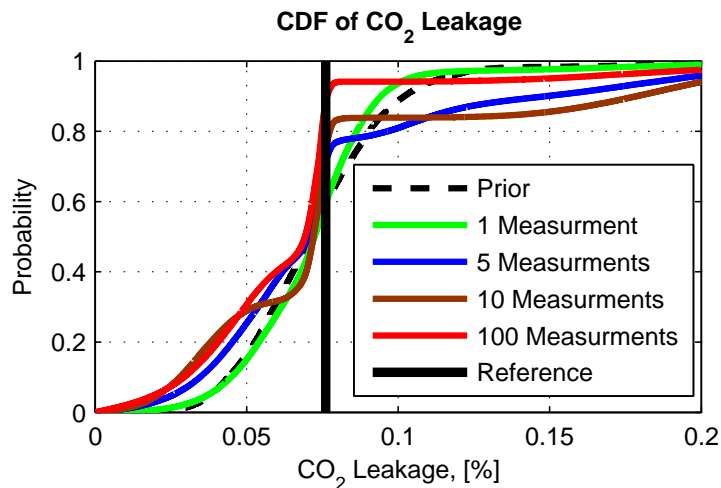
Figure 11.7.: Cumulative density function of CO_2 leakage: prior and posterior prediction after 1000 days injection using 1, 5, 10 and 100 measurements for Case 1

In order to illustrate the efficiency of the presented history matching framework, we will provide some details regarding the involved CPU time. Here we considered a simple Benchmark model which demands approximately 10 minutes of CPU time without parallelization on a contemporary PC. To obtain the response surface ($N = 3$ and $d = 3$), we performed 20 runs of the original Benchmark model, resulting in 200 min. overall simulation time. One evaluation of the obtained response surface for history matching demands approximately 0.2 milliseconds. In this example, we used $N_p = 100.000$ particles for Bootstrap filtering. Hence, for the overall procedure of history matching, we used about 3.34 hours ($20 \times 600sec. + 100.000 \times 0.0002sec.$, i.e. it takes only about 20 sec. only to perform the actual Bayesian updating step) of CPU time. Alternatively, direct application of Bootstrap filtering to the original Benchmark model would demand approximately 695 days ($100.000 \times 600sec.$) of CPU time. It is evident that such direct application of Bootstrap filtering (or similar methods) seems to be not feasible even when involving parallelization, especially for realistic and complex models. Thus, the presented framework provides an extremely efficient way for handling the task of history matching, with very accurate statistical methods. Without the PCE-based response surface, nobody would ever even consider applying an approach like Bootstrap filtering on this type of model.

11.5. Iterative aPC-based bootstrap filtering

In this Section, we will re-consider the problem of history matching for the case where the true properties of the system are very far from the prior (Case 2 defined in Section 11.3.2, see Figure 11.8). This is a very challenging problem for updating, because the assumed prior distribution does not adequately cover the domain of interest in hindsight. Thus, the aPC-based response surface used in BF is fitted to a distant and poorly chosen region within the parameter space. That means, it represents an expansion around a point or distribution far away from the region of interest, and hence cannot be expected to be accurate. Theoretically, the approximation could be improved significantly using a higher degree of expansion, leading to a very high number of model evaluations. Unfortunately, this would ask for a very high computational power, as can be seen from equation (11.6). As an alternative, we will re-iterate the response surface, keeping it at third order.

We will initially perform the same steps as in Section 11.4, i.e. we will project the original model (equation (11.11)) onto a response surface via aPC and then apply BF on the obtained surrogate model. Let us denote this initial step as 0th iteration. The corresponding results are
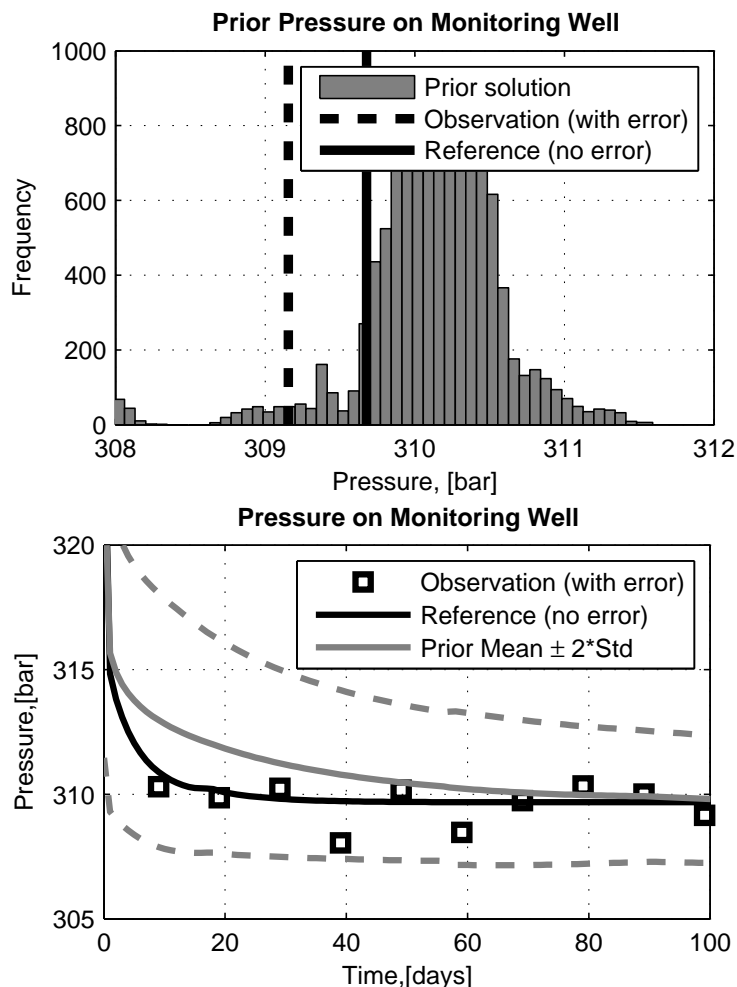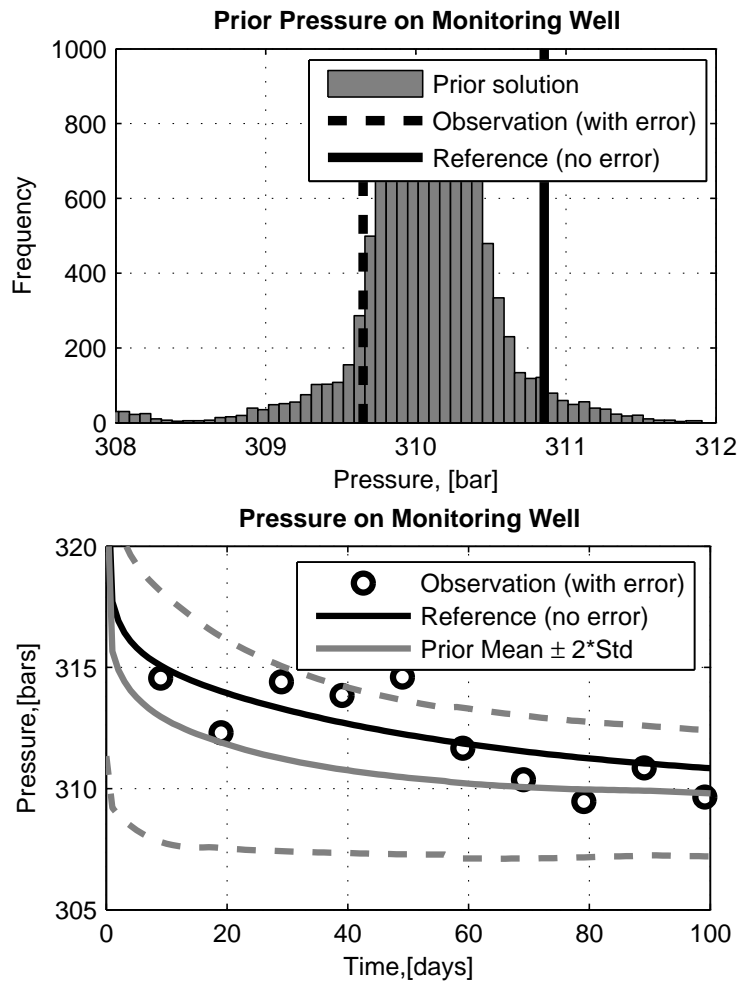
Figure 11.8.: Prior pressure at monitoring well: pressure histogram after 100 days of injection (upper plot) and correspondence to observation values during 100 days (bottom plot) for Case 2

presented by green lines in Figure 11.9 for the posterior distribution of model parameters, and for the updated pressure at the monitoring well in Figure 11.10.

We will improve this initial 0th prediction by applying an iterative approach. The idea we will pursue here is to improve the response surface in the parameter domain where the respective previous step indicated a high posterior probability density, because this is the alleged (best current guess for the) parameter region of interest. For that, in each iteration, we will include new integration points for the projection onto the orthonormal basis. These additional integration points for projection are located in the space of uncertain parameters, where the current iterate of the posterior density is largest. For all new integration points,
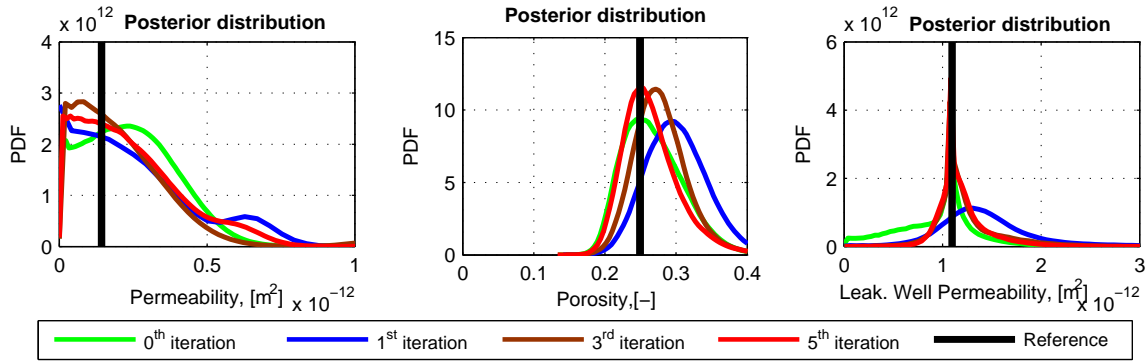
Figure 11.9.: Posterior distribution of modeling data: absolute permeability, porosity and leakage well permeability

we will run the original model (equation (11.11)) to obtain the corresponding model outputs Ω. Then we will perform a new projection of the model onto the orthonormal basis (see Section 11.2.1) using all cumulatively available collocation points within the least-squares collocation method (e.g.(Chen et al., 2009; Moritz, 1978)) instead of simple collocation. The extension of the original probabilistic collocation method by least-squares collocation during the iterations is very useful for complex applied tasks, because it keeps the black-box property. Also, the initial PCM in the 0th iteration can be understood as the special case of least-squares at the limit of a determined problem.

The residual of the least-squares collocation method in each integration point is not equal to zero in comparison to the classical probabilistic collocation technique, where the residuals are zero at all collocation points. However, it provides an integration rule which minimizes the overall residuals (Liou and Lin, 1991) in all collocation points. In that way, we can cumulatively incorporate all available (i.e original and additional) integration points. Thus, the updated response surface contains more accurate information about the system behavior in all alleged regions of interest. In practice, it is improved in the region where the posterior distributions during iteration assign a strong weighting. This bears some conceptual similarity to approaches like the shifted Hermite chaos (e.g., (Paffrath and Wever, 2007)) for evaluation of failure probabilities. The proposed iteration scheme can be very useful for history matching because it can help to avoid the dangerous consequences of specifying a misleading prior distribution. Our method does not change the meaning or importance of such prior information in Bayesian updating, but merely protects the response surface from being expanded in an inappropriate alleged region of interest.

In each iteration step, the number of new integration points can vary from 1 up to any desired

number of such points. In our concept, the modeler can specify the order of local refinement (0th, 1st, 2nd, etc.). According to that order, a new polynomial basis can be constructed using the posterior distribution obtained from the previous iteration. Then, a new full tensor set of optimal integration points can be computed from the respective roots of the new polynomial basis, again based on the optimal integration theory (Villadsen and Michelsen, 1978).

Obviously, using the full tensor grid would be very costly, especially when the number of parameters and the order of local refinement are high. To reduce the number of model evaluations, we suggest again using the principles behind the PCM method. This means to add just a limited number of points from the suggested tensor grid, selected to cover the high probability region of the current posterior. Thus, the modeler has some freedom to choose the number of evaluations of the original model according to the available computational resources. Evidently, a large number of additional integration points will lead to a more robust projection in the region of the possible posterior. However, it is not guaranteed that a large effort for local refinement will immediately lead to final results, because the refinement can be placed poorly in the parameter space at early iteration steps. Thus, we suggest to use a moderate number of new integration points in each iteration step.

Once the response surface is improved by adding new collocation points, the history matching using Bootstrap filtering is applied again. This procedure is repeated until the matched model matches the observation values with a desired precision, or until no significant changes can be observed between the iterations any more. Overall, the iterative accumulation of collocation points leads to an efficient estimation of the posterior distribution. This refinement idea could also be extended to different methods which use response surfaces for inverse modeling (e.g. EnKF or MCMC combined with PCE, etc.).

In Case study 2, we choose a 1st order refinement for 3 parameters, and add 4 highly probable integration points out of the 8 available ones in each iteration step. The results of our iterations are presented in Figures 11.9 and 11.10 by green, blue, brown and red lines for the 0th, 1st, 3rd and 5th iteration, respectively. According to our observations, the posterior improves quite fast (with only a few iterations). In our case study, we stopped at the 5th iteration, because all posterior distributions were stabilized, i.e. the posterior distributions in the ith iteration did not differ significantly from the posteriors in the $(i-1)^{th}$ iteration.

After matching the model via the iterative version of aPC-based Bootstrap filtering, we perform a forecast. As in Section 11.4, we predict the CDF of the CO_2 leakage rate after 1000 days. The posterior CDF (after the 5th iteration) is significantly different from the prior CDF, see Figure 11.11, and also differs strongly from the results after the 0th and 1st iterations.
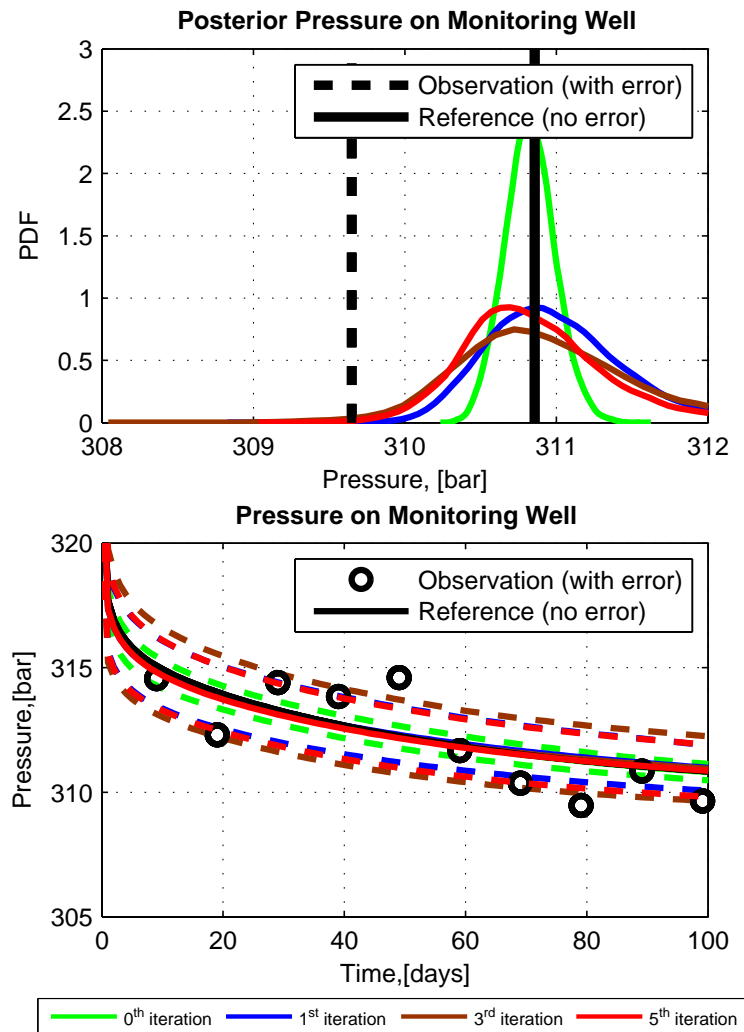
Figure 11.10.: Posterior pressure at monitoring well: pressure distribution after 100 days of injection (upper plot) and corresponding matching of pressure (mean $\pm\ 2\ \cdot$ std) to observation values during 100 days (bottom plot).

The question how to optimize the overall projection in the iteration approach and, in more general, under the changing probability measure from prior to posterior, is a more general and wider challenge than what can be addressed within the scope of the study. Such a procedure should find a balance between the number of iterations and the number of new integration points per iteration. If the prior assumption is strong offset against the posterior, then the computational effort for a very accurate expansion according to the prior distribution would be spent almost in vain. To overcome this drawback, the iterative framework presented here could first perform a cheap expansion using a low order. Then, using the flexibility of aPC, one could construct expansions of higher orders during the iterations, as the knowledge on
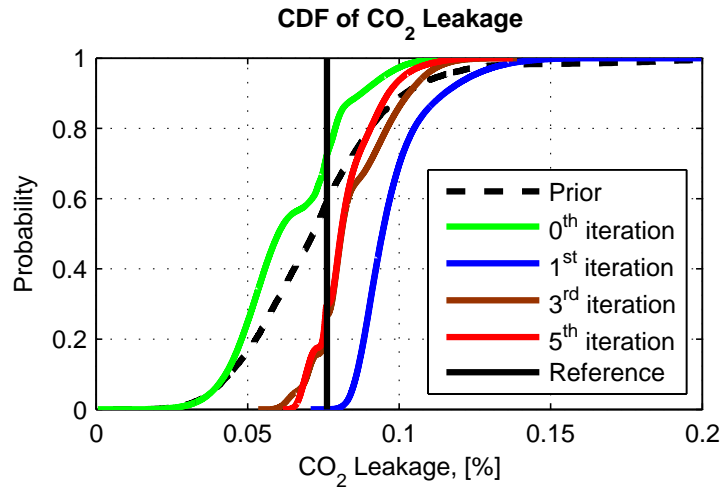
Figure 11.11.: Cumulative density function of CO_2 leakage: prior and posterior prediction during iterations after 1000 days.

the posterior improves and the increasing set of collocation points becomes more trustworthy to provide an accurate high-order response surface in the current alleged posterior region. Altogether, this poses significant challenges for future research.

11.6. Summary and conclusions

The present study deals with history matching of mathematical models to observation values. We propose an advanced framework for history matching based on response surface attained via the polynomial chaos expansion (PCE) and strict Bayesian principles. The reduced model represented by the response surface captures the model's dependence on textcolorblackits three physically most relevant parameters. For constructing the response surface, we used the aPC, a recent generalization of PCE theory, because it has very good convergence properties and tends to represent efficiently the underlying model according to the available statistical information. The aPC allows accommodating for a wide range of prior distributions of model parameters. The applied technique for projection onto the polynomial basis is totally non-intrusive, i.e. it is black-box compatible with arbitrary commercial simulation codes. Then, we perform Bayesian updating via Bootstrap filtering in order to match the obtained reduced polynomial model to past or real-time observations of system behavior. Basically, with Bootstrap filtering, we follow a direct implementation of Bayes theorem.

The combination of high-order expansion and Bootstrap filtering accounts for the non-

linearity of both the forward model and of the inversion. It takes into consideration higher-order statistical moments in comparison to (Ensemble) Kalman Filters. Hence, our method is more accurate than linearized inversion rules or related second-order moment approaches based on implicit lineralizations or on multi-Gaussian assumptions. The usually high computational costs of accurate filtering become very feasible in our suggested method by combining it with a response surface framework. Thanks to the computational efficiency of the aPC, Bayesian updating for history matching becomes an interactive task and could even be applied to real-time problems with complex large-scale real-world models in future works. textcolorblack The key contribution of our approach is that the response surface can be prepared in expansive off-line computation, but can served for Bayesian updating to new incoming data withing seconds.

The efficiency and power of Bayesian updating strongly depends on the accuracy of prior information. In our aPC-based methodology, the model parameter distributions can be determined from arbitrary available information (modeler's experience, expert opinion, general prior information or field data) and reflects the uncertainty or expected range of variation of input parameters. The polynomial basis of the aPC is able to adapt to arbitrary shapes of these parameter distributions. The presented methodology approximates the original model best where the prior probability density of the parameters is highest. In the case where the prior guess is highly inaccurate and strongly offset against the posterior, we suggest to use an iterative procedure, which helps to overcome this drawback even with small costs. We propose to account for the posterior on each iteration step and increase the precision of expansion around the current iterate of the posterior distribution.

A direct and straightforward application example of the proposed methodology is illustrated using a CO_2 benchmark problem. In this example, we found highly satisfying accuracy and computational efficiency. However, our methodology is not restricted to this example, as both the polynomial chaos expansion and Bootstrap filtering require no specific properties for the forward model or the inversion task. The only restriction is that the forward model has to be approximated well by the aPC.

Part IV.

Application to storage of energy relevant gases in geological formations

Modeling of environmental systems often demands such tasks as prediction of possible scenarios, design of planned engineering actions, estimation of risks, calibration of the model using available data etc. The common core of all tasks is multi-query evaluation of the underlying model with varying modeling parameters. Regardless of that, the overall modeling procedure should be performed within an acceptable time frame. Moreover, the modeling process in the field of environmental engineering is a chain of many tasks. Like any chain, it is only as strong as its weakest link. Estimation of uncertainty is part of the chain and underestimation of its influence can weaken that link immensely. Chapter 5 has illustrated that sources of modeling error can have different nature and their impact can be relevant for the overall modeling procedure. Thus, an adequate combination of a reliable physical model with advanced stochastic tools offers a powerful base for the forecasting of environmental behaviors. The current Part IV contains applications of model reduction technique to problems of energy relevant gas storages in geological formations.

Many environmental projects should be designed the way that even improbable dangers stay below acceptable risk levels. Chapter 12 discusses an integrative approach to robust design under uncertainty and probabilistic risk assessment for CO_2 storage in geological formations. It shows, that the probability of failure and negative impacts is an interplay between controllable engineering aspects and uncertain aspects of the system. Chapter 13 deals with assessment of pressure increasing in the storage which might result in caprock failure or brine displacement. Namely, Chapter 13 presents a study on the influence of the dimension of the high permeable channel and the permeability on pressure evolution in a sand channel system during carbon dioxide injection.

It is well known that quantification of uncertainty in underground systems requires assumption on the probability distributions of all model parameters, which posing a huge demand on data availability or requiring highly subjective assumptions on distribution shapes. Chapter 14 presents a minimally subjective approach for uncertainty quantification, based on a new and purely data-driven version of polynomial chaos expansion. However, analysis of uncertainty requires consideration of different kinds of uncertainties associated with the applied models. Chapter 15 distinguishes between scenario and statistical uncertainty. In particular, Chapter 15 estimates the risk of brine discharge into freshwater aquifers due to CO_2 injection into geological formations. A realistic large-scale problem of CO_2 storage is investigated in Chapter 16, which makes use of global sensitivity analysis and risk assessment based on chaos expansion framework. The influence of uncertain parameters that control the structural heterogeneities, such as barriers, aggradation angle, fault transmissibility and aquifer external support, have been analyzed in this study.

12. Robust design and probabilistic risk assessment for CO_2 storage

CO_2 storage in geological formations is currently being discussed intensively as a technology with a high potential for mitigating CO_2 emissions. However, any large-scale application requires a thorough analysis of the potential risks. Current numerical simulation models are too expensive for probabalistic risk analysis or stochastic approaches based on a brute-force approach of repeated simulation. Even single deterministic simulations may require parallel high-performance computing. The multiphase flow processes involved are too non-linear for quasi-linear error propagation and other simplified stochastic tools. As an alternative approach, we propose a massive stochastic model reduction based on the probabilistic collocation method. The model response is projected onto a higher-order orthogonal basis of polynomials to approximate dependence on uncertain parameters (porosity, permeability etc.) and design parameters (injection rate, depth etc.). This allows for a non-linear propagation of model uncertainty affecting the predicted risk, ensures fast computation and provides a powerful tool for combining design variables and uncertain variables into one approach based on an integrative response surface. Thus, the design task of finding optimal injection regimes explicitly includes uncertainty, which leads to robust designs with a minimum failure probability. We validate our proposed stochastic approach by Monte Carlo simulation using a common 3D benchmark problem. A reasonable compromise between computational efforts and precision was reached already with second-order polynomials. In our case study, the proposed approach yields a significant computational speed-up by a factor of 100 compared with the Monte Carlo evaluation. We demonstrate that, due to the non-linearity of the flow and transport processes during CO_2 injection, including uncertainty in the analysis

leads to a systematic and significant shift of the predicted leakage rates towards higher values compared with deterministic simulations, affecting both risk estimates and the design of injection scenarios.

12.1. Introduction

12.1.1. Modeling carbon dioxide storage in geological formations

It is highly likely that carbon dioxide (CO_2) emissions are influencing the global climate (IPCC, 2005). Carbon capture and storage (CCS) has been proposed as an interim technology to reduce CO_2 emissions. This technology comprises capturing CO_2 at industrial facilities, compressing it into a fluid or supercritical state and disposing of it in deep underground formations. Before the technology can be implemented on a scale that can become relevant for the climate, the risks of injecting CO_2 into the subsurface need to be analyzed thoroughly. Examples of these risks are CO_2 leaking back to the atmosphere or into drinking-water aquifers, overpressures in the formations etc.

In recent years, a great deal of research was directed towards understanding the physical processes in CO_2 storage. In order to describe the space-time evolution of the CO_2 plume and the influence of potentially leaky wells, (semi-)analytical solutions were derived by Nordbotten et al. (Nordbotten et al., 2005a, 2009). A comparison study of various simplifying semi-analytical models with complex numerical models was discussed by Ebigbo et al. (Ebigbo et al., 2007). The analysis in (Birkholzer et al., 2009) focused on the effects of large-scale CO_2 leakage through low-permeability layers. Various optimization strategies for characterizing surface leakage using near-surface measurement approaches were developed in (Cortis et al., 2008). It is evident that multiphysical and multiphase non-linear models for describing CO_2 flow became extremely complex due to the instability of the flow, phase transitions, geological-formation properties, chemical effects etc. Thus, a current question to be addressed by research is how many different effects and processes need to be included in the analysis. Kopp et al. (Kopp et al., 2009) performed a sensitivity analysis of different parameters and forces, including the effects of reservoir parameters such as depth, temperature, absolute and relative permeability, as well as capillary effects. Recently, Class et al. (Class et al., 2009) published a benchmark study, comparing a number of mathematical and numerical models with different complecity applied to specific 3D problems on the reservoir scale

in the context of simulating processes during CO_2 injection in geological formations on a time scale up to 50 years.

12.1.2. Options for uncertainty analysis

Modeling underground CO_2 storage involves many conceptual and quantitative uncertainties (Hansson and Bryngelsson, 2009a). The lack of information about distributed properties leads to model uncertainties up to a level where the quantification of uncertainties becomes the dominant question in application tasks and may override the influence of secondary physical processes. In the development of CO_2 injection as a large-scale interim solution, our ability to quantify its uncertainties and risks will play a key role. Unfortunately, only sensitivity analyses (Birkholzer et al., 2009), (Kopp et al., 2009) and no probabilistic risk assessment for CCS has been applied up to the present. Fault-tree analyses have been used to identify risks through different factors (Wildenborg et al., 2005), but have not yielded quantitative risk information.

The multiphase flow and transport processes involved are strongly non-linear, including phase changes about the supercritical state and effects such as gravity-induced fingering, convective mixing etc. This eliminates quasi-linear and other simplified stochastic tools from the list of reliable options. Current numerical simulation models are inadequate for stochastic simulation techniques based on brute-force Monte Carlo simulation and related approaches (e.g. (Maltz and Hitzl, 1979)), because even single deterministic simulations may require parallel high-performance computing. Thus, the necessity for reasonably fast and non-linear stochastic approaches for modeling CO_2 sequestration poses a research task that needs to be investigated as soon as possible. In the current study, we suggest and apply a massive stochastic model reduction technique based on non-intrusive polynomial expansion, as explained below.

A large number of studies for diverse applications are based on the polynomial chaos expansion introduced by Wiener (Wiener, 1938). The chaos expansion offers an efficient accurate high-order way of including non-linear effects in stochastic analysis. The chances and limitations of polynomial chaos expansion techniques were discussed in (Augustin et al., 2008). This technique can mainly be sub-divided into intrusive and non-intrusive approaches. The intrusive approach requires manipulation of the governing equations and can often provide semi-analytical tools for stochastic analysis. The best-known method from this group is the stochastic Galerkin technique, which has been applied in studies like (Xiu and Karniadakis,

2003) for modeling uncertainties in flow simulations. However, because the necessary symbolic manipulations of the model equations may become very complex, non-intrusive approaches like the probabilistic collocation method (Isukapalli et al., 1998), (Li and Zhang, 2007) have been receiving increasing attention. The non-probabilistic collocation method was developed by Villadsen and Michelsen (Villadsen and Michelsen, 1978) and has become more popular in recent years for different stochastic applications, for example (Huang et al., 2007), (Foo and Karniadakis, 2010), (Shi et al., 2009), (Lin and Tartakovsky, 2009). Reviews of the mathematical theory of the aforementioned methods, focusing on their strengths and limitations, are presented in (Jakeman and Roberts, 2008). In the current chapter, the non-intrusive probabilistic collocation technique will be considered.

12.1.3. Purpose of the work

This chapter has two principal purposes. The first is to develop a reasonably accurate probabilistic risk-assessment method at acceptable computational costs. The challenge here is to find a compromise between computational effort and a reasonable approximation of the physical processes. To this end, we apply for the first time the probabilistic collocation method to the problem of CO$_2$ sequestration, obtaining an accurate high-order tool for probabilistic risk assessment. The second purpose is to develop a tool for the optimal design of CO$_2$ injection regimes. To this end, we present a novel framework, which projects all design parameters and uncertain parameters onto a single integrative response surface. This integrates the design task into the stochastic model, as presented in section 12.2, allowing us to find robust designs with controlled failure risks. We also draw attention to the possible significant impacts of incorporating uncertainty into risk predictions and design tasks, in comparison to deterministic modeling approaches (section 12.3.4).

12.1.4. Approach

The current work provides a massive stochastic model reduction via the polynomial chaos expansion, as presented in section 12.2.3. We employ the collocation method (section 12.2.4) to project the model response surface onto the polynomial chaos. This technique allows for non-linear propagation of uncertainties in the risk analysis at low computational costs. Our variables of interest (see section 12.3.2) include uncertain parameters such as porosity, permeability etc., and a list of design parameters (injection rate, depth etc.).

We validate our proposed approach by Monte Carlo simulation in section 12.3.3, using a small version of the 3D benchmark study defined by Class et al. (Class et al., 2009). Further, we apply the integrative probabilistic collocation method to the optimal design of a CO_2 injection regime based on the same scenario.

12.2. The integrative probabilistic collocation method

12.2.1. Stochastic response surface methods

Stochastic response surface methods (Isukapalli et al., 1998) deal with the characterization of uncertainties in systems describing the dependence of model output on the uncertain input parameters. In the case of CO_2 storage, the uncertainty input parameters are porosity, permeability etc. This uncertainty has a non-trivial influence on the model output, like the spatial distribution of pressure, gas saturation, amount of CO_2 leakage etc. In the present chapter, we construct the stochastic response surface in a closed polynomial form. Obviously, a response surface can be constructed in different ways, e.g. can be constructed directly on dense Cartesian grid of input parameters with a much higher computational efforts. In that way, the response surface methodology has been applied, e.g., to design an experiment for CO_2 miscibility behavior in (Ghomian et al., 2008). In the current chapter we also would like to explore methodology which demands minimum number of model evaluation to construct response surface (see section 12.2.3).

12.2.2. Novel integrative concept for analysis

Integrative modeling tasks We classify model parameters in two classes: design or control parameters that can be chosen by the operator of a system, and uncertain parameters that describe our (incomplete) knowledge of the system properties. In the CO_2 injection problem, the latter include, for example, permeability, porosity etc. On the other hand, the system's performance and failure probability will also depend on design parameters such as the injection rate or depth. Evidently, the decision-making for design parameters will depend on the interplay between the response to injection strategy, system uncertainty and, finally, the probability of failure.

Integrative response surface In the present chapter, we introduce an integrative concept, in which all parameters group into one integrative, transparent and efficient structure. Actually, we project all design and uncertain parameters onto a single integrative model response surface. It is a multidimensional surface and contains the integral information about the system behavior under all possible conditions at all points in space and time. Thus, we expand the notion of stochastic response surfaces introduced in (Isukapalli et al., 1998) to integrative response surfaces forming an effective basis for robust design under uncertainty.

12.2.3. Polynomial chaos expansion

Multivariable input Let us introduce a formal model $\Omega = f(\omega)$. We investigate the influence of uncertain and design parameters ω on the model output Ω (integrative response surface approach). This means that our considered model has a multivariable input ω:

$$\omega_1, ..., \omega_{N_U}, \omega_{N_U+1}, ... \omega_{N_U+N_D} \tag{12.1}$$

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

Chaos expansion According to (Wiener, 1938), the model output Ω can be represented by the polynomial chaos expansion (PCE):

$$\Omega(\omega) = \alpha_0 + \sum_{i=1}^{N} \alpha_i H_1(\omega_i)$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{i} \alpha_{ij} H_2(\omega_i, \omega_j)$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{i} \sum_{k=1}^{j} \alpha_{ijk} H_3(\omega_i, \omega_j, \omega_k)$$

$$+ ... \tag{12.2}$$

where $\{H_1, H_2, H_3, ...\}$ form an orthogonal polynomial basis for ω and the coefficients α quantify the dependence of the model output Ω on the input parameters ω.

Hermite polynomials basis In this chapter, we project the stochastic model response surface onto the orthogonal basis of Hermite polynomials defined by:

$$H_d(\omega_{i_1}, ..., \omega_{i_d}) = (-1)^d e^{\frac{1}{2}\omega^T \omega} \frac{\partial^d}{\partial \omega_{i_1} ... \partial \omega_{i_d}} \left[e^{\frac{1}{2}\omega^T \omega} \right] \tag{12.3}$$

where d is the polynomial degree. Hermite polynomials are an optimal basis for normally distributed variables. Obviously, the distribution of physical parameters cannot always be assumed to be normal. Thus, we apply a transformation Θ from physical space ω^{Ph} to normal space ω :

$$\omega = \Theta(\omega^{Ph}) \tag{12.4}$$

where $\omega^{Ph} = \{\omega_1^{Ph}, \omega_2^{Ph}, ..., \omega_N^{Ph}\}^T$ consists of non-Gaussian distributed physical parameters. As an alternative approach, the polynomial basis could be chosen from the Askey scheme for selected non-Gaussian distributions (Askey and Wilson, 1985) or from other generalized options (Xiu and Karniadakis, 2003).

Weighting function   Even little information about reality can help to construct probability density functions of uncertain parameters, for example a log-normal distribution of permeability or empirical correlations between porosity and permeability. In other cases, the methods of maximum entropy (Jaynes, 1982) and minimum relative entropy (Woodbury and Ulrych, 1993) can be used to construct a probability distribution from sparse information. Of course design parameters do not have probabilistic distributions, but suitable weighting functions for such parameters can be described by user-defined feasibility functions that select the feasible range or preferences of the designing engineer concerning the values of design parameters. Feasibility functions provide a freedom for scenario analysis and can be used as an entry point for monetary punishment terms.

Space-time dependent expansion   We consider a system which is space-distributed and time-dependent, with model output $\Omega(X, \omega)$, where the vector X consists of three space coordinates and time $X = \{x, y, z, t\}$. Hence, the coefficients α in expansion (12.2) are determined for each point in space and time. In the case of CO_2 sequestration (see illustrations in section 12.3), the output Ω corresponds to CO_2 saturations, pressures in the reservoir, or, foe example, total CO_2 leakage to the surface. Here, saturation and pressure are three-dimensional and time-dependent, whereas total CO_2 leakage is just time-dependent.

In practice, only a finite number of terms in the expansion (12.2) is considered, leading to:

$$\Omega(X, \omega) = \sum_{j=1}^{P} c_j(X) \Psi_j(\omega) \tag{12.5}$$

where the coefficients $c_j(X)$ are deterministic space- and time-dependent functions, and $\Psi_j(\omega)$ is a simplified notation of the multi-dimensional cross-terms in equation (12.2). The

number of terms P in expansion (12.5) depends on the total number of input parameters N and the order d of the expansion, according to the following combinatory formula:

$$P = \frac{(N+d)!}{N!d!} \tag{12.6}$$

12.2.4. Collocation method

The remaining task is to evaluate the yet unknown coefficients $c_i(X)$ in expansion (12.5). The goal of diverse methods is to determine $c_i(X)$ from the point of view of weighted-residual minimization or projection. In this chapter, we use the collocation method.

Non-intrusive collocation method Various intrusive or non-intrusive techniques can be used to find the polynomial coefficients of this expansion. We apply the non-intrusive probabilistic collocation method (PCM) to find the coefficients, because it does not impose rearrangement, modification or simplification on the initial flow model. The basic idea of collocation was described in the book (Villadsen and Michelsen, 1978) by Villadsen and Michelsen in 1978 as a solution of deterministic differential equations by polynomial approximation as an alternative to the popular and effective perturbation method. In 1996, the collocation method was applied to the uncertainty analysis of a simple ocean model by Webster (Webster et al., 1996). During the same period, the collocation method was applied for probabilistic analysis to environmental and biological systems (Isukapalli et al., 1998), although it was not a popular tool for stochastic modeling at that time. Only in the last few years has the use of collocation methods increased in probabilistic analysis, especially where model complexity is very high with pronounced non-linearity. The uncertainty of steady-state single-phase flow in random porous media was analyzed in (Li and Zhang, 2007) by a combination of the Karhunen-Loeve expansion and the probabilistic collocation method (PCM). The method was later extended for the non-stationary formulation of non-linear flow in heterogeneous unconfined aquifers (Shi et al., 2009). Recently, (Lin and Tartakovsky, 2009) demonstrated that PCM on sparse grids can efficiently simulate solute transport in randomly heterogeneous porous media with large variances. In (Huang et al., 2007), PCM was applied to general random field problems. The multi-element PCM was tested by Foo et al. (Foo and Karniadakis, 2010) on a stochastic diffusion problem with various random input distributions, and the authors showed that the PCM error for estimating mean and variance is two orders of magnitude lower than the error obtained with Monte Carlo using the same number of model evaluations.
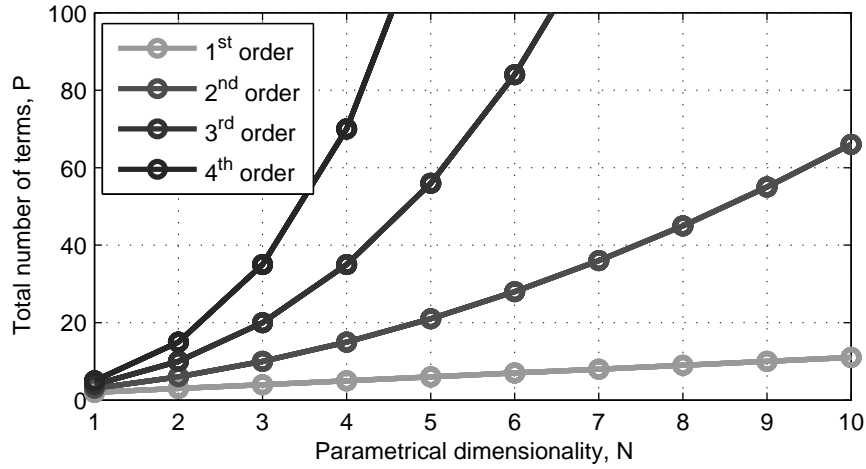
Figure 12.1.: Computational costs. Number of terms and model evaluations in polynomial chaos expansion of different orders

Collocation approach The main idea of the collocation method (Villadsen and Michelsen, 1978) consists of matching the unknown coefficients c_i to model outputs Ω_i evaluated with different parameters sets ω_i. The distinctive feature of non-intrusive approaches such as PCM is that any simulation model can be considered a "black-box", i.e. commercial software can be used without any modifications required.

Collocation formulation A given N-dimensional set of parameters ω_i is called a collocation point in the multi-dimensional parameter space. The number of collocation points ω_i is equal to the number P of unknown coefficients c_i in the formulated chaos expansion (12.5), requiring P model evaluations in total (Figure 12.1). This leads to the following system of linear equations:

$$M_\Psi(\omega)V_c(X) = V_\Omega(X) \tag{12.7}$$

where V_c is the $P \times 1$ vector of unknown coefficients c_i, the $P \times 1$ vector V_Ω contains the model output $\Omega_i(\omega_i)$ for each collocation point ω_i, and the $P \times P$ matrix M_Ψ contains the P Hermite polynomials $H_j(\omega_i)$ evaluated at the P collocation points. The vectors $V_c(X)$ and $V_\Omega(X)$ are space- and time-dependent, whereas the matrix $M_\Psi(\omega)$ does not depend on space and time and can be generated once for the given expansion degree and parameter number.

Selection of collocation points The solution $V_c(X)$ of system (12.7) depends on the selection of collocation points. According to Villadsen and Michelsen (Villadsen and

Michelsen, 1978), the optimal choice of collocation points corresponds to the roots of the polynomial of one degree higher than the order used in the chaos expansion. The resulting polynomial surface is exact at the collocation points, and is a polynomial inter - or extrapolation within and outside the domain with collocation points, respectively. The collocation method provides the same results as Galerkin projections for the case of one input parameter, and hence is regarded as an optimal method (Villadsen and Michelsen, 1978). For the multi-dimensional case, the number of available points is larger than the necessary number of collocation points.

There are two principal strategies for selecting collocation points: totally random and selection based on the distribution of the input parameters. Theoretically, any choice of collocation points from the available roots should give adequate estimates of the polynomial chaos expansion coefficients (Isukapalli et al., 1998). However, different combinations of collocation points will provide different estimates of the probability density function of the model output. In this work, we use the second option. We focus on the best estimation of the mean stochastic characteristics, and therefore select the collocation points from the most probable regions of the input parameters' distribution (Li and Zhang, 2007).

Post-processing computations Once the coefficients have been obtained, the initial model can be represented by the polynomials in equation (12.5). This means that the computational cost for evaluating diverse statistical quantities is now very low. The mean value of model output can be obtained directly via:

$$\mu_\Omega(X) = c_1(X) \tag{12.8}$$

This means that each model output $\Omega(X)$ has its own coefficients in space-time and coefficient $c_1(X)$ represents the mean value of that output $\Omega(X)$. The variance can be computed from:

$$\sigma_\Omega^2(X) = \sum_{i=2}^{P} c_i^2(X) \langle \Psi_i(\omega) \rangle^2 \tag{12.9}$$

In order to evaluate more complex statistical quantities, Monte Carlo simulation can be performed directly on the polynomial obtained. In this way, the probability density function and cumulative distribution function of the model output can be computed easily. All output statistics depend on the design parameters, and can be evaluated from the integrative response surface at low computational costs for any desired values of the design parameters.

12.3. Case study: Robust design and risk assessment for CO_2 storage

12.3.1. Benchmark description

We consider the benchmark leakage problem of injected CO_2 into overlying formations through a leaky well defined by Class et al. (Class et al., 2009). Figure 5.1 illustrates a 2D section of the 3D domain under consideration. CO_2 is injected into a deep aquifer, spreads within the aquifer and, upon reaching a leaky well, rises up to a shallower aquifer. A quantification of the leakage rate which depends on the pressure build-up in the aquifer due to injection and on the plume evolution is the goal of the simulation. At the lateral boundaries, constant boundary conditions are imposed. The leaky well is at the center of the domain and the injection well is 100m away. Both aquifers are 30m thick and the separating aquitard has a thickness of 100m. The leaky well is modeled as a porous medium with a higher permeability than the formation.

The benchmark study used the following assumptions (Class et al., 2009): fluid properties

Parameter	Value
CO_2 density, ρ_g	479kg/m^3
Brine density, ρ_w	1045kg/m^3
CO_2 viscosity, μ_g	$3.950 \cdot 10^{-5}$Pa s
Brine viscosity, μ_w	$2.535 \cdot 10^{-4}$Pa s
Aquifer permeability, K_A	$2 \cdot 10^{-14}$m^2
Aquifer thickness	30m
Aquitard thickness	100m
Leaky well permeability, K_L	$1 \cdot 10^{-12}$m^2
Porosity, ϕ	0.15
Leaky & injection well radius	0.15m
Injection rate	8.87kg/s (1600m^3/d)
Distance between wells	100m
Dimensions of model domain	$1000m \times 1000m \times 160m$
Simulation time, t	1000days

Table 12.1.: Simulation parameters

such as density and viscosity are constant; all processes are isothermal; CO$_2$ and brine are two separate and immiscible phases; mutual dissolution is neglected; the pressure conditions at the lateral boundaries are constant over time; the formation is isotropic; capillary pressure is negligible. The initial conditions in the domain include a hydrostatic pressure distribution which is dependent on the brine density. The aquifers are initially filled with brine. The initial pressure at the bottom of the domain (at 3000m depth) is 3.086×10^7 Pa. The lateral boundary conditions are constant Dirichlet conditions and equal to the initial conditions. All other boundaries are no-flow boundaries. CO$_2$ is injected at a constant rate of 8.87 kg/s, which corresponds to $1600 m^3/d$. Simulation time is 1000 days. The relevant parameters used for the simulation are given in Table 12.1.

Mass balances of the two phases and the multiphase version of Darcy's law give the following system of differential equations:

$$-\phi\frac{\partial S_g}{\partial t} - \nabla \cdot \left\{ \frac{k_{rw}}{\mu_w} \mathbf{K} \cdot (\nabla p - \rho_w \mathbf{g}) \right\} - q_w = 0$$

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

(12.10)

which is constrained by the equation

$$S_w + S_g = 1. \tag{12.11}$$

The subscripts w and g stand for the brine (water) phase and the CO$_2$-rich (gas) phase respectively. The primary variables in equation (12.10) are the gas-phase saturation S_g and pressure p. S_w is the brine-phase saturation. The relative permeabilities k_{rw} and k_{rg} are secondary variables and linear functions of S_w and S_g, ($k_{rw} = S_w = 1 - S_g$; $k_{rg} = S_g$), $\mathbf{g}$ is the gravity vector, $\mathbf{K}$ is the absolute permeability tensor and q_w, q_g are sources/sinks.

The benchmark problem was simulated using the DuMuX simulator. DuMuX (Flemisch et al., 2007) is a multi-scale multi-physics toolbox for the simulation of flow and transport processes in porous media. Its main intention is to provide a framework for the easy and efficient implementation of models for porous media flow problems, ranging from problem formulation, selection of spatial and temporal discretisation schemes, as well as non-linear solvers, to general concepts for model coupling.

12.3.2. Description of case studies

Case studies 1 and 2 We consider two scenarios within the described benchmark: case 1, validation on a small test problem (section 12.3.3), and case 2, application to a large problem combined with the design task of finding an optimal injection regime (section 12.3.4).

The first case study focuses on the validation of the proposed collocation approach by Monte Carlo simulation. With this in mind, we consider a small version of the benchmark problem where the simulation time is equal to 30 days, and the numerical grid is coarse (1183 nodes). This simplification is imposed by the expensive Monte Carlo approach. The second case study demonstrates our integrative approach for robust design under uncertainty on the original-size benchmark problem with a fine grid (65985 nodes) and a simulation time of up to 1000 days.

Uncertain parameters	*Probability distribution function*
Reservoir porosity, ϕ	Log-normal distribution: $$f(\phi) = \frac{1}{\phi\sigma\sqrt{2\pi}}\exp\left[-\frac{(ln(\phi)-\mu)^2}{2\sigma^2}\right], \phi > 0,$$ with $\mu = -1.8971$, $\sigma = 0.2$
Reservoir permeability, K_A	Analytical permeability-porosity correlation with beta distribution for error ε: $$f(\phi,\varepsilon) = \xi_1\phi^{\xi_2}\left[1+\xi_3*\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)}\left(\frac{\varepsilon}{\xi_3}\right)^{\alpha-1}\left(1-\left(\frac{\varepsilon}{\xi_3}\right)\right)^{\beta-1}\right],$$ with $\xi_1 = 4.0929\times10^{-11}$, $\xi_2 = 3.6555$, $\xi_3 = -2$, $\alpha = 3$, $\beta = 3$ and Γ is the Gamma function
Leaky well permeability, K_L	Log-normal distribution: $$f(K_L) = \frac{1}{K_L\sigma\sqrt{2K_L}}\exp\left[-\frac{(ln(K_L)-\mu)^2}{2\sigma^2}\right], K_L > 0,$$ with $\mu = -27.6310$, $\sigma = 0.3679$

Table 12.2.: Distribution of uncertain parameters

Design and control parameters	*Feasibility function*
Injection rate, Q_{CO_2}	$$f(Q_{CO_2}) = \frac{1}{Q_{CO_2}\sigma\sqrt{2\pi}}\exp\left[-\frac{(ln(Q_{CO_2})-\mu)^2}{2\sigma^2}\right], Q_{CO_2} > 0,$$ with $\mu = 2.1827$, $\sigma = 0.2$
Size of the screening interval, W_z	$$f(W_z) = \xi*\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)}\left(\frac{W_z}{\xi}\right)^{\alpha-1}\left(1-\left(\frac{W_z}{\xi}\right)\right)^{\beta-1},$$ with $\xi = 30$, $\alpha = 2$, $\beta = 2$ and Γ is the Gamma function

Table 12.3.: Distribution of design and control parameters

Uncertain and design parameters   In both cases, we consider three uncertain parameters: reservoir absolute permeability, reservoir porosity and permeability of the leaky well. Their distributions were considered around the corresponding benchmark values defined in Table 12.1. The assumed probability distribution functions for the parameters, required for the selection of the collocation points, are presented in Table 12.2. Actually, Table 12.2 contains mathematical functions fitted to observed distributions of the uncertain parameters. The correlation between absolute permeability and porosity was reconstructed from the U.S. National Petroleum Council Public Database (more than 1200 reservoirs), see also (Kopp et al., 2009). For case 2, we also included two design parameters for describing the injection strategy: the CO$_2$ injection rate (fluctuating around 8.87 kg/s) and the size of the screening interval (up to 30m). The choice of the design parameters in this study is only exemplary and serves to demonstrate how engineering decision-making can be supported by the approach presented here. Both the injection rate and the screening interval directly affect the ratio of forces in the reservoir during the injection. The chosen feasibility functions of design and control parameters are presented in Table 12.3. The choice of feasibility functions is arbitrary, and modelers have full freedom to introduce feasibility functions and to weight them according to their personal experience or preferences. For example, we chose a beta distribution for the size of the screening interval (see Table 12.3) with lower and upper bounds of z=0[m] and z=30[m], respectively, reflecting the physical bounds of the reservoir.

Quantities of interest   Our model outputs of interest are:

(a) Physical characteristics of flow: spatially distributed pressure and saturation as a function of time and the CO_2 leakage through the leaky well as a function of time. CO$_2$ leakage rate is defined in the benchmark study as the CO$_2$ mass flux midway between the top and bottom aquifer divided by the injection rate, in percent.

(b) Stochastic characteristics of flow: mean values, variance and cumulative distribution function.

12.3.3. Case 1: Validation on a small test problem

First, we validate the PCM method against traditional Monte Carlo simulations. We performed a Monte Carlo simulation with 1000 realizations and P simulations (see equation 12.6) for the collocation approach within the simplified benchmark problem. We repeated
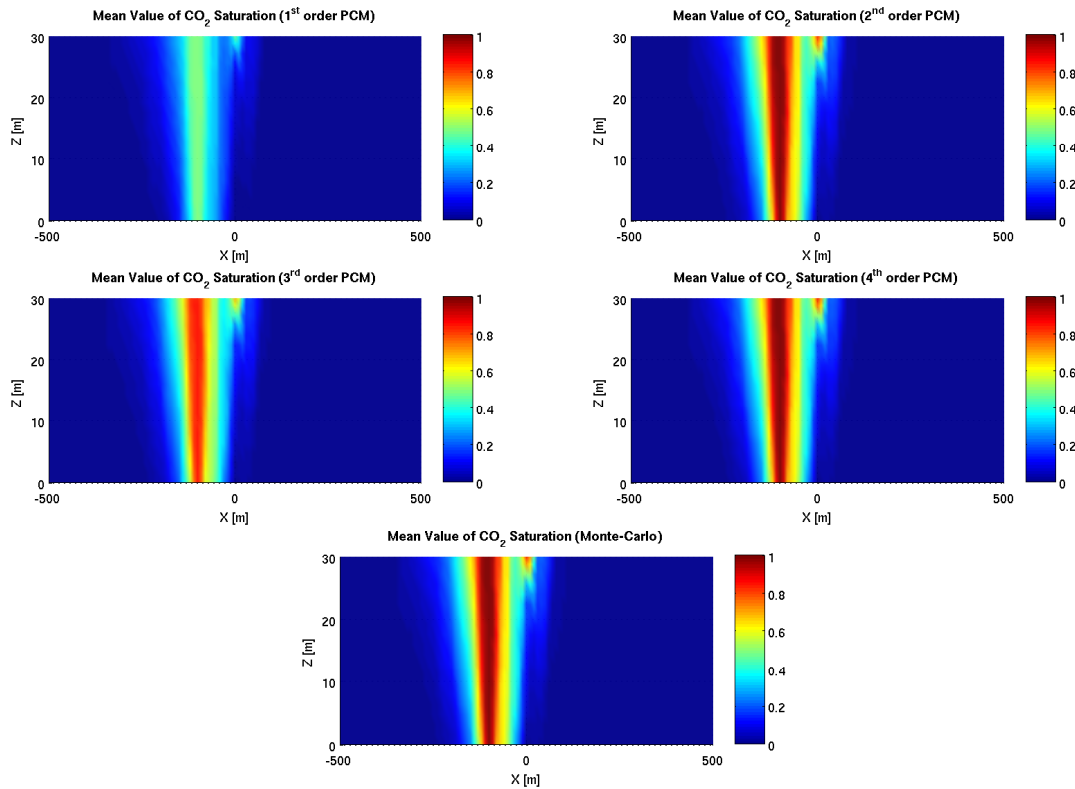
Figure 12.2.: 2D section plot (y=0[m]) for mean value of CO_2 saturation

the comparison study for different degrees of chaos expansion, such as first order (4 samples), second order (10 samples), third order (20 samples) and fourth order (35 samples).

The strongest CO_2 infiltration processes occur between the injection well and the leaky well. Therefore, we show the mean values of CO_2 saturation after 30 days in the 2D section for y=0[m] (Figure 12.2), where leakage well is located at the position x=0[m]. The bottom plot in Figure 12.2 illustrates the reference mean value which was calculated on the basis of 1000 Monte Carlo samples. The other four plots were obtained by the PCM with different degrees of expansion. Evidently, the linear approximation is not adequate to represent the non-linear behavior of multiphase flow. The second-order expansion shows high accuracy at very low computational costs. The higher orders of chaos expansion show even better accuracies (Figure 12.2), but the convergence is not uniform (see the third order), because of a special property of the collocation method that will be discussed in section 12.3.5.

The most integrative characteristic of the overall process is the total leakage of CO_2. Figure 12.3 and Figure 12.4 show the mean value and standard deviation of CO_2 leakage rate as functions of time, respectively. To compare a quantitative characteristic for probabilistic
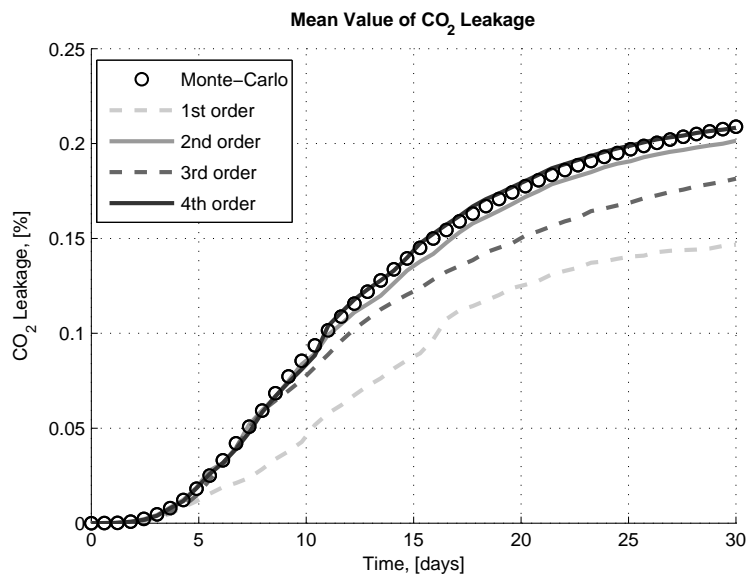
Figure 12.3.: Mean value of CO_2 leakage rate
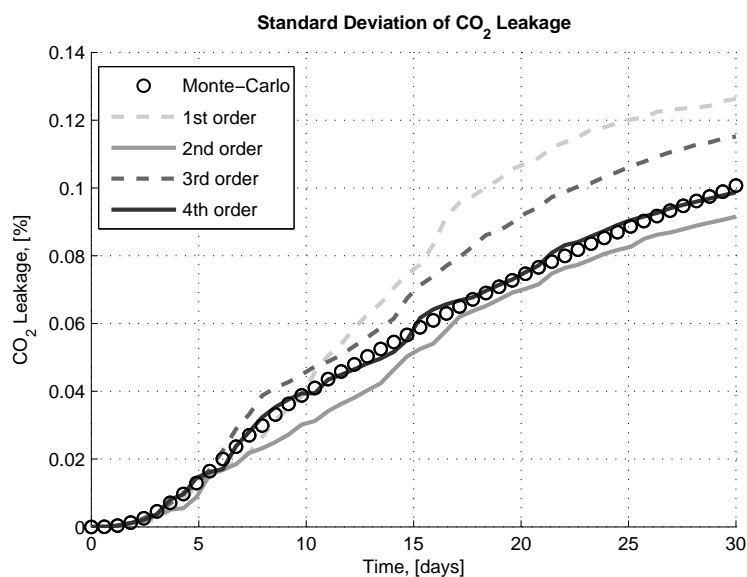
Figure 12.4.: Standard deviation of CO_2 leakage rate

risk assessment, we also computed the cumulative density function of CO_2 leakage rate after 30 days (Figure 12.5). The cumulative density function represents the probability that the random variable is less than or equal to a particular value. For example, the CO_2 leakage rate takes on a value less than 0.3% with a probability of about 80% (see Figure 12.5). In addition, Figure 12.5 shows the value of the CO_2 leakage rate obtained by the determinis-
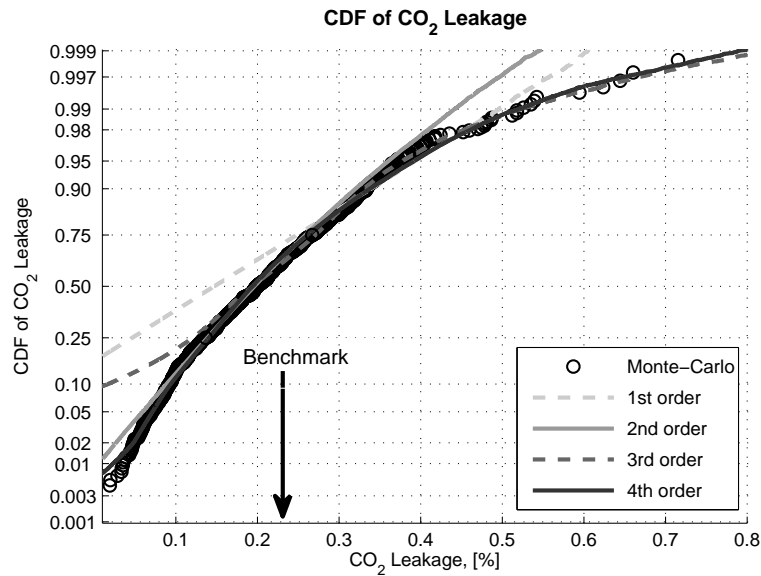
Figure 12.5.: Cumulative distribution function of CO_2 leakage rate after 30 days

tic benchmark prediction. Such a deterministic computation does not take any uncertainty into consideration, and hence cannot quantify a probabilistic risk. The difference between deterministic and stochastic prediction increases with time (see details in section 12.3.4).

Because the collocation approach provides an efficient polynomial representation of the model output, we were able to perform 100,000 Monte Carlo runs on the polynomial in order to construct the cumulative distribution function in virtually no time. The great advantage consists in constructing smooth cumulative distribution functions even in regions where the probability density is very low and Monte Carlo becomes inaccurate (see Figure 12.5).

In summary, the PCM approach provides an effective tool for a probabilistic risk assessment of CO_2 storage; however, this risk assessment is still based on a priori knowledge of probability distributions, in this case on the distribution of the parameters ϕ, K_a, K_l. For our purposes, the second degree of expansion turned out to be sufficiently accurate and can be considered the cheapest reasonable approximation for non-linear transport processes in CO_2 storage.

12.3.4. Case 2: Application to a large problem

We now consider the original-size benchmark problem and additionally include design parameters. Using the IPCM conception, we construct an integrative response surface (section
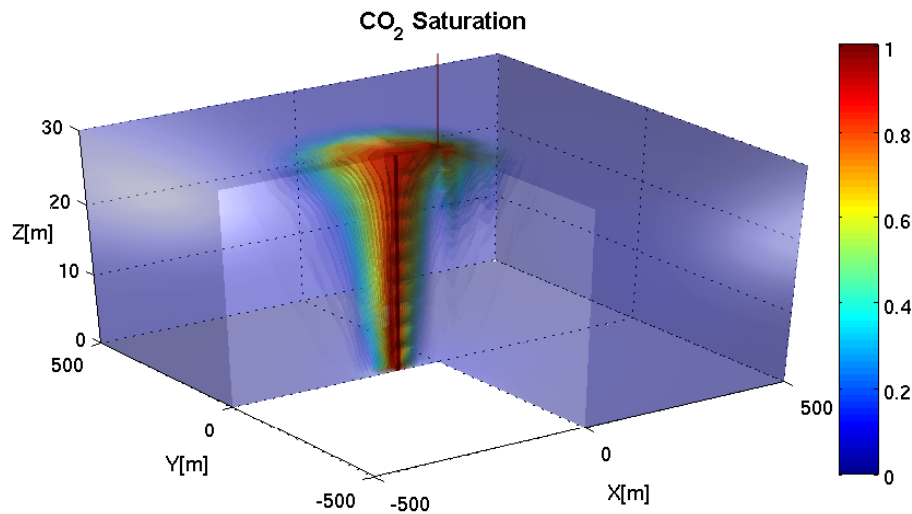
Figure 12.6.: Distribution of CO$_2$ in the reservoir after 100 days

12.2.2) of the model output in the combined 5D parametrical space (ϕ, K_A, K_L, Q_{CO_2}, W_z) and 4D physical space (X,Y,Z,t).

Figure 12.6 illustrates the saturation of CO$_2$ after 100 days obtained by 2nd order IPCM, averaged over the three uncertain parameters, and evaluated at the original benchmark values of the design parameters. Figure 12.7 shows the dynamics of the cumulative distribution function of caprock pressure (100, 500 and 1000 days). This provides quantitative information about caprock pressure exceedance probabilities and can be directly used for predicting critical behavior of the storage formation. In this case, the pressure in the reservoir and thus also at the caprock decreases over time as a result of the boundary conditions in the problem description, cf. (Ebigbo et al., 2007). For the same reason, the leakage rate also decreases after an early peak which marks the arrival of the CO$_2$ at the leaky well. The probability of unacceptable CO$_2$ leakage (Figure 12.8) can be explored to quantify the probability of punishment actions when the CO$_2$ leakage towards the surface exceeds acceptable limits.

All the above results had the values of design parameters fixed to their original values when information was extracted from the integrative response surface. Figure 12.9 demonstrates how the injection rate and screening interval influence the leakage rate of CO$_2$. An important advantage of IPCM is that parameter uncertainty is easily included in such predictions. The top surface in Figure 12.9 is the CO$_2$ leakage rate expected after 100 days as a function of the design parameters, averaged over the uncertain parameters. The bottom surface in Figure 12.9 is the CO$_2$ leakage rate using the expected values of the uncertain parameters
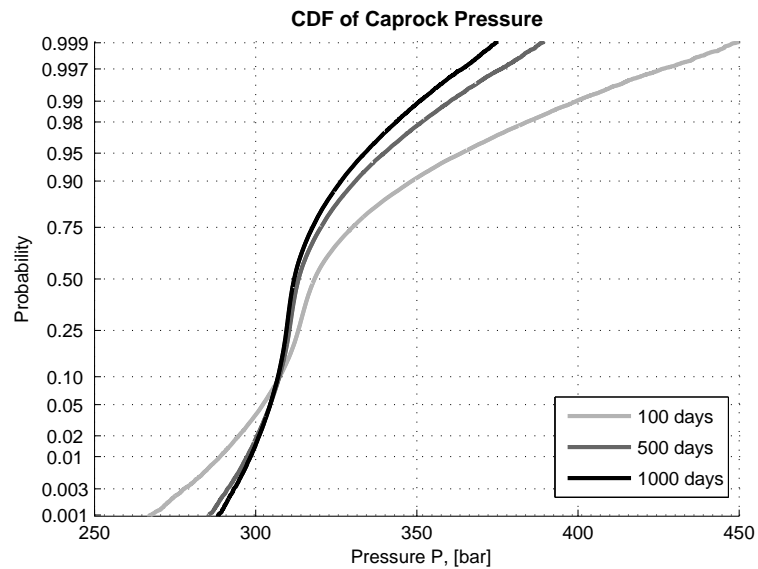
Figure 12.7.: Cumulative distribution function of caprock pressure after 100, 500 and 1000 days
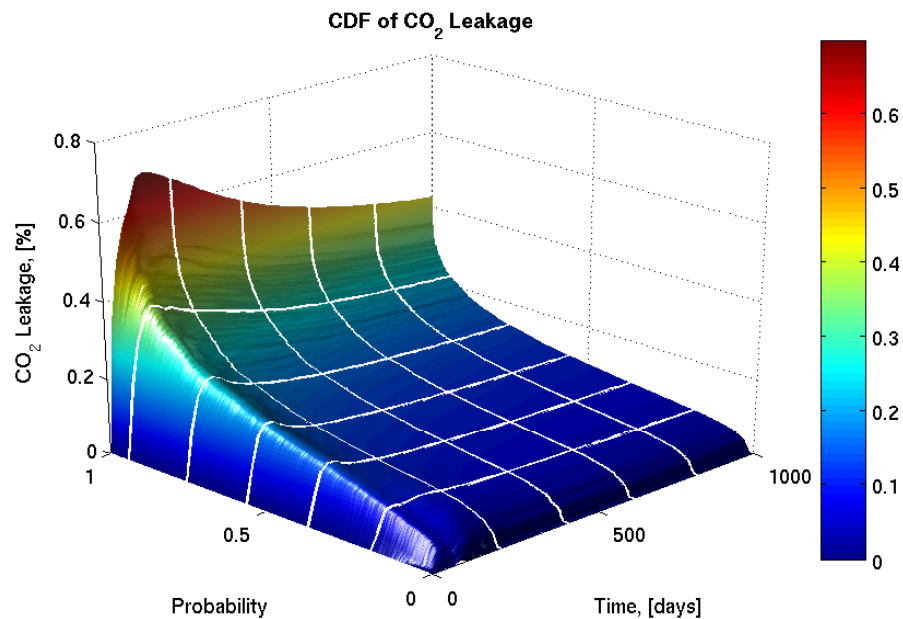
Figure 12.8.: Time dynamics for the cumulative distribution function of CO_2 leakage rate

(see Table 12.1), i.e. as in deterministic simulations. It is easy to see that the impact can be extremely important for non-linear systems (here, a factor of about two), especially in long-term simulations.
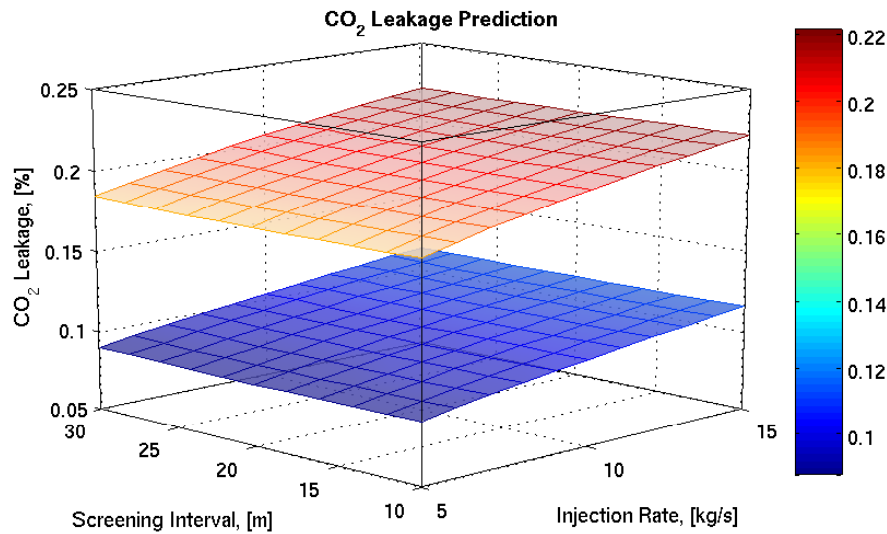
Figure 12.9.: Influence of design parameters on prediction of CO_2 leakage rate after 1000 days: top surface - expected CO_2 leakage rate (average over uncertain parameters); bottom surface - CO_2 leakage rate evaluated pseudo-deterministically with expected values of parameters

In a similar fashion, the dependence of the leakage probability or any other statistical characteristics on design parameters can be evaluated, so that the injection regime can be chosen according to a maximum allowable failure probability. Figure 12.10 illustrates the choice of design parameters based on the caprock pressure after 1000 days. In this test case, a critical caprock pressure equal to 330 bar was chosen at a significance level of 5%, i.e. the maximal acceptable probability of failure is set to 0.05 (solid black line on surface). Figure 12.10 demonstrates acceptable strategies of injection where the caprock pressure does not exceed the limit of 330 bar, which corresponds to an injection-induced pressure build-up of about 40 bar.

In this way, the proposed approach provides a constructive solution to the problem of robust design under uncertainty and provides valuable support for risk-informed decision making.

12.3.5. Discussion

The simulation of complex large-scale systems, such as multiphase flow problems in underground reservoirs, is limited by computational power. This leads to simplifications of models on different levels, such as coarse numerical schemes, neglecting uncertainty, simplifying
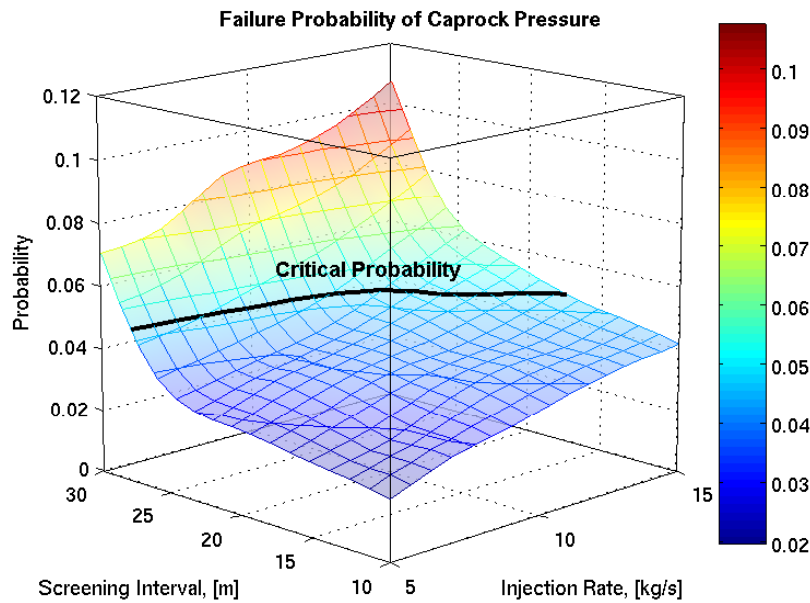
Figure 12.10.: Choice of design parameters based on caprock pressure after 1000 days: critical pressure 330 bar at a significance level of 5 %

or neglecting physical processes etc., especially for large-scale simulations. The greatest challenge to modeling consists of finding a healthy and reasonable compromise between an accurate system representation and computational efforts. In section 12.3.4, we showed that neglecting uncertainty can be a strong simplification for modeling CO$_2$ sequestration, and the consequences can be stronger than when neglecting several physical phenomena (e.g. phase transition, convective mixing, capillary forces etc.).

The underlying idea of the work presented here relies on the application of the polynomial expansion even to discontinuous functions such as CO$_2$ saturation. In fact, this causes a smoothing of the saturation front in parametrical space. However, it conserves the discontinuity of physical processes in space and time. This leads to a reasonably accurate approximation even for discontinuous variables such as saturation, but offers a field of future research to find discontinuous expansion bases.

From the practical point of view, the computational costs of IPCM are proportional to the number of terms in the chaos expansion multiplied by the time of a single simulation run. In section 12.3, we presented two small-scale examples, where the computational time for single simulation runs was about 1 hour and 20 hours for the coarse (section 12.3.3) and the fine (section 12.3.4) model, respectively. Large-scale models can be feasible as well, because

of the evident potential for trivial parallelization (see section 12.2) and because of possible incorporation with readily developed high-performance computing codes for simulating the system (see (Class et al., 2009), (Flemisch et al., 2007), (Bastian et al., 2008), (Lacroix et al., 2001)). The total number of model evaluations depends on the desired degree of expansion and the number of considered parameters (see section 12.2). Theoretically, there is no limitation to the maximum number of parameters, however, in reality, the number of parameters is limited by the number of model evaluations according to the available computational power. The presented methodology has strong advantages over Monte Carlo, when the number of parameters remains below approximately 100 (see e.g. (Xiu and Karniadakis, 2003), (Li and Zhang, 2007)). The fastest estimation could be computed using a linear approach; however, the precision of such linearization for non-linear flow is not acceptable. On the other hand, too high degrees of expansion (here: more than 4th order) quickly become too expensive, and cannot be considered a realistic alternative to Monte Carlo approaches for large numbers of parameters. Thus, we see the compromise between reasonable precision and fast computation using the 2nd, 3rd and 4th orders. However, the 3rd order of expansion showed greater inaccuracy than the 2nd order. This is based on the fact that not one of the collocation points for 3rd order coincides with the central point of largest probability (zeroth root in normal space), see also (Li and Zhang, 2007). Hence, no information about the most probable region enters the statistical computations. This is a propriety of numerical integration which is caused by the transformation from Gaussian space to physical space. To avoid this problem, an advanced technique is proposed in (Oladyshkin et al., 2011a). As a reasonable compromise, we recommend the second order as the cheapest, but still sufficiently accurate approach for evaluating non-linear processes.

12.4. Conclusions

In this work, we provide a massive stochastic model reduction via the polynomial chaos expansion. We use the "black-box" collocation technique to project the space and time model response surface onto a higher-order orthogonal basis of polynomials. This allows for the non-linear propagation of parameters uncertainty affecting the predicted quantities, ensures fast computation and provides a powerful tool for joining design variables and uncertain variables into one approach based on an integrative response surface in the form of an explicit polynomial expression. This offers fast evaluation for statistical quantities and their dependence on design or control parameters.

We recommend the second order of polynomial expansion as a reasonable compromise between computational effort and accuracy. The proposed stochastic approach was validated on the basis of Monte Carlo simulation using a common 3D benchmark problem. In this case study, our proposed approach yielded a significant speed-up of 100: 1000 Benchmark runs for the Monte Carlo evaluation were comparable in accuracy with 10 benchmark runs using the probabilistic collocation method.

A specific novelty is that we project all the design parameters and uncertain parameters onto one single integrative response surface. Based on this integrative concept, the design task explicitly includes uncertainty, which leads to robust designs with minimum failure probability. Thus, the integrative response surface provides a powerful tool for probabilistic prediction and robust design of non-linear systems and provides valuable support for risk-informed management decisions.

We demonstrated that neglecting parametric uncertainty constitutes a strong simplification for modeling CO_2 sequestration. Due to the non-linearity of CO_2 infiltration, including uncertainty leads to a systematic and significant shift of the predicted leakage rates (here towards higher values), affecting both risk estimates and the design of injection scenarios.

13. Pressure assessment in a sand channel system during CO$_2$ injection

Bibliographic Note: The content of this chapter is based on the following original article: Walter L., Oladyshkin S., Class H., Darcis M., Helmig R., A study on pressure evolution in a sand channel system during CO$_2$ injection. Energy Procedia, Elsevier, V. 4, P. 3722-3729, 2011.

Risk assessment and feasibility studies based on numerical simulations are essential for a realisation of large scale carbon capture and storage projects. The numerical simulation of CO$_2$ storage in deep saline aquifers, which is focused on in this study, is very demanding with respect to computational costs. During CO$_2$ injection it is important to observe the pressure increase since it might result in caprock failure, reduction of storage capacity or brine displacement. For many reservoir parameters like e.g. the permeability only few information exists. More over the geological structure of the reservoir, e.g. the size and distribution of high permeable sand channels is often not known in detail. To deal with the resulting uncertainties the integrative probabilistic collocation method is applied to a channel system scenario. The influence of the dimension of the high permeable channel and the permeability on the pressure evolution is investigated in detail. Additionally, the maximum allowable injection rate is predicted with the method.

13.1. Introduction

CO$_2$ storage in saline aquifers is recently intensively investigated as one option to reduce greenhouse gas emissions. One challenging task concerning this issue is the development of methods for assessing the risks during CO$_2$ injection and the post-injection phase. CO$_2$

leakage and brine displacement and infiltration into higher (drinking water) aquifers far away from the injection site are the major risks, which need to be considered and minimized during CO_2 storage. The major problem for performing risk assessment is the uncertainty in many reservoir parameters like permeability, porosity, heterogeneity, since there is often only few information available for these parameters. To cover all these uncertainties numerous Monte Carlo simulations would be required, which is very time consuming and causes high computational cost. The paper (Kopp et al., 2010) already presented a risk assessment method for a leakage scenario, where a limited number of most relevant parameters was used and risk was estimated by considering their range of variability. To reduce the computational cost we use a massive stochastic model reduction based on the integrative probabilistic collocation method, which was developed for CO_2 storage in (Oladyshkin et al., 2011b). The idea behind this method is simply to replace the full model by a response surface that is able to represent the model output (e.g. leakage, pressure) dependent on the chosen uncertain and design parameters. The method is tested here on a simplified example for assessing the pressure increase depending on different uncertain parameters. The pressure increase is important to consider since too large pressures might lead to caprock failure near the injection well, which might result in new leakage pathways. Far away from injection, the pressure increase also has to be taken into account because it might influence the pressure in other storage reservoirs and might reduce the overall storage capacity. Additionally, brine displacement due to pressure increase has to be investigated since infiltration into higher (drinking) water aquifers need to be avoided. In this scenario it is focused on the influence of the structure of the reservoir therefore the permeability and the dimension of a high permeable sand channel are varied.

13.2. Integrative probabilistic collocation method

The integrative probabilistic collocation method (IPCM) is described in detail in (Oladyshkin et al., 2011b). Here we will only present the three main steps. The overall basis of this method is the stochastic response surface approach, which consists of (1) the expression of the uncertain input variables as standard distributed random variables, (2) the formulation of a polynomial chaos expansion for the output variables and (3) the determination of the coefficients in the expansion (see (Isukapalli et al., 1998)). In this work, we use the polynomial chaos expansion with Hermite polynomials as a basis. To determine the unknown coefficients (step 3) we apply the probabilistic collocation method. The resulting surface is

called integrative response surface, which contains all design and uncertain parameters.

As already mentioned, normally distributed variables are needed for the chosen method. However the uncertain parameters as the permeability or the channel dimensions are not normally distributed since no negative values occur. The construction of a reliable probability density function of the uncertain parameters can be performed with the help of even little information, e.g. bore hole information about the permeability. For the method, the physical distribution (e.g. lognormal for the permeability) has to be transformed to normal space.

In the second step, the output can be represented by the polynomial chaos expansion (Wiener, 1938). The model output Ω depending on the input parameters can be represented by the expansion as:

$$\Omega(\omega) = \sum_{j=1}^{P} c_j \Psi_j(\omega) \qquad (13.1)$$

The coefficients c_j quantify the dependence of the model output Ω on the input parameters ω. The function Ψ_j is a simplified notation of the multi-variate orthogonal polynomial basis for ω. The formulas are presented in (Oladyshkin et al., 2011b). The CO₂ storage problem is solved at each grid point in space and time, therefore the polynomial form depends on the three spacial dimensions and the time. The number of terms P, which have to be solved, depends on the number of uncertain input parameters N and the order of the expansion d. We choose the second order polynomial expansion due to a compromise between computational time and adequate accuracy (see (Oladyshkin et al., 2011b)).

$$P = \frac{(N+d)!}{N!d!} \qquad (13.2)$$

For the third step, different methods can be applied to determine the unknown coefficients. In our case the coefficients are obtained by the collocation technique. A set of parameters with the dimension N is called a collocation point. The number of required collocation points is the same as the number of terms and unknown coefficients P. This means that P model runs are needed for obtaining the unknown coefficients. If the coefficients are determined, a set of polynomials for calculating the output as pressure and saturation are available. The optimal choice of collocation points are the roots of the polynomial one degree higher than the order used in the chaos expansion. The polynomial surface is exact directly at the collocation points and a polynomial extrapolation or interpolation in between. For our multidimensional problem the number of available collocation points is higher than the required number for calculating the coefficients. One possibility for choosing the collocation points is a totally

random selection. In this work, the other option, where the points are selected from the most probable regions of the input parameters is used. After the determination of the unknown coefficients a set of polynomials is available for post-processing computations. This means that the model output depending on the uncertain and design input parameters is projected on one single integrative response surface. The mean values and the variance can also be easily computed with the polynomials. For obtaining further statistical parameters, Monte Carlo simulation on the polynomials can be performed. For more details see (Oladyshkin et al., 2011b), (Isukapalli et al., 1998) and (Li and Zhang, 2007).

13.3. The channel scenario

In order to investigate the influence of different reservoir structures like e.g. high permeable channels on the pressure increase during CO_2 injection, the IPCM is applied for modelling a channel scenario. In reservoirs with high permeable sand channels CO_2 will mainly spread inside these channels. Depending on the dimension and the permeability of the channel the pressure increase in the reservoir will change. In the presented scenario no site specific data is used, however the scenario can be easily assigned to realistic data, which will be subject of future work. Figure 13.1 shows the model domain with the variable channel enclosed by a larger domain with lower permeability. The outer domain beyond the channel has the function to reduce the influence of the boundary conditions on the pressure field. For boundary conditions no flow (Neumann) conditions are applied on all sides except the back end of the domain. Here constant head boundary conditions (Dirichlet) are set. The CO_2 injection takes place over the whole height of the permeable channel on the left front edge of the channel (see red cycle in Figure 13.1). This means that the injection area changes with changing channel height and the CO_2 will mainly spread into the permeable channel. The simulation period is two years of injection.

As uncertain parameters, the permeability, the length, the width and the height of the channel are taken since for these parameters often only few information exists. Additionally the injection rate is chosen as a design parameter. Finally, five variable input parameters ($N = 5$) and the second order of polynoms ($d = 2$) results in 21 terms in chaos expansion (equation (13.2)) and 21 simulation runs ($P = 21$, see equation (13.2)). Figure 13.2 shows the distributions of the uncertain parameters and the feasibility function for the injection rate. Taking the injection rate as a design parameter into account can help to investigate how the pressure increase reacts with changing injection rate. Therewith the maximum possible injection rate for a
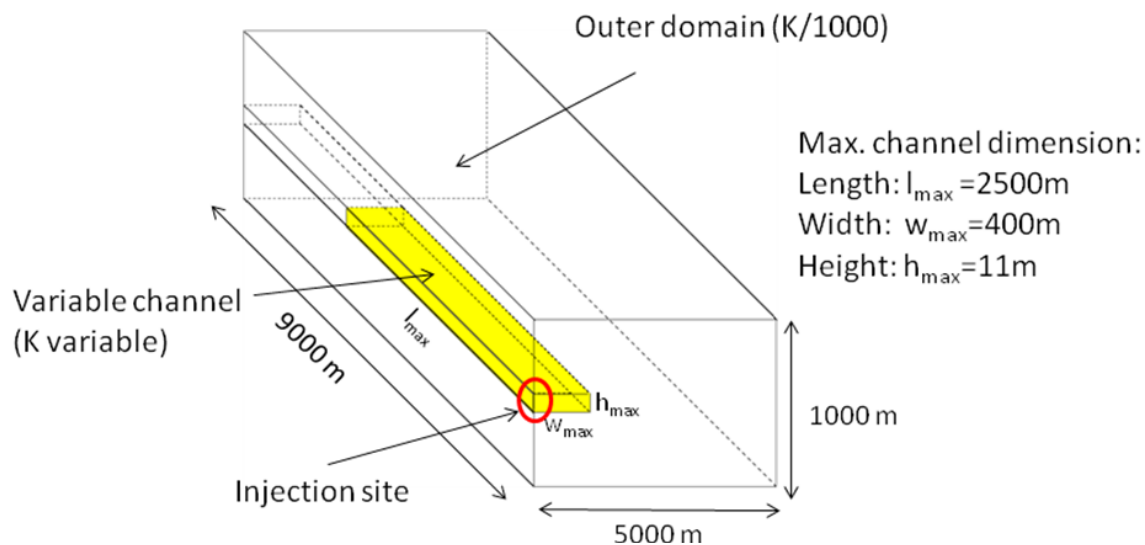
Figure 13.1.: The simulation domain with variable channel (Walter et al., 2011).

given threshold value for maximum allowable pressure increase at the caprock can be easily estimated. To construct the response surface the model should be evaluated 21 times with 21 different parameter sets. For the simulation we use DuMuX, a multiphase-multicomponent simulator for flow in porous media (see (Flemisch et al., 2007), http://www.dumux.de). A two phase (2p) model, which does not account for compositional effects, is chosen. Compositional effects are neglected here, since we mainly focus on the pressure increase during the injection period and not on the long term plume behavior. However, the pressure peak is slightly overestimated using a 2p model without compositional effects, thus the assessment is conservative.

13.4. Results

In this section different possible evaluation methods are shown for a demonstration of the method and an investigation of the influence of uncertain and design parameters on the pressure increase. For evaluation, two interesting points of the domain are chosen for measuring the pressure. One is located directly at the injection well and the second one is located in a distance of 350 m to the injection well (along the length of the channel). At both locations the pressure is measured in the overlying caprock (2 m above the channel) since the pressure in the caprock is most interesting for risk assessment as it was discussed in Section 13.1. In Figure 13.3, the mean pressure increase obtained with the IPCM, during two years
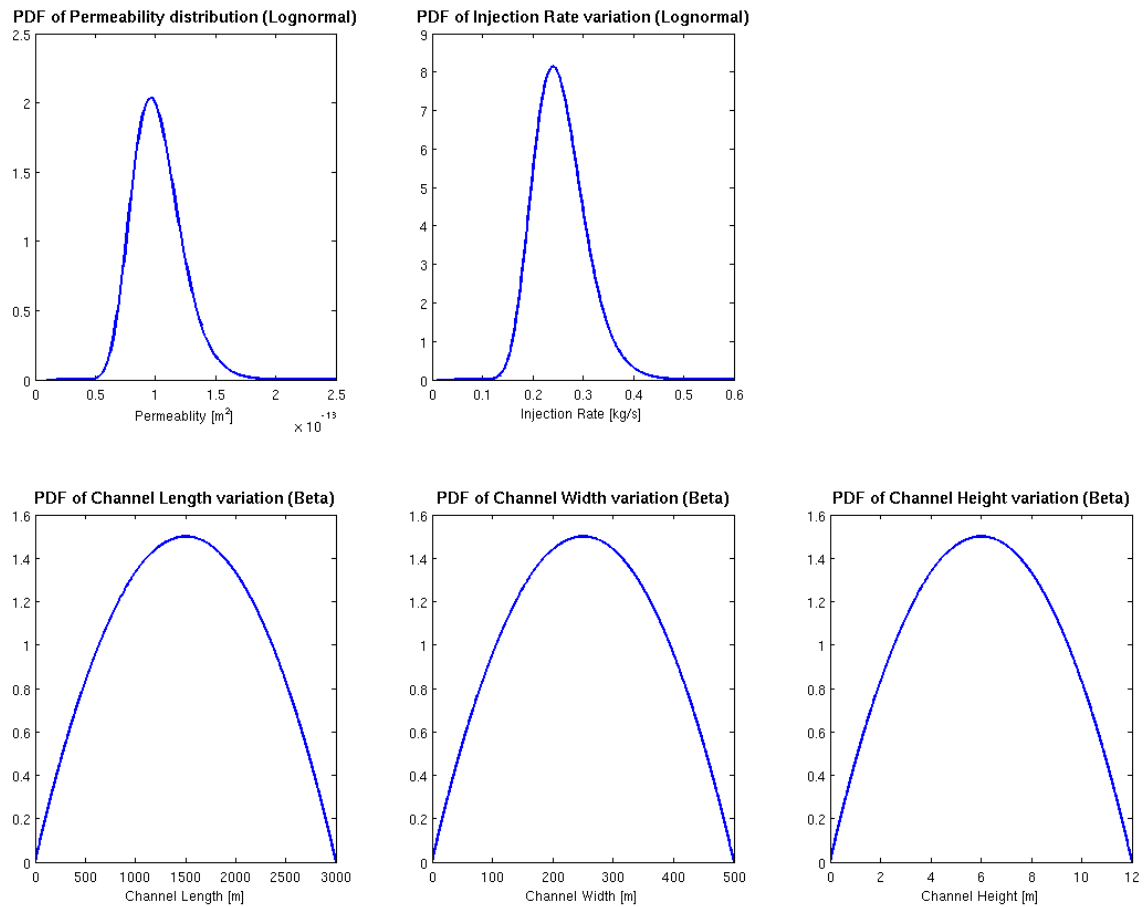
Figure 13.2.: Distributions of the uncertain parameters and feasibility function for the injection rate

of injection at the two measurement locations is shown. At the beginning of the injection the pressure increases strongly. After 100 days only a slightly further increase occurs. When injection starts the relative permeability is very small, therefore a high resistance results in high pressure peaks. With increasing saturation the relative permeability and thus the resistance is decreasing.

Interesting to see, is how the pressure signal is affected by changing uncertain parameters. With this method the dependencies can be easily presented. The left plot in Figure 13.4 shows the response surface in 3 dimensions, which is obtained by varying the injection rate and the permeability. For the dimensions of the channel mean values are constructed using 1000 Monte Carlo simulation on the obtained polynomials. This surface easily shows that a higher permeability decreases the pressure due to lower resistance and a higher injection
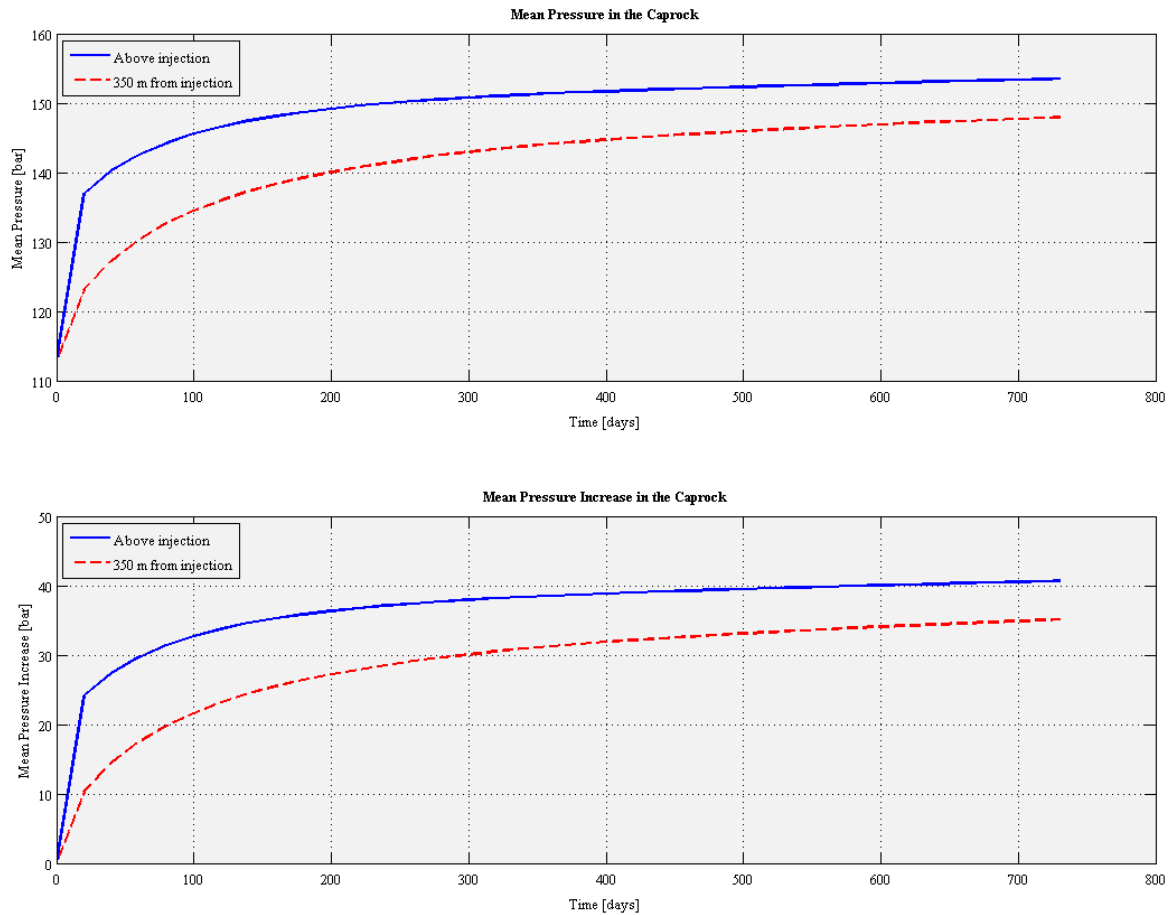
Figure 13.3.: Mean pressure increase at two different locations in the caprock

rate increases the pressure. For a more detailed analysis the dependence on the injection rate is plotted as a 2D curve (see the right plot in Figure 13.4). Now the permeability is also set to the mean value. For increasing injection rate one would expect a linear increase of the pressure. However the non linear slope in the right plot of Figure 13.4 results from the compressibility of the fluid phases. With changing pressure the volume also changes, therefore the pressure injection rate curve shows a slightly nonlinear run.

The Figure 13.5 shows how the pressure is increased with variable channel dimensions at the two measurement locations. In each figure one dimension (length or height) is variable and for the remaining input parameters the mean values are set. Obviously, the pressure increase falls with larger dimensions since a larger high permeable region with less resistance for the CO$_2$ exists. As already shown in Figure 13.3 the pressure increase is lower with proceeding time. In overall comparison, the pressure increase 350 m away from injection is up to 20
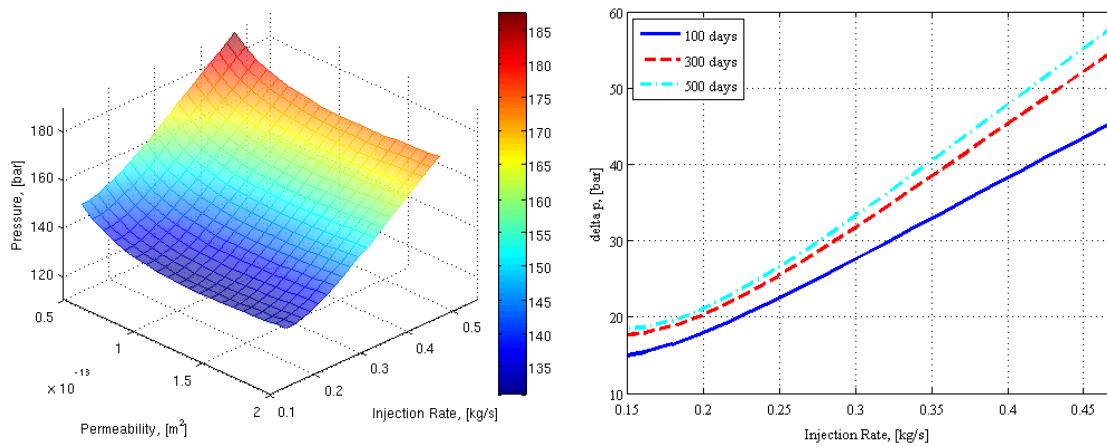
Figure 13.4.: Influence of permeability and injection rate (left) and influence of the injection
rate (right) on the pressure at injection

bar lower than at the injection well. The dependence of pressure on the channel width is not
shown since it shows a similar run as the dependence on the channel length.
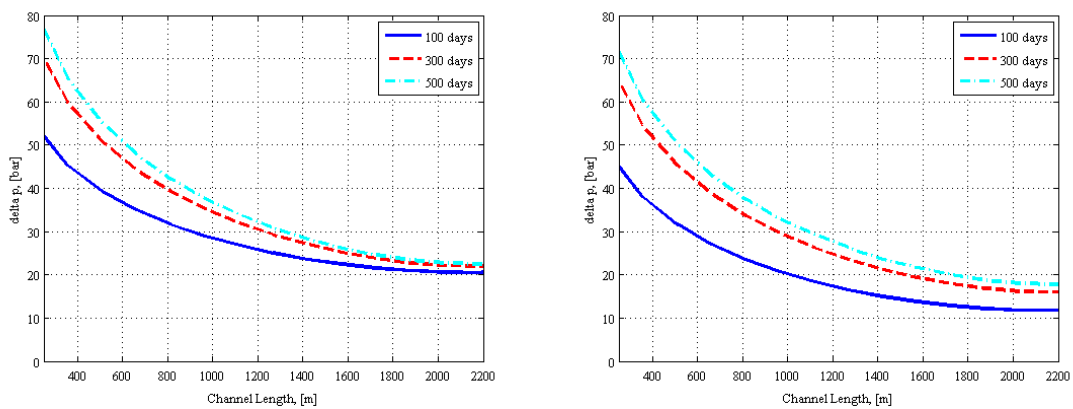
Figure 13.5.: Influence of the channel length: a) at injection and b) in 350 m distance from
injection along the channel

An interesting point is the horizontal run of the pressure curve with channel length larger
than 1800 m (see the left plot in Figure 13.6). This means that in this case study, the pressure
at the injection well does not further decrease with larger channel length than 1800 m. The
pressure gradient at the end of the channel is very small when long channels are considered.
One reason is that the ratio of compression and displacement becomes smaller with longer
channels. A change in the channel length higher than 1800 m results only in very small
pressure changes. This change has no influence on the injection well pressure due to the
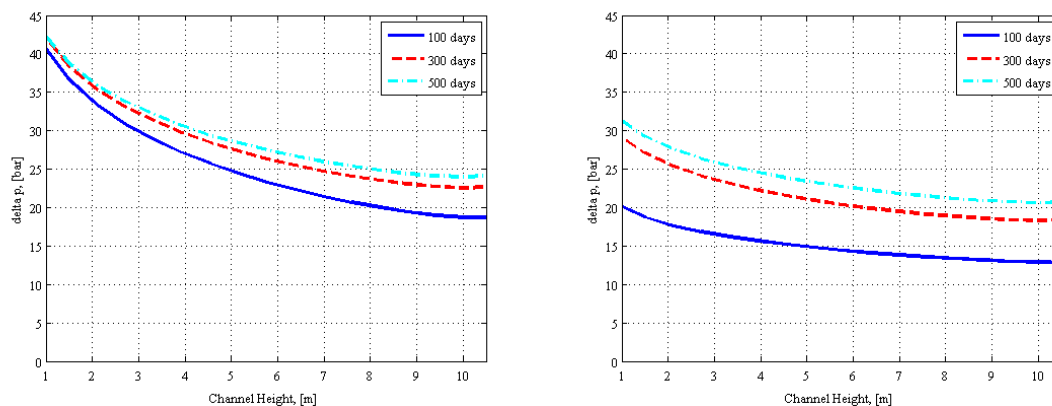
Figure 13.6.: Influence of the channel height: c) at injection and d) in 350 m distance from injection along the channel

small value and the long distance from injection.

Another interesting point is the difference in pressure increase with time between channels with smaller and higher height (see the right plot in Figure 13.6 and Figure 13.7). In a thinner channel the same amount of CO_2 is injected into a smaller volume (see Section 13.3). So, the pressure peak at the beginning of injection is much higher. The CO_2 front moves faster through channels with smaller height. Due to the long distance between the front and the injection well and the low pressure at the front in comparison with the high injection pressure, a further pressure increase at the front does not influence the injection pressure. The presented curves demonstrate the strong influence of changes in the storage site structure (here the channel dimensions) in the domain.

Another powerful tool, which can be provided by IPCM, is the cumulative distribution function (CDF). For the determination of CDFs Monte Carlo has to be performed on the generated polynomials. One possible assessment of CDFs is exemplary shown in Figure 13.8. For CO_2 injection one critical parameter is the maximum possible pressure in the caprock. This threshold value is determined in a way that any damage of the caprock is avoided. In this fictive scenario a caprock threshold value of 60 bar is chosen. CDFs with three different injection rates are performed. For all other uncertain parameters, the mean values determined from Monte Carlo on the polynomials are taken. The significance level is 5%, which means that the overpressure is accepted if it is below the threshold value with a probability of 95%. The threshold value of 60 bar overpressure is only not exceeded with a probability of 95% by the case with the smallest injection rate (0.15kg/s). If higher injection rates are applied this is not further warranted. With this method the maximum allowable injection rate can be
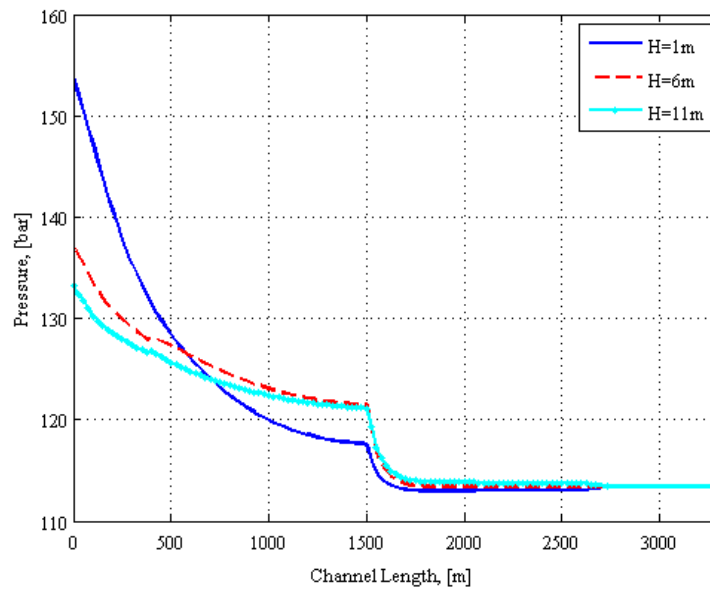
Figure 13.7.: Section along the channel, with three different channel heights (H) and mean values for the remaining parameters
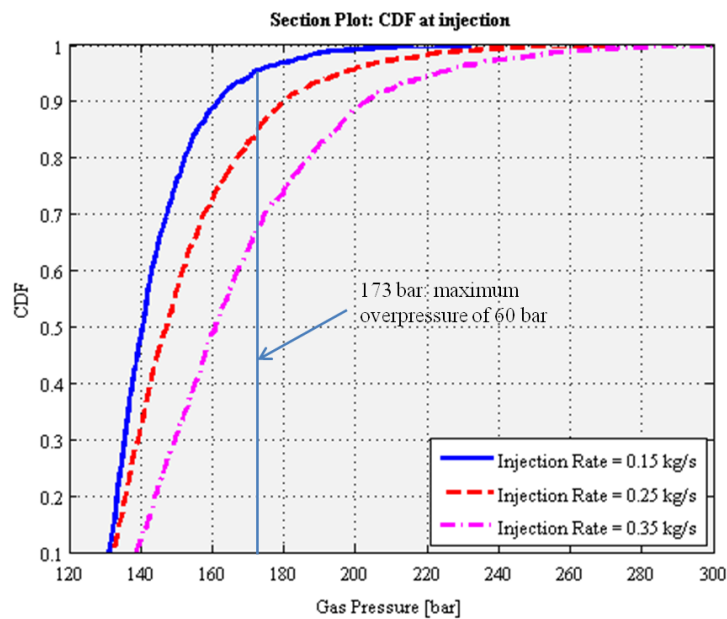
Figure 13.8.: Cumulative distribution function for pressure at injection (initial pressure: 113 bar)

easily accessed by using CDFs. Thus, probabilistic informed prediction can be performed, i.e. robust design.

13.5. Conclusions

In this work, detailed analyses on the pressure evolution in a channel system with low computational cost are demonstrated. Several powerful tools for investigations of the influence of different uncertain and design parameters on the pressure increase are performed. The shown channel scenario clearly demonstrates how the structure of the storage site strongly influences the pressure increase in the domain. Several characteristic behaviours and their visibility in the results of the probabilistic collocation method were discussed and a method to determine the maximum allowable injection rate was presented.

In the future this method is planned to be applied to much larger and more realistic CO_2 storage sites for assessing the different risks during injection and in post-injection phase. For risk assessment of pressure increase and brine displacement far away from injection, large domains and long time periods must be considered (see (Birkholzer and Zhou, 2009) and (Schäfer et al., 2011)). In order to deal with these large domains and long simulation times, the idea is to further reduce the computational cost by model coupling in time and space ((Darcis et al., 2009) and (Darcis et al., 2011)).

14. Data-driven uncertainty quantification of CO$_2$ storage

Model-based uncertainty analysis can help to judge the potentials and hazard in many engineering applications better. This requires to specify the probability distributions of all model parameters, posing a huge demand on data availability or requiring highly subjective assumptions on distribution shapes to compensate for missing data. We present a minimally subjective approach for uncertainty quantification in data-sparse situations, based on a new and purely data-driven version of polynomial chaos expansion (PCE). It avoids the subjectivity that is otherwise introduced when choosing among a small limited number of theoretical distribution shapes to represent natural phenomena: we only demand the existence of a finite number of statistical moments, and do not require knowledge or even the existence of probability density functions for input parameters. In a small fictitious example with independent experts, otherwise, we demonstrate that this subjectivity can easily lead to substantial prediction bias, and that the subjective choice of distribution shapes has a similar relevance as uncertainties due to physical conceptualization, numerical codes and parameter uncertainty. With our approach we can directly and most flexibly use raw data sets available from global databases or soft information from experts in the form of arbitrary distributions or statistical moments. We illustrate and validate our proposed approach by a comparison with a Monte Carlo simulation using a common 3D benchmark problem for CO$_2$ injection, which is a low-parametric homogeneous system. We obtain a significant computational speed-up compared with Monte Carlo as well as high accuracy even for small orders of expansion, and show how our novel approach helps overcome subjectivity.

14.1. Introduction

14.1.1. Modeling carbon dioxide storage

CO$_2$ storage in geological formations is currently being discussed intensively as an interim technology with a high potential for mitigating CO$_2$ emissions. In recent years, greate research efforts have been directed towards understanding the physical processes in CO$_2$ storage (e.g. (IPCC, 2005)). The multiphase flow and transport processes involved are strongly non-linear. They include phase changes in the region of th critical point and effects such as gravity-induced fingering and convective mixing as well as geo-chemical and geo-mechanical processes etc. In order to describe the space-time evolution of the CO$_2$ plume and the influence of potentially leaky abandoned wells, (semi-) analytical solutions have been derived by Nordbotten et al. (Nordbotten et al., 2005a, 2009). A comparison study of various simplifying semi-analytical models with complex simulations models was discussed by Ebigbo et al. (Ebigbo et al., 2007). The analysis in (Birkholzer et al., 2009) focused on the effects of large-scale CO$_2$ leakage through low-permeability layers. Various optimization strategies for monitoring surface leakage using near-surface measurement approaches were developed in (Cortis et al., 2008). Recently, Class et al. (Class et al., 2009) published a benchmark study comparing a number of mathematical and numerical models of varying complexity. In the current chapter, we would like to attract attention to the often ignored or forgotten fact, that efforts invested in improved physical conceptualization, numerical codes and stochastic modeling can easily be overwhelmed by error through human subjectivity in data interpretation and in histogram analysis at a very early stage of modeling (see Section 5.3). The approach developed in this chapter will help overcome this problem, especially in data-sparse situations when only global databases for similar sites and expert opinions are available. A direct application of our proposed framework can be considered best at the stage of site screening, when considering relatively low-parametric large-scale models, prior knowledge and measured data on large scales. In more detailed site-specific studies, more resolved structural models will be more useful, as defined site investigation and new data allow the spatial resolution of geological heterogeneity and Bayesian updating.

Modeling underground CO$_2$ storage involves many conceptual and quantitative uncertainties (Hansson and Bryngelsson, 2009a). The lack of information on subsurface properties may lead, depending on the specific question at hand, to parameter uncertainties up to a level where the uncertainties dominate or even override the influence of secondary physical processes (see (Oladyshkin et al., 2010)). A significant part of the scientific community

still refrains from considering uncertainty in modeling, but the corresponding arguments are discussed and rejected one by one in (Pappenberger and Beven, 2006). Still, current numerical simulation models are inadequate for stochastic simulation techniques based on brute-force Monte Carlo simulation and related approaches (e.g. (Maltz and Hitzl, 1979), (Robert and Casella, 2004)), because even single deterministic simulations may require parallel high-performance computing. The necessity for reasonably fast stochastic approaches for modeling CO_2 sequestration poses an urgent research task that needs to be investigated as soon as possible. The current chapter follows a recent line of development that uses a massive stochastic model reduction based on the polynomial chaos expansion technique, but focuses additionally on the impact and minimization of human subjectivity when assigning distribution shapes to represent input parameter uncertainty in stochastic modeling.

14.1.2. Uncertainty and subjectivity in data interpretation

Stochastic models require specifying probability density functions (PDFs) for all uncertain input parameters. Unfortunately, precise information on distribution shapes is very rare in realistic applications, such as underground reservoir simulations, groundwater modeling, etc. Applied research on real-world systems, like modeling CO_2 storage within site-screening procedures, often faces the problem of immensely limited information about the model parameters involved, e.g. reservoir permeability, porosity, etc. With only limited data available or even a total absence of data, not even probability density functions representing the lack of knowledge can be easily inferred in a justified manner. Moreover, even if some amount of data is available, the statistical distribution of the corresponding model parameters can be nontrivial, e.g. bounded, skewed, multi-modal or discontinuous. Several examples of real data types can hardly be described by known families of theoretical statistical distributions. A straightforward example is the world-wide reservoir-permeability and -porosity data available from public databases (e.g. http://www.npc.org) presented in Figure 14.1, illustrating a possible empirical distribution shape of reservoir parameters. Sophisticated approaches to assigning distributions to data include maximum likelihood or moment-fitting of selected parametric distributions that satisfy the personal perception of the modeler. Due to the typically low numbers of parameters (e.g. two or three), this resembles a subjective and highly restrictive assumption on the distribution shape, its higher-order moments, and its extreme-value tails. Methods of maximum entropy (Jaynes, 1982) and minimum relative entropy (Woodbury and Ulrych, 1993) are often used in the engineering sciences to construct probability distributions from sparse prior information (mostly in the form of a few statistical

moments and bounds). The required information may be available from different sites with supposedly similar conditions. Although these two methods are designed to minimize subjectivity, they are heavily debated within the statistics community (e.g. (Sambucini, 2007)) and in fact still impose a specific assumption on distribution shape. Thus, in any case, the attempt to construct probability density functions from samples of limited size or from sparse information introduces additional room for subjectivity into the analysis.

In our example application to CO$_2$ injection (see Section 14.3.4), we will provide a striking example of bias introduced by subjectivity. We will relate possible errors due to this type of subjectivity to the magnitude of error introduced by numerical schemes, reduced physical complexity and uncertain parameter values. This will illustrate that subjectivity in assigning distribution shapes is not only a philosophical problem, but can severely bias or even invalidate all subsequent results, including uncertainty quantification and risk assessment.

We conclude that applied tasks demand the direct handling of arbitrary distribution shapes and sparse data sets without additional assumptions. To make this possible, we claim that it is not even necessary to define the available prior information with a probability density function. Instead, we will show that the available information can be used directly and most purely, when employing a certain class of expansion techniques for stochastic modeling.

14.1.3. Approach and contributions

The purpose of the current chapter is to work with a highly parsimonic and yet purely data-driven description of uncertainty. We apply a very recent polynomial chaos expansion (PCE) technique that can be applied to arbitrary distributions of data, which is called the arbitrary polynomial chaos (aPC). PCE techniques and the aPC are discussed in detail in Section 14.2. The aPC can use unspecified probability measures for model parameters that are implicitly and incompletely defined via a counted number of statistical moments only, and requires no additional information. In fact, our own derivation of aPC will explicitly show and then exploit the fact that statistical moments are the only source of information that is propagated in all types of polynomial expansion-based stochastic approaches. Thus, exact probability density functions for uncertain input parameters do not have to be known and do not even have to exist. This avoids the necessity of assuming subjectively or speculating on the exact shapes of probability distributions.

The new freedom in distribution shapes gained with aPC opens the path to accessing with PCE even those applications where data samples of limited size merely allow the inference

of a few moments, and one would not be able to construct a probability density function without introducing subjective assumptions and hence a dangerous sources of bias.

After understanding that only moments matter when considering polynomial expansion, Maximum Entropy (ME) methods are indeed minimally subjective in the sense that they require only the same moments for defining distribution shapes that are also required to construct a PCE. Since both ME and our method work with the same information units, they are suitable for combination.

There are only very few studies that have used aPC before, and they can only be found in mathematical stochastics (Soize and Ghanem, 2004), (Ghanem and Doostan, 2006) and aerospace engineering (Witteveen and Bijl, 2006), (Witteveen et al., 2007). These studies focused on proofs of existence, constructing the basis by Gram-Schmidt orthogonalization and related, quite theoretical issues. A notable exception is the very recent study (Li et al., 2011) in field of petroleum engineering. That study did not discuss, however, the aPC in the light of data availability in applications and subjectivity. A detailed mathematical discussion and some necessary proofs to support the statements made in Section 14.2 of our study will be available in the paper (Oladyshkin and Nowak, 2012a).

14.2. Data-driven analysis

14.2.1. A short review of polynomial chaos expansions

A large number of stochastic simulations tools (e.g. (Ghanem and Spanos, 1993), (Matthies and Keese., 2005), (Wan and Karniadakis, 2006), (Wan and Karniadakis, 2009), (Xiu and Karniadakis, 2003), (Xiu and Karniadakis, 2002a)) for diverse applications are based on the polynomial chaos expansion (PCE) first introduced by Wiener (Wiener, 1938) in 1938. The chaos expansion offers an efficient and accurate high-order way of including non-linear effects in stochastic analyses. In simple words, the dependency of model output on input parameters is approximated by a high-dimensional polynomial. This is achieved by projecting the model response surface onto a basis of polynomials which is orthogonal in the probabilistic parameter space (e.g. (Li and Zhang, 2007)). The benefits and limitations of polynomial chaos and related expansion techniques are discussed in (Augustin et al., 2008).

PCE techniques can mainly be sub-divided into intrusive and non-intrusive approaches for the projection integrals involved. Intrusive approaches require manipulation of the govern-

ing equations and can sometimes provide semi-analytical solutions for stochastic analyses of simple problems. The best-known method from this group is the stochastic Galerkin technique, which has also been applied in studies for modeling uncertainties in flow problems (e.g. (Ghanem and Spanos, 1993), (Matthies and Keese., 2005), (Xiu and Karniadakis, 2003)). However, the necessary symbolic manipulations may become very complex and analytically cumbersome, and cannot easily be implemented in commercial codes. For this reason, non-intrusive approaches like sparse quadrature (Keese and Matthies, 2003) and the probabilistic collocation method (PCM) (Isukapalli et al., 1998; Li and Zhang, 2007) have been receiving quickly increasing attention. In a simple sense, PCM can be interpreted as a smart (mathematically optimal) interpolation rule of model output between different parameter sets. The polynomial interpolation may be interpreted as a response surface of the model. It is based on a minimal and optimally chosen set of model evaluations, each with a defined set of model parameters (called collocation points). Reviews of the mathematical theory behind intrusive and non-intrusive PCE approaches are presented in (Jakeman and Roberts, 2008). Uncertainty characterization can also be extended to robust design tasks in uncertain systems, when including design variables in the PCE approach (Oladyshkin et al., 2011b).

The classical polynomial chaos is based on Hermite polynomials, which are optimal for normally distributed random variables (Wiener, 1938), (Cameron and Martin, 1947). Unfortunately, natural phenomena are not often that simple, and the distribution of physical parameters often cannot be considered Gaussian. Simple examples of non-Gaussian phenomena are the non-negativity of reservoir permeability or the boundedness of porosity. In such cases, it is possible to put into conformity the physical variable with a normal variable by an adequate transformation called Gaussian anamorphosis (Wackernagel, 1998). The anamorphosis function can be constructed through cumulative distribution functions and is bijective, i.e. provides one-to-one correspondence between both spaces. However, using transformed (normal) variables in the expansion cannot be considered an optimal choice because it leads to slow convergence of the expansion (e.g. (Xiu and Karniadakis, 2002a), (Xiu and Karniadakis, 2003)).

In recent years, the classical PCE technique was extended to the generalized polynomial chaos (gPC) (Wan and Karniadakis, 2006), which can directly use a counted number of theoretical statistical distributions, such as Gamma, Beta, Uniform, etc. (see the Askey scheme in (Askey and Wilson, 1985)). However, application tasks demand further adaptation of the chaos expansion technique to a larger spectrum of distributions. In (Wan and Karniadakis, 2006), the authors presented a multi-element generalized polynomial chaos

(ME-gPC) method, decomposing the random space into local elements, and subsequently implementing gPC locally within the individual elements. The ME-gPC is the first adaptive piece-wise approach helping to deal with the discontinuity of probability distributions and providing the desired adaptation to a wide spectrum of distributions. Up to the present, all implementations of PCE require the random variables to be statistically independent. However, PCE also can be applied to correlated variables, if correlation can be removed (or minimized) by adequate linear or nonlinear transformation.

14.2.2. Polynomial representation of models

This Section deals with the representation of models by polynomials, which is the basis for uncertainty propagation from input parameters to model output (prediction) within all PCE techniques. The case of CO_2 storage serves as an example throughout this chapter. Here, the significant uncertain input parameters are porosity, permeability, etc. (Kopp et al., 2009). This uncertainty has a non-trivial influence on the model output. In the case of CO_2 storage, the model output may, for example correspond to CO_2 saturations, pressures in the reservoir, the total CO_2 leakage to the surface, or the amount of displaced saltwater (brine) contaminating fresh-water aquifers above.

In order to quantify this influence, we perform non-linear uncertainty propagation from model parameters to the predicted quantities of interest. To do so, we project the output of an adequately chosen simulation model onto a higher-order orthogonal basis of polynomials via the probabilistic collocation method (PCM) (Li and Zhang, 2007). Let us introduce a formal model $\Omega = f(\omega)$. We investigate the influence of uncertain input parameters ω on the model output Ω. The total number of input parameters is equal to N, and ω can be considered a vector $\omega = \{\omega_1, ..., \omega_N\}$. According to polynomial chaos theory (Wiener, 1938), the model output Ω can be represented by the polynomial expansion (PCE):

$$\Omega(\omega) = \sum_{i=1}^{\infty} c_i \Psi_i(\omega). \tag{14.1}$$

Here, the coefficients c_i quantify the dependence of the model output Ω on the input parameters ω. Because the system we consider is space-distributed and time-dependent, the model output should be written $\Omega(X, \omega)$, where the vector $X = \{x, y, z, t\}$ consists of three space coordinates and time. This means that the coefficients c_j in expansion (14.1) have to be determined for each desired point X_k in space and time, i.e. $c_j(X_k)$.

In practice, only a finite number M of terms in the above expansion is considered. The function Ψ_i is a simplified notation of the multi-variate orthogonal polynomial basis for ω (see Section 14.2.3). The number M of terms in expansion (14.1) depends on the total number of input parameters N and the order d of the expansion, according to the combinatory formula $M = (N+d)!/(N!d!)$.

The remaining task of PCE methods consists in evaluating the as yet unknown coefficients c_j in expansion (14.1). In this chapter, we use the non-intrusive probabilistic collocation method (Oladyshkin et al., 2010; Li and Zhang, 2007), which requires evaluating the model Ω with M different sets of parameters ω that are called collocation points.

According to Villadsen and Michelsen (Villadsen and Michelsen, 1978), the optimal choice of collocation points corresponds to the roots of the polynomial of one degree higher $(d+1)$ than the order used in the chaos expansion (d).

In the current chapter, we present a data-driven polynomial basis (Section 14.2.3), which also defines data-driven positions of the collocation points for PCM. Our concept to avoid subjectivity, however, is not at all limited to PCM, but will apply to all PCE-related methods and other methods that work with polynomial expansion.

14.2.3. Data-driven polynomial basis

For simplicity, we will consider only one random variable ω_j from the vector $\omega = \{\omega_1, ..., \omega_N\}$ in the following derivation. In this section, we will show how to construct the data-driven orthogonal polynomial basis for each individual component ω_j. The multi-dimensional basis for $\omega = \{\omega_1, ..., \omega_N\}$ is obtained from the individual univariate bases by simple multiplication of all bases $\left\{P_j^{(0)}, ..., P_j^{(d)}\right\}$, $j = 1, ..N$.

The basis of degree d for parameter ω_j is defined as $\left\{P_j^{(0)}, ..., P_j^{(d)}\right\}$. The polynomial $P_j^{(k)}(\omega_j)$ of degree k in the variable ω_j is defined as:

$$P_j^{(k)}(\omega_j) = \sum_{i=0}^{k} p_{i,j}^{(k)} \omega_j^i, \quad k = \overline{0, d}, \quad j = \overline{0, N}. \tag{14.2}$$

Here, $p_{i,j}^{(k)}$ are the coefficients in $P_j^{(k)}(\omega_j)$. The main idea of our data-driven approach consists of constructing the coefficients $p_{i,j}^{(k)}$ in such a way that the polynomials in equation (14.2) form a basis that is orthogonal in arbitrarily given distributions of data. The arbitrary distributions can be either discrete, continuous, or discretized continuous and can be

specified either through some statistical moments, analytically as PDF/CDF, numerically as histograms, or theoretically through the even more general format of probability measures. Most importantly, the distribution can be specified directly via raw data, without any need for PDF estimation in intermediate steps.

We briefly summarize the theory of the arbitrary polynomial chaos from our own point of view. An orthogonal polynomial basis up to order d can be constructively defined for any arbitrary probability measure, given that ω_j has finite moments up to order $2d - 1$. Starting within the definition of orthogonality and some matrix algebra, it can be shown that the unknown polynomial coefficients $p_{i,j}^{(k)}$ can be defined from the following matrix equation:

$$
\begin{bmatrix}
\mu_{0,j} & \mu_{1,j} & \cdots & \mu_{k,j} \\
\mu_{1,j} & \mu_{2,j} & \cdots & \mu_{k+1,j} \\
\cdots & \cdots & \cdots & \cdots \\
\mu_{k-1,j} & \mu_{k,j} & \cdots & \mu_{2k-1,j} \\
0 & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
p_{0,j}^{(k)} \\
p_{1,j}^{(k)} \\
\cdots \\
p_{k-1,j}^{(k)} \\
p_{k,j}^{(k)}
\end{bmatrix}
=
\begin{bmatrix}
0 \\
0 \\
\cdots \\
0 \\
1
\end{bmatrix} .
\tag{14.3}
$$

Here the $\mu_{i,j}$ are the non-central (raw) statistical moments of order i for random variable ω_j. Written in this form, it becomes evident that moments are the only required form of information on input distributions for constructing the basis and the collocation points.

If the moments of ω_j are evaluated directly from a data set of limited size, there need to be k or more distinct values in the data set. The respective proofs are provided in the paper (Oladyshkin and Nowak, 2012a). All moments are always finite if no element of the data set is infinite. An analytical explicit form of the coefficients $p_{i,j}^{(k)}$ for moderate degrees of polynomials and more detailed properties of the basis $\left\{ P_j^{(0)}, \ldots, P_j^{(d)} \right\}$ are also given in (Oladyshkin and Nowak, 2012a). We would like to remark that any PDF contains stronger information in comparison to raw data or other weaker probability measures. As a consequence, the availability of a PDF is a sufficient condition for the proposed framework; however, the condition that a finite number of moments exist is still necessary.

The orthogonal polynomial basis above can be used directly for analysis. However, a normalized basis has further useful properties. For example, the mean and variance of $\Omega(\omega)$ according to the expansion (14.1) is given by simple analytical relations (see equation 14.6), thanks to the orthonormality property. The orthonormal polynomial basis can be obtained as:

$$
\widehat{P}_j^{(k)}(\omega_j) = P_j^{(k)} / \left\| P_j^{(k)} \right\| ,
\tag{14.4}
$$

where the norm of the polynomial $\left\|P_j^k\right\|$ for space of events Λ (where $\omega_j \in \Lambda$) with probability measure Γ is defined as:

$$\left\|P_j^{(k)}\right\|^2 = \int_{\omega_j \in \Lambda} \left[P_j^{(k)}(\xi)\right]^2 d\Gamma(\omega_j). \tag{14.5}$$

The normalization by $\left\|P_j^{(k)}\right\|$ for $k = d$ additionally requires the finiteness of the $2d$-th moment.

14.2.4. Evaluation of statistics

In our approach, the statistics of the model output are based directly on the model and the specified moments of input data. If a model output $\Omega(\omega)$ is expanded in the normalized polynomial basis (14.4), then characteristic statistical quantities of $\Omega(\omega)$ can be evaluated directly. For example, the mean and variance of $\Omega(\omega)$ are given by the following simple analytical relations:

$$\texttt{mean}(\Omega) = c_1, \quad \texttt{var}(\Omega) = \sum_{j=2}^{N} c_j^2, \tag{14.6}$$

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

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

As an alternative to the analytical relations above, arbitrary output statistics can be evaluated via Monte Carlo analysis of the polynomial response surface that results from the expansion (14.1). This is very fast, because the polynomial surface is much faster to evaluate than the original model equations. However, the latter approach will ask for full knowledge of the probability density function involved, which we try to avoid. If Monte Carlo cannot be avoided, e.g. for computing exceedance probability in risk analysis, we suggest using maximum entropy PDFs, or alternatively, a sufficiently large data set (if available) that can be used directly as Monte-Carlo realizations of input parameters.

14.2.5. Small data sets and robustness

When the input data set is small, the sample moments are only uncertain estimates of real moments. Hence, a direct application of the method presented becomes less robust. In that case, it would be useful to apply some standard methods to achieve robustness in the estimation of moments, such as bootstrapping (e.g. (Efron, 1987)). Moreover, especially in such cases, expert opinion can be very useful. In our proposed approach, an expert will have total freedom of data interpretation (not restricted to the selection among standard PDFs) and can provide much more sophisticated information (e.g. lower and higher moments, etc.). According to our approach, expert opinion (in a most general sense) will be incorporated directly without any additional transformation or additional subjectivity when translating it to the stochastic numerical framework.

14.3. Data-driven uncertainty quantification for CO$_2$ storage

14.3.1. Statistical problem formulation

To illustrate our proposed methodology, we will consider three parameters uncertain: reservoir absolute permeability, reservoir porosity and permeability of the leaky well. Figure 14.1
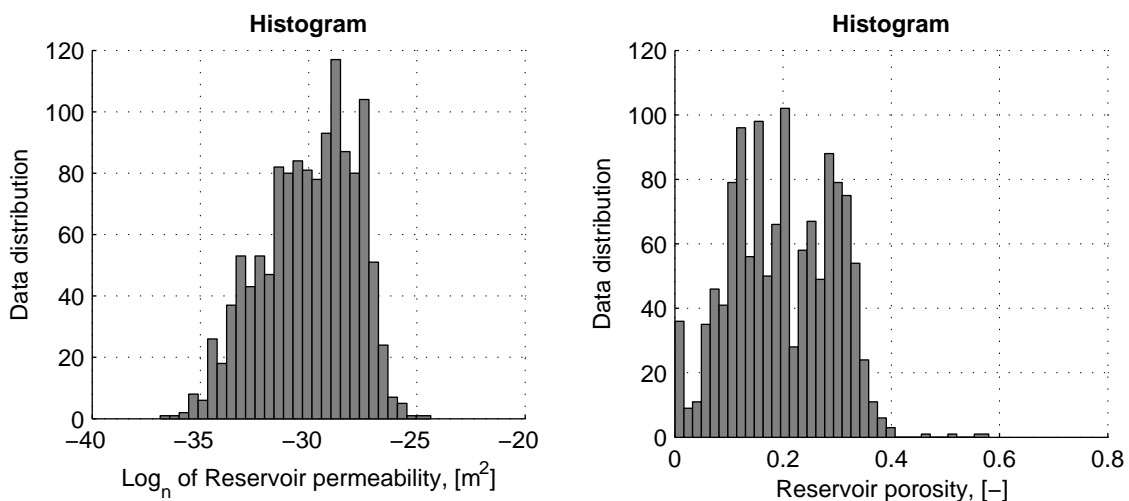

Figure 14.1.: Distributions of reservoir data taken from U.S. National Petroleum Council Public Database

demonstrates the distributions of the material properties which we use in our illustration. Here, the distributions of absolute permeability and porosity were taken from the U.S. National Petroleum Council Public Database (which includes 1270 reservoirs), see also (Kopp et al., 2009). This choice reflects the situation of site screening, where site-specific data and data that allow heterogeneity within geological units to be resolved are not yet available. Instead, we use data sets that represent macroscopic properties of supposedly similar sites as prior knowledge.

Unfortunately, such information is not available for all sources of uncertainty, like in our case the permeability of the leaky well. In such cases, additional assumptions become necessary. The classical way would be to introduce a theoretical (parametric) probability distribution, e.g. with an assumed mean and variance. Here, we could introduce as expert opinion the lognormal distribution for the leaky-well permeability, with parameters defined from the benchmark values (Class et al., 2009), see Figure 14.2. As discussed in the introduction, establishing a full theoretical probability density function involves a strong assumption on all higher moments up to infinite order, and assumes implicit knowledge of the exact shape, e.g. also of the extreme-value tails. Our approach strongly alleviates this situation, because it can handle a set of moments directly (e.g. the mean, variance, skewness, peakedness), without any further assumptions on higher-order moments and without having to introduce a PDF. We will use this freedom, and obtain only a small number of required moments directly from a large database or via experts, without asking for a full PDF. For evaluating arbitrary complex output statistics without having a sufficiently large raw data set, we would recommend the Maximum Entropy PDF also shown in Figure 14.2 to draw a sufficiently large sample for Monte Carlo analysis of the polynomial.

In the test case considered, we apply a second-order expansion, which requiring knowing the first four moments. Figure 14.2 illustrates possible distributions of stochastic variables, where the solid line represents a theoretical PDF and the dashed lines represent a small collection of alternative maximum entropy PDFs with the same first four moments. Our method does not require a choice between these alternatives, but only uses common information in the form of the required moments (here: up to order four). The discussion in Section 14.2.3 shows that the existence of finite moments (up to a certain required order) is a necessary and sufficient condition for the proposed framework. Usually, this condition is easily fulfilled for a large spectrum of practical applications, especially for moderate degrees of expansion.

Formally, the uncertain parameters presented in Figures 14.1 and 14.2 correspond to the components of the vector $\omega = \{\omega_1, \omega_2, \omega_3\}$ and represent the input parameters of Equation
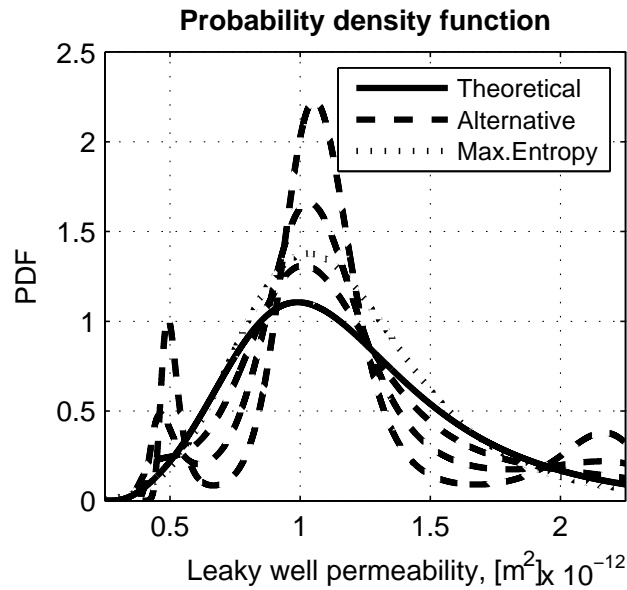
Figure 14.2.: Assumed stochastic distribution: theoretical PDF (solid line), some alternative PDFs (dashed lines) with the same first four moments and Maximum Entropy PDF with the same first four moments (dotted line).

(14.1). The model output quantities Ω considered here are pressure and saturation values as a function of space and time, and the CO$_2$ leakage rate through the leaky well as a function only of time.

14.3.2. Application and test of the data-driven approach

The procedure for quantifing uncertainty in CO$_2$ storage can be divided into three main steps. The first step is to construct the polynomial basis (see Section 14.2.3) according to the data considered in Section 14.3.1. The second step is to set up the chaos expansion and obtain the required coefficients c_i using the non-intrusive probabilistic collocation method summarized in Section 14.2.2. In the third step, we evaluate all desired output statistics according to Section 14.2.4.

In the current work, we also apply the conventional PCE approach for direct comparison with the data-driven approach. The principal difference lies in the first step, i.e. in constructing the polynomial basis. The conventional approach is based on Hermite polynomials, which are optimal for Gaussian random variables. Because the random variables ω are not distributed Gaussian in compliance with the Hermite polynomial basis, an additional conversion is required. A large number of methods are based on transformations of the model variables ω_{Ph}

from physical space (shown in Figure 14.1) to corresponding normal variables ω_N via Gaussian anamorphosis (Wackernagel, 1998). This would implicitly define an exact PDF, which once again is an unjustifiably strong assumption. In contrast, the data-driven approach is based directly on the considered moments of the distributions or raw data of the uncertain input variables ω (see Section 14.2) in physical space.

The numerical part of our case study focuses on the validation of the proposed approach by comparison with Monte Carlo simulation. Due to the high computational demand of Monte Carlo reference simulations, we consider only a small version of the original benchmark problem, where the simulation time is limited to 30 days, and the numerical grid is coarse (1183 nodes). This simplification is imposed only by the expensive Monte Carlo approach, while PCE methods can easily handle much larger problems.

14.3.3. Computational efficiency and convergence

The data-driven polynomial chaos expansion presented in Section 14.2 provides a simple but powerful tool for stochastic modeling and, in this case study, for the probabilistic risk assessment of CO$_2$ storage. Before discussing the mentioned issues of subjectivity in Section 14.3.4, we will first demonstrate the computational efficiency of the data-driven polynomial basis, and we will show the improved convergence compared with the conventional polynomial chaos approach.
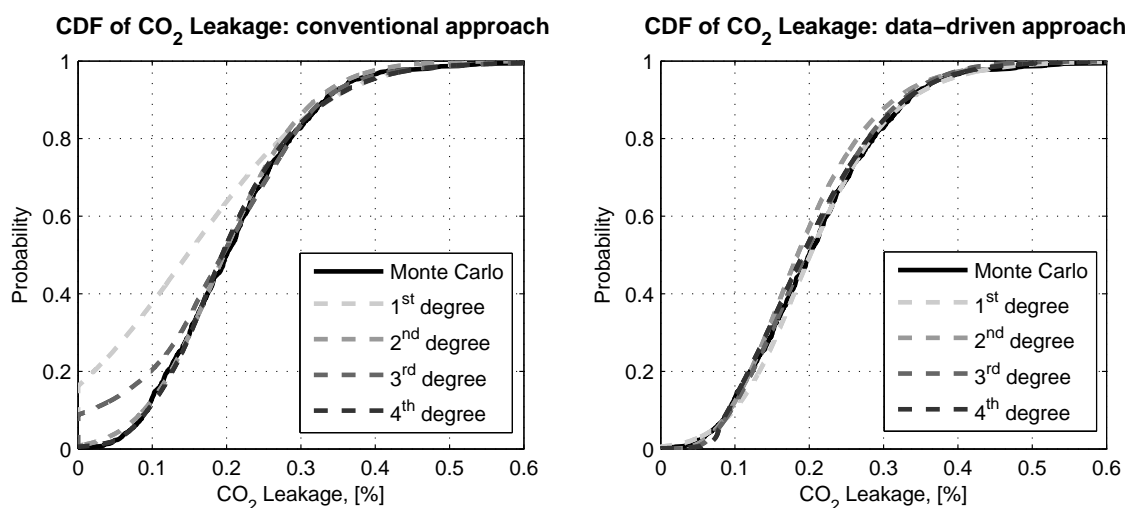

Figure 14.3.: Cumulative distribution function of CO$_2$ leakage rate after 30 days: conventional approach (left plot) and data-driven (right plot)

The most integrative characteristic of the overall benchmark problem is the total leakage of CO$_2$. To compare a quantitative characteristic that is most important in probabilistic risk assessment, we computed the cumulative probability function of the CO$_2$ leakage rate after 30 days (Figure 14.3). The cumulative density function represents the probability that the CO$_2$ leakage is less than or equal to a particular value. The left plot in Figure 14.3 corresponds to the conventional approach, i.e. involving Gaussian anamorphosis and Hermite polynomials. The right plot in Figure 14.3 corresponds to our data-driven approach without additional transformation. The convergence of both types of chaos expansions (dashed lines) was validated by traditional Monte Carlo simulations (solid lines) with 1270 realizations, where we sampled directly from the available data sets. We repeated the comparison study for different degrees of the chaos expansion, such as first order (4 evaluations), second order (10 evaluations), third order (20 evaluations) and fourth order (35 evaluations).

Figure 14.3 demonstrates that the data-driven approach provides fast convergence. Even small degrees of the data-driven expansion (even the linear one in our specific case study) ensure adequate representation of the stochastic processes in the considered multiphase flow system. This fact can be very useful for fully resolved and complex real-world application challenges, where computational costs are very high even for a single model evaluation.

Convergence with the conventional approach (here: based on Hermite polynomials) strongly depends on the non-linearity of the required transformation from ω_{Ph} to ω_N (Oladyshkin and Nowak, 2012a). There are two sources of slow convergence (or errors) for chaos expansions based on non-data-driven polynomial bases. First, the transformation from physical space to normal space introduces additional non-linear behavior into the overall problem, which will require a higher order of expansion to obtain comparable accuracy. Second, for non-intrusive methods that rely on numerical integration to obtain the coefficients c_i, the accuracy of numerical integration strongly depends on the choice of integration or collocation points. For example, in Gauss-Hermite integration (Abramowitz and Stegun, 1965), the polynomial basis of degree d defines the positions of integration points by the roots of the polynomial of degree $d + 1$. These integration points are optimal only if the weighting function (here: the probability measure) and the polynomial basis are in direct correspondence. Any non-linear transformation from ω_{Ph} to ω_N destroys this direct relation. Thus, using a non-data-driven polynomial basis leads to a non-optimal placement of integration points, which causes a reduced accuracy of the integration (see (Oladyshkin and Nowak, 2012a)). As a consequence, in comparison with the classical PCE, the same order of precision can be achieved with a smaller degree of expansion. From the practical point of view, the computational costs of our data-driven chaos expansion combined with PCM are proportional to the

number of terms in the chaos expansion multiplied by the time of a single model evaluation. The number of terms is $M = (N+d)!/(N!d!)$, where N is number of model parameters and d is the degree of expansion. The number M quickly increases with the degree d of expansion and the N number of parameters, but the framework is very efficient for low-parametric systems. In order to translate these advantages to heterogeneous systems, a combination with the KL-expansion (e.g. (Li and Zhang, 2007)) can be very useful. Moreover, an initial sensitivity analysis can be applied to keep the number of analyzed parameters and their respective degree of expansion at a low and efficient level.

14.3.4. Impact of subjectivity

To illustrate the drastic impact of subjectivity that can be introduced into analysis, we perform the following simple experiment. The sample data (reservoir permeability and porosity) presented in Figure 14.1 are distributed to five different and independent experts. The task of each expert consisted of constructing a theoretical probability density function for each parameter, which in their opinion would describe the statistics of the raw data best. As a common way of data description, all experts provided two-parametrical distributions for each input parameter (permeability and porosity), which are presented in Table 14.1. Table 14.1 demonstrates that all experts proposed very different assumptions and techniques to match the permeability and porosity distributions. The responses of all experts were collected and used as input for modeling the benchmark problem defined in Sections 5.2 and 14.3.1.

The results of this experiment are illustrated in Figure 14.4 using second-order polynomial

Number	Assumption on permeability	Assumption on porosity
Expert 1	Manual fit of Log-Normal distribution to first two moments	Manual fit of Beta distribution to first two moments
Expert 2	Maximum likelihood fitting for Beta distribution	Maximum likelihood fitting for Normal distribution
Expert 3	Log transformation and visual fitting by Normal in distribution	Maximum likelihood fitting Beta distribution
Expert 4	Visual fitting by BoxCox transformation of Log-Normal distribution	Maximum likelihood fitting for Log-normal distribution
Expert 5	Visual fitting of Log-Normal distribution	Visual fitting of Log-Normal distribution

Table 14.1.: Expert opinions for the distributions of input data.
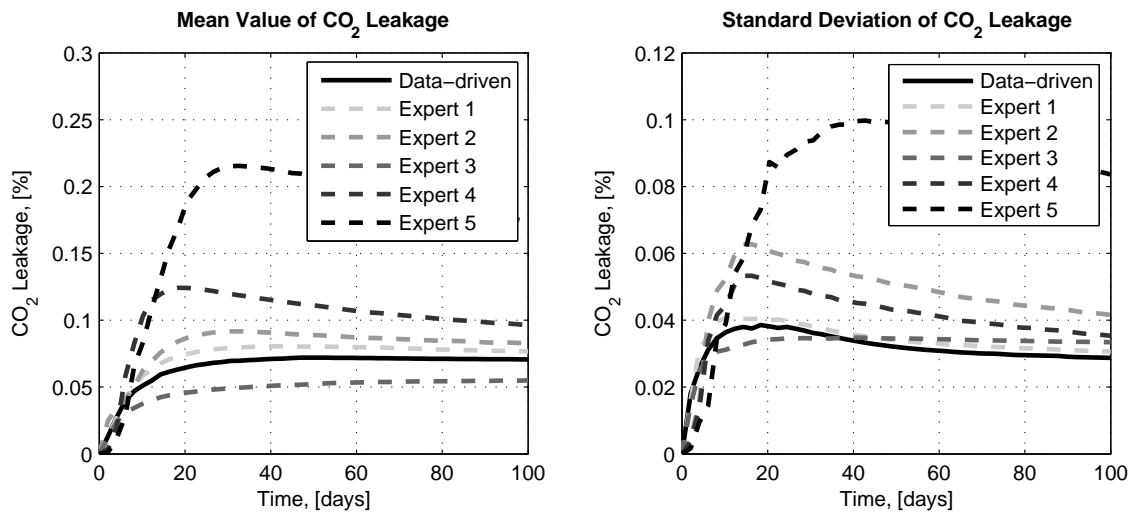
Figure 14.4.: Estimation of mean value (left plot) and standard deviation (right plot) of the CO$_2$ leakage rate: expert opinions (dashed lines) and data-driven approach (solid lines)

chaos expansion. It shows the resulting mean value (left plot) and standard deviation (right plot) of the CO$_2$ leakage rate over time. Here, dashed lines correspond to the results based on subjective expert opinions, and solid lines correspond to the results from our purely data-driven approach. All test cases presented in Figure 14.4 were performed under the same conditions in all other aspects. The differences in predicting leakage rates based on the different experts are a consequence of their different interpretations and opinions on how to treat the data. This example clearly demonstrates that the room for subjectivity (when assigning theoretical stochastic distributions to real data and thus modifying their higher-order moments) can lead to significant differences in the predicted values.

In this specific case, individual expert subjectivity even results in a systematic bias on average for all experts. The reason is that most of the experts assigned a log-normal distribution to reservoir permeability and a normal distribution to porosity, which is a systematic overestimation of higher-order moments compared with the raw data. As one can easily see from Figure 14.4, one of the experts provides a very good match to the leakage values obtained directed from the raw data. This expert's PDFs have a very precise approximation of the first four moments of the data set, which is sufficient for the second-order expansion in the example under consideration to produce identical results.

The data-driven polynomial chaos expansion we propose here for stochastic analysis is based directly on the moments of sampled data without intermediate steps of data re-interpretation.

This avoids the subjectivity usually introduced when choosing among a small limited number of theoretical distributions to represent a natural phenomenon, and so avoids the problem illustrated in Figure 14.4.

14.4. Conclusions

In this work, we presented a data-driven approach for stochastic modeling and uncertainty quantification, based on polynomial chaos expansion (PCE), and applied it to CO$_2$ storage in deep geological formations. PCE allows a non-linear propagation of parameter uncertainty onto the predicted quantities. The data-driven approach provides a response surface based on a global orthonormal polynomial basis for arbitrary distributions. The method only demands the existence of a finite number of moments, and does not require the exact knowledge or even existence of probability density functions. The arbitrary distributions can be either discrete, continuous, or discretized continuous and can be specified either through a few statistical moments, analytically as PDF/CDF, numerically as a histogram, or theoretically through the even more general format of a probability measure.

Thus, the data-driven approach provides freedom for modeling even physical systems with unknown probability distribution functions, when only data sets of very limited size or only little prior knowledge is available. It offers a new approach to defining parameter uncertainty in stochastic analysis that avoids the subjectivity of assigning theoretical probability distributions. Such theoretical distributions are often not sufficiently supported by the available data and thus can introduce assumptions on higher-order moments (and extreme-value tails) that are only based on the chosen distribution. Especially non-linear models are highly sensitive to such subjectively chosen higher-order moments or shapes of distribution tails. An interesting aspect is that only moments up to twice the order of expansion matter. Therefore, any PDFs fitted to input data will lead to the same statistical moments for model output, if the PDFs coincide in the required moments up to twice the order of expansion. For the same reason, if one still desires to fit a single PDF to the input data set, we recommend maximum entropy or minimum relative entropy methods applied such that the moments relevant for PCE are matched. This differs drastically from fitting low-parametric distributions to only lower moments of the available data, because this would modify the remaining relevant moments up to twice the order of expansion.

In a small fictitious example, we asked independent experts to select and fit parametric distributions to two raw data sets. The example demonstrated that subjectivity in data interpreta-

tion can easily lead to prediction bias. With this example, we also illustrated that subjectivity can be a source of error that plays in the same league as physical conceptualization, numerical code development, and stochastic modeling. Modeling is a chain of many tasks. Like any chain, it is only as strong as its weakest link. Our results indicate clearly that the statistical treatment of input data is part of the chain, and that the subjectivity in assuming theoretical curves can weaken that link immensely.

Our test case illustrated the method for a law-parametric homogenized system. This is most useful for site-screening procedures. When moving towards a more refined model with heterogeneity, our approach can easily be adapted by the often invoked combination with the KL-expansion.

We use the "black-box" collocation technique to project the output of a simulated model ("response surface") onto a higher-order orthogonal basis of polynomials, which then can be used for a wide range of statistical modeling tasks. The data-driven stochastic approach was validated on the basis of Monte Carlo simulation using a common 3D benchmark problem. The proposed approach yields a significant computational speed-up compared with Monte Carlo, and provides faster convergence than conventional polynomial chaos expansions. Even for small degrees of expansion, the data-driven expansion can be very accurate, which can save a lot of computational power for probabilistic risk analysis. The available computational resources can thus be used towards large-scale, more finely resolved and more complex simulation models in real-world application challenges. Our data-driven version of PCE can also be applied for robust design problems under uncertainty (Oladyshkin et al., 2011b).

In summary, the chaos expansion approach presented here provides an effective tool for a probabilistic risk assessment of CO_2 storage and other problems that suffer from data sparsity. Our method is based directly on raw data or other arbitrary sources of information without auxiliary assumptions. This increases the efficiency of chaos expansion and minimizes subjectivity, providing valuable support for risk-informed decision making as well as for robust design and control, allowing a better assessment and reduction of risk.

15. Risk estimation of brine migration resulting from CO_2 injection

Comprehensive risk assessment is a major task for large-scale projects such as geological storage of CO_2. Basic hazards are damage to the integrity of caprocks, leakage of CO_2, or reduction of groundwater quality due to intrusion of fluids. A particular focus of this study is on salinisation of freshwater aquifers resulting from displaced brine. Quantifying risk on the basis of numerical simulations requires consideration of different kinds of uncertainties associated with the applied models. This study distinguishes between different levels of uncertainty, among which scenario uncertainty and statistical uncertainty are particularly accentuated. Addressing scenario uncertainty involves expert opinion on relevant geological features such as caprock properties, faults, distinct geological layers, etc. This is considered in this work by 6 different scenarios having different characteristic geological features. On the other hand, Monte Carlo methods are a classical approach to address statistical uncertainty. Since this is not feasible for large-scale 3D models including complex physics, a model reduction on the basis of the Integrative Probabilistic Collocation Method is proposed. It is shown that, dependent on the data availability, both levels of uncertainty can be equally significant. The presented study gives estimates for the risk of brine discharge into freshwater aquifers due to CO_2 injection into geological formations and resultant salt concentrations in the overlying drinking water aquifers.

15.1. Introduction

15.1.1. CO$_2$ storage and associated risks

Carbon capture and geological storage (CCS) is currently being discussed as one possible technology for reducing emissions of the greenhouse gas CO$_2$ (carbon dioxide) into the atmosphere. CCS comprises three main parts: (i) the capture of CO$_2$o, e.g. from fossil- or biomass-fueled power plants, (ii) the transportation of the CO$_2$ to a suitable storage site, and (iii) the safe storage in deep geological formations. Possible geological storage sites are depleted oil or gas reservoirs, coal beds, and saline aquifers.

This study focuses on CO$_2$ storage in saline aquifers because these formations have the highest storage potential worldwide ((IPCC, 2005)). Various trapping mechanisms in the porous rock are responsible for storing the CO$_2$ safely: structural trapping by a closed caprock, residual trapping in the porous rock, solubility trapping in the water phase, and geochemical trapping by reactions with the rock matrix ((IPCC, 2005)).

To assess the safety of CO$_2$ storage it is important to quantify the risks to health, safety, and the environment. The basic hazards that need to be considered for CO$_2$ storage in saline aquifers are CO$_2$ leakage, brine displacement and infiltration, and structural failure. If CO$_2$ leaks from the reservoir, it could rise to shallower aquifers and dissolve partly or fully in freshwater (drinking-water resources) or leak back to the atmosphere. Structural failure could occur due to high pressure peaks at the injection well and result in new leakage pathways through the caprock. It is highly likely that the injected CO$_2$ will displace brine initially. Injected CO$_2$ (partly) dissolves in the brine. Non-dissolved gaseous (supercritical) CO$_2$ requires pore space. The pore space is gained by compression of the fluids and the rock and by displacement of the resident brine. The displaced brine could move into shallower aquifers through fractures, faults, or abandoned wells and subsequently infiltrate and pollute freshwater aquifers with salt and other contaminants. Obtaining estimates for damage events and their likelihood is of major interest to water suppliers, and this study will therefore focus on the salt transport and potential damage to freshwater resources.

15.1.2. Risk assessment for geological CO_2 storage - State of research

A few studies have already presented systematic risk assessments for realistic CO_2 storage scenarios. (Benson et al., 2002) and (Lewicki et al., 2007) discussed the lessons which can be learned from natural or industrial analogues for CO_2 storage. (Lewicki et al., 2007) identified the features, events, and processes from these analogues and compared them with CO_2 storage. (Maul et al., 2007) discussed similarities and differences between radioactive waste disposal and CO_2 storage for performance assessment. (Oldenburg, 2008) and (Oldenburg et al., 2009) developed a screening and ranking method and a certification framework, respectively, for selecting suitable storage sites on the basis of health, safety, and environmental risk resulting from CO_2 or brine leakage. (Kopp et al., 2010) developed a method for assessing a CO_2 leakage scenario, where the most relevant parameters and dependencies of all other parameters were used to estimate the risk, thereby considering their range of variability. (Siirila et al., 2012) presented a quantitative methodology to assess the risks to human health, that can occur if CO_2 leaks into groundwater.

It is important to consider the pressure evolution during CO_2 storage since high pressure peaks at the injection well can damage the caprock. Moreover, the far-field pressure build-up is an indicator of brine displacement and possible brine infiltration into water aquifers. (Birkholzer et al., 2009) and (Schäfer et al., 2011) investigated the pressure build-up resulting from CO_2 injection and concluded that it is very important to consider the far-field pressure evolution, especially for brine displacement. (Lemieux, 2011) summarizes the current scientific knowledge of the potential impact of CO_2 storage in saline aquifers on groundwater resources.

15.1.3. Uncertainties

Risk assessment requires dealing with uncertainties. In a broader sense, uncertainty can be defined as the absence of information or knowledge ((Rowe, 1994)). In the field of uncertainty quantification, there is often distinguished between two categories, i.e. epistemic and aleatory uncertainty (e.g., (Helton, 1994), (Hoffman and Hammonds, 1994), (Wojtkiewicz et al., 2001)). Epistemic uncertainty is also denoted as reducible, subjective, or type B uncertainty and results from a lack of information about the system. Such an uncertainty can be reduced by further research or measurements. Aleatory uncertainty is also known as variability, irreducible, stochastic, or type A uncertainty and results from the fact that a system can
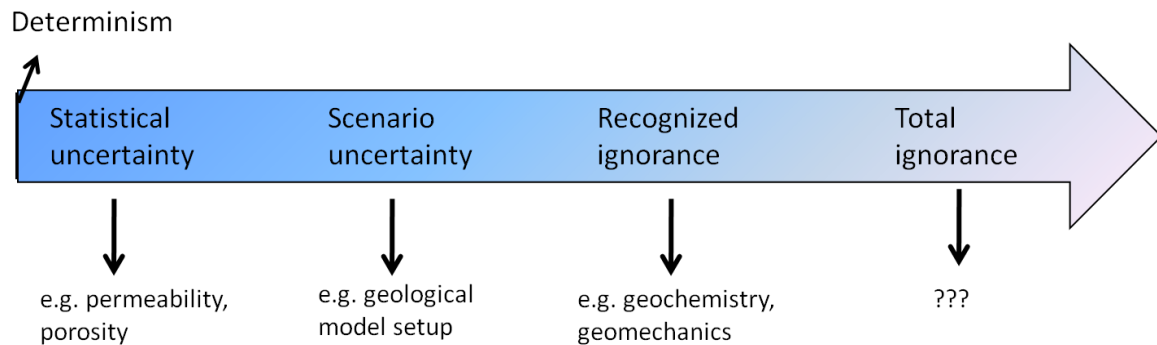
Figure 15.1.: Levels of uncertainty after (Walker et al., 2003).

react in many different unpredictable ways. Some authors additionally distinguish between dimensions of uncertainties (e.g., (Rowe, 1994) and (Walker et al., 2003)).

(Walker et al., 2003) divides uncertainties not only into different natures (epistemic and aleatory) but also into the location and the level of uncertainty. The location identifies where the uncertainty is situated within the model complex, whereas the levels categorise the uncertainty into ranges of knowledge. In this study, we follow the approach of (Walker et al., 2003) and categorize the uncertainties into levels. Levels of knowledge range from determinism to total ignorance (see Figure 15.1). Determinism means that everything is known exactly, which would be an ideal situation. Total ignorance is the other extreme, where nothing is known. The levels in between are statistical uncertainty, scenario uncertainty and recognized ignorance.

Statistical uncertainties can still be described by statistical terms (e.g., probability density function, or histogram). In the case of CO_2 storage, such uncertainties can be found, for example, in permeability, porosity, or anisotropy values. The next level is scenario uncertainty (or conceptual uncertainty). Such uncertainties, e.g. the geological layering considered in the model or a gap in the caprock cannot be described sufficiently by statistical terms. Adequate scenarios have to be chosen properly by considering the opinion of experts in the field of interest, e.g., geologists. Measurements and site exploration typically reduce uncertainty. For example, drilling an exploration well together with a seismic campaign might substantially improve the knowledge of number and orientation of distinct geological features, while still leaving much uncertainty of their specific properties. When considering uncertainty due to geological features, there can be a continuous transition between these two categories of uncertainty. The modeler must consider both the unknown natural conditions and interpretation of available expert opinion. Recognized ignorance is defined by (Walker et al., 2003)

as fundamental uncertainty about the mechanisms and functionalities. Assumptions made in the model such as neglecting compositional effects or thermal effects are examples for this level.

For a given system, all these levels of uncertainties can be found. All levels of uncertainty discussed in this paper, except for total ignorance, can be also classified into the commonly used concept of epistemic and aleatory uncertainty, (the nature dimension in (Walker et al., 2003)). The considered examples for uncertainties as permeability, the geological model set-up, or geochemistry are epistemic uncertainties since more measurements or more sophisticated models reduce uncertainty. However, in engineering practice there will always remain an aleatory uncertainty in parameters like permeability or porosity due to insufficient measurements.

15.1.4. Scope of this work

In this study, a comprehensive risk assessment methodology is developed and demonstrated for the estimation of the potential for brine displacement and brine infiltration into a drinking water aquifer. The proposed method can be transferred to other hazardous events such as CO$_2$ leakage and to other problems such as natural-gas storage or radioactive-waste disposal. The approach to divide uncertainties into different levels proposed in (Walker et al., 2003) is adapted to CO$_2$ storage and the uncertainties in the geological structure are investigated on two different levels: statistical and scenario uncertainty. We deal with both by performing numerical simulations of the CO$_2$ storage process.

Scenario uncertainty is investigated by simulating 6 scenarios with different geological setup and subsequent ranking according to the resulting damage. Accounting for statistical uncertainties requires further the testing of parameter spaces, for example with Monte Carlo simulations. Performing Monte Carlo simulations with complex multiphase flow models would incur huge computational costs. We reduce them with a massive stochastic model reduction via polynomial chaos expansion combined with the probabilistic collocation method, presented in (Oladyshkin et al., 2011b) and (Oladyshkin et al., 2011a). The method provides estimates of probabilities of certain hazardous events. By combining them with an estimation of damage, a quantification of risk can be obtained.

In order to quantify the impact of brine migration on a resource like potable groundwater, it is necessary to consider both brine migration from the reservoir and its mixing with freshwater

resources. Therefore, we discuss additional approximations for assessing the salt concentrations in the water aquifer and for investigating salt water upconing at water-production wells. The aim of these simple analytical approximations is to emphasize that a comprehensive risk assessment requires a set of consistent methods which allow for the consideration of various levels of uncertainties and operational strategies. The methods are linked to achieve an integral risk assessment methodology.

The chapter includes the following sections: Section 15.2 provides a description of the simulation model; Section 15.3 shows results for 6 different scenarios; and Section 15.5 characterizes the risk of brine breakthrough. The chapter concludes with discussions in Section 15.4.

15.2. Simulation model

15.2.1. Model set-up

The model set-up is based on the assumption that drinking water is pumped from a depth of 500 m. Such a deep pumping depth is considered because it is assumed that municipal authorities wish to protect all drinking water resources above a CO_2 reservoir and the deepest system would be those first polluted if leakage occurred. It is further assumed that the freshwater aquifer is located above the CO_2 storage reservoir, separated by 7 geological layers in between. The storage reservoir is located in a depth of 817.5 m. At this depth, pressure and temperature are high enough to assure supercritical conditions. For CO_2 storage it is important that the CO_2 is in supercritical state because it has a higher density than in gaseous or liquid state, which increases the storage capacity and safety.

The model domain is confined at the bottom by the storage reservoir, into which the CO_2 is injected, and at the top by the freshwater aquifer, with a caprock and low permeable layers (aquitards) in between. All layers, except for the freshwater aquifer, initially contain brine of equal constant salinity (0.1 kg salt/kg solution). Compared to the most probable situation of salinity increasing with depth, this is a conservative assumption. Figure 15.2 shows a schematic view of the domain for the reference scenario.

A radially symmetric 3D model domain is set up with an inner radius of 1 m representing the injection well. The domain has a height of 350 m and an outer radius of 10 km (see Figure 15.3). The simulated domain represents 1/12 (30°) of a circle, which means that

extensive values like brine discharge or injection rate need to be multiplied by 12 to account for the entire domain. The upper and lower boundaries as well as the segments' symmetric boundaries are modeled with closed no-flow boundary conditions. Closed boundaries on the top and the bottom of the reservoir lead to higher pressures than would be obtained with open boundaries. This assumption corresponds to the existence of aquitards of very low permeabilities at top and bottom. The outer boundary at 10 km distance is a hydraulically open constant-head boundary. This has also influence on the pressure in the domain since the prescribed injection rate causes primarily a pressure gradient. Thus, dependent on the size and the traveling speed of the pressure signal, a fixed head at the boundary can lead to an underestimation of pressure in the domain during injection. Therefore, to minimize errors induced by this boundary condition, the distance to the investigated locations should be large. The CO$_2$ injection period is 25 years at a rate of 1 Mt CO$_2$/a. The actual injection rate into the radially symmetric section corresponds to 1/12 Mt/a.

A two-phase (2p) model implemented in the simulator *DuMu^X* ((Flemisch et al., 2011)) including only the phases CO$_2$ and brine is applied, which means that the model does not distinguish between salt water (brine) and freshwater. This is considered justified if, like here, a primary goal is just to identify geological setups having high potentials for brine displacement. The output of interest is the brine discharge into the water aquifer. For the 2p
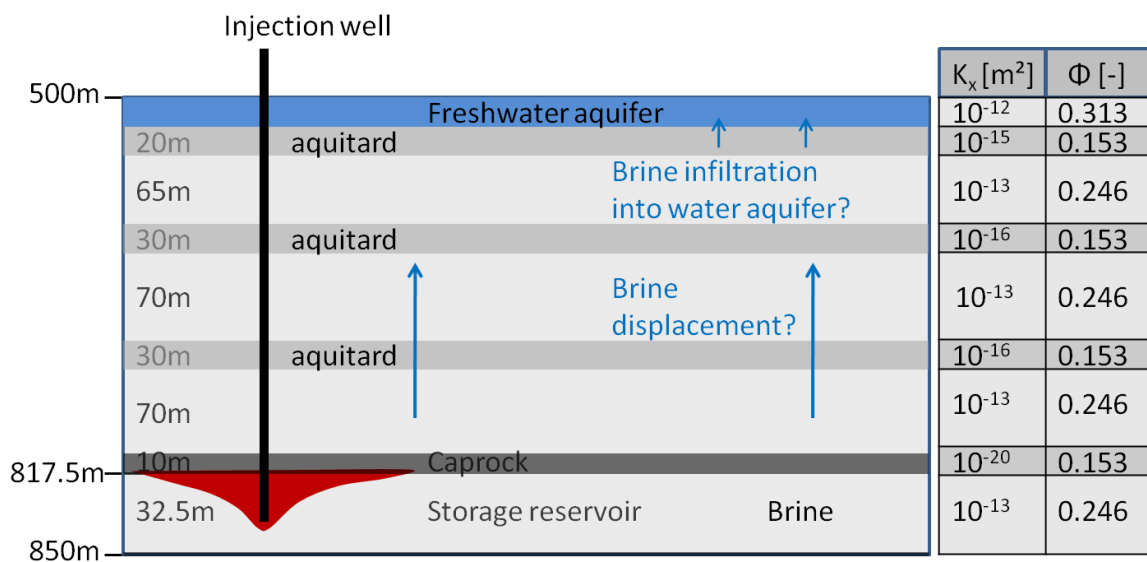
Figure 15.2.: Schematic cross-sectional view of the reference scenario (Walter et al., 2012) to estimate the risk of brine infiltration into a freshwater aquifer including permeability and porosity values for all layers ($K_z = 0.1 \cdot K_x$).
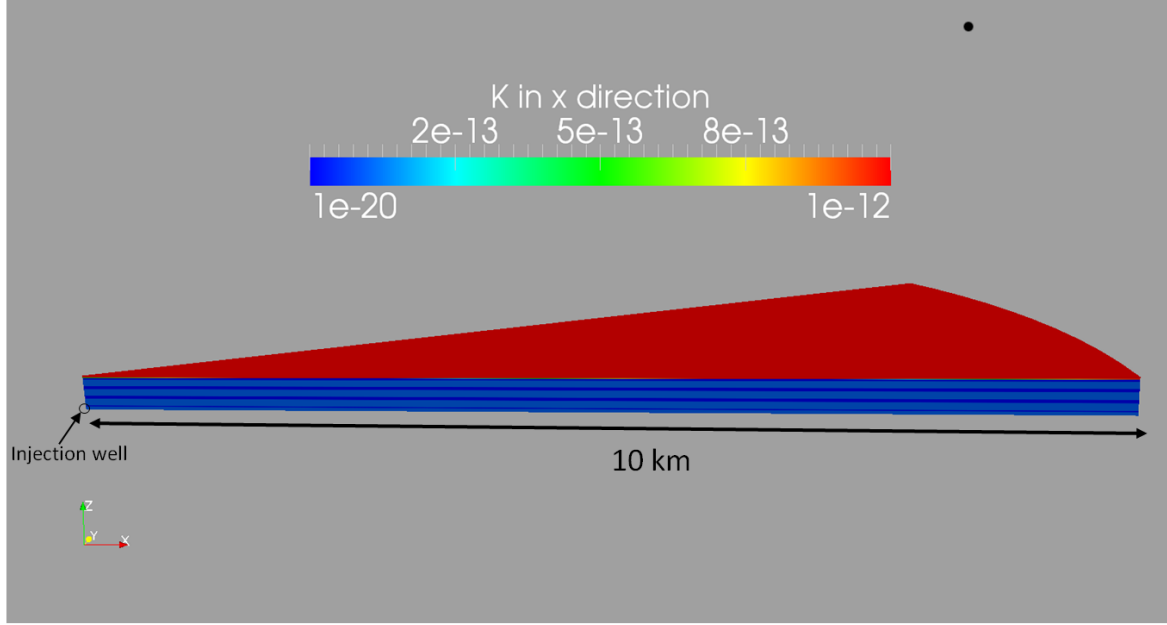
Figure 15.3.: Model domain (Walter et al., 2012) (1/12 segment of a circle).

model (no compositional effects considered), only advective discharge is calculated, which is defined for each phase as:

$$\text{Discharge} = \mathbf{K} \cdot ((\text{grad} p_\alpha) - \rho_\alpha \, \mathbf{g}) \, \rho_\alpha \, \lambda_\alpha A \cdot \mathbf{n} \,, \tag{15.1}$$

with the discharge in [kg/s], the intrinsic permeability $\mathbf{K}$ in [m^2], the potential gradient $(\text{grad} p_\alpha - \rho_\alpha g)$ in [Pa/m], the density ρ_α in [kg/m^3], the mobility λ_α in [1/Pa s], the surface area A in [m^2], and the unit normal vector $\mathbf{n}$. For the equations of the $2p$ model, we refer to (Flemisch et al., 2011) and for information about the fluid properties to (Bielinski, 2006). To account for rock compressibility, the following additional equation is applied in the model:

$$\Phi(p_g) = \Phi_{init}(1 + C(p_g - p_g^{ref})) \,, \tag{15.2}$$

where Φ [-] is the effective porosity, Φ_{init} [-] the initial porosity, C the rock compressibility of $4.5 \cdot 10^{-10}$ 1/Pa (see (Birkholzer et al., 2009)), p_g [Pa] the gas pressure, and p_g^{ref} [Pa] the initial reference pressure.

The horizontal permeabilities and porosities for the reference scenarios are shown in Figure 15.2. The anisotropy for all layers is set to 0.1 ($K_z = 0.1 \cdot K_x$). The porosity is related to the

different permeabilities of the layers using ((Holtz, 2002) and (Kumar et al., 2005)):

$$\Phi_{init} = \sqrt[9.61]{\frac{k_x}{7e^7[mD]}}, \qquad (15.3)$$

where k_x is the horizontal permeability in [mD]. This relation is compared with real data for permeability and porosity determined from core samples from the Ketzin site located near Berlin in Germany ((Norden et al., 2010)) and it fits well with the sampled data. For permeabilities lower than 1 mD (10^{-15}m^2), the relation is not verified by data. However, we also use this relation to determine the porosity in layers of low permeability and the caprock since the values obtained are reasonable. (Birkholzer et al., 2009) use similar values for the aquitards in their simulations. Based on this reference scenario, different variations are chosen to estimate which geological structure is associated with the highest risk of brine infiltration into the freshwater aquifer, see Section 15.3.

15.2.2. Arbitrary chaos expansion via probabilistic collocation method

A possible way of dealing with the statistical uncertainties is performing Monte Carlo simulations. Handeling this with acceptable computationally costs is possible with a reduced model, for example, using the polynomial chaos expansion technique (PCE). The principal idea of the classical chaos expansion, originally introduced by (Wiener, 1938), is to replace the full model by a response surface that is able to represent the model output (e.g. leakage, pressure) dependent on the chosen uncertain parameters. Therefore, the output of a simulation model is projected onto a higher order orthogonal basis of polynomials. The non-intrusive (i.e. "black-box") approach for the definition of the PCE form has lately been receiving increasing attention. The PCE technique combined with probabilistic collocation method (PCM) is described in detail in (Oladyshkin et al., 2011b), and an application is presented in (Walter et al., 2011).

In the current chapter we apply the most recent generalization of the chaos expansion technique ((Oladyshkin and Nowak, 2012a)), which accounts for arbitrary parameter input. The so-called arbitrary polynomial chaos expansion (aPC) approach provides improved convergence in comparison to classical PCE techniques, when it is applied to input distributions that fall outside of the range of classical PCE ((Oladyshkin and Nowak, 2012a)). Moreover, it allows to align the complexity and the order of analysis with the reliability and detail level

of statistical information on the input parameters. The fidelity of the reduced model has been tested via full model using classical PCE techniques in (Oladyshkin et al., 2011b) and using recent generalization aPC in (Oladyshkin et al., 2011a) for the CO_2 benchmark presented by (Class et al., 2009).

Here, only the main steps are summarized. The simulation output Ω can be projected onto an orthogonal basis of polynomials, i.e. the model output can be represented by the polynomial chaos expansion ((Oladyshkin et al., 2011a)). Thus, any model output Ω depending on the input parameters ω can be represented by the expansion as:

$$\Omega(\omega,x,y,z,t) = \sum_{j=1}^{P} c_j(x,y,z,t)\Psi_j(\omega).$$ (15.4)

The coefficients c_j quantify the dependence of the model output Ω on the input parameters ω. Here, Ψ_j is a simplified notation for the multi-variate orthogonal polynomial basis (polynomials of more than one variable) for the input parameters $\omega = \{\omega_1, \omega_2, ...\}$.

The CO_2 storage problem is solved at each grid point in space and time. Therefore, the polynomial form depends on the three spatial dimensions and the time. For the construction of the data-driven orthogonal polynomial basis, the only information required is the moments of the input distribution. If the basis is constructed, the unknown coefficients c_j have to be determined to represent the model output with the polynomials. The number of unknown coefficients P, which have to be defined, depends on the number of uncertain input parameters N and the desired order of the expansion d:

$$P = \frac{(N+d)!}{N!d!}.$$ (15.5)

The remaining step is to determine the unknown coefficients. In this work, the coefficients are obtained by the probabilistic collocation technique. The (full complexity) model has to be run P times with different sets of collocation points to evaluate the unknown coefficients c_j. The optimal choice of collocation points is the roots of the polynomial that is one degree higher than the order used in the chaos expansion. After the determination of the unknown coefficients, a set of polynomials forming a response surface is available for estimating the model output dependent on the uncertain model input. In the example for brine displacement during CO_2 storage presented here, the model outputs Ω are the pressure, the CO_2 saturation and the brine discharge into the water aquifer. The uncertain input parameters are the permeability of the reservoir, fault permeability, and anisotropy of all layers, which are described in Section 15.3.2.

15.3. Modeling brine displacement and migration

As explained earlier, the focus of this work is on a methodology applied to estimating the potential for brine displacement and migration for different geological situations. In this section, scenario uncertainty and statistical uncertainty are investigated and discussed.

15.3.1. Impact of scenario uncertainty

Scenario uncertainty is addressed by numerical simulations of various well-defined scenarios. Brine infiltration into the shallower freshwater aquifer is obtained by simulating the CO_2 injection process into a deep geological reservoir; in this case, we simulated a 25 years injection period and 25 years of the post-injection period. Various geological features such as fractures, faults, or seal weaknesses can influence the success of storage and are considered to explore uncertainties attributed to different scenarios. They influence the amount of brine infiltrating into a shallow freshwater aquifer.

Figure 15.4 shows the scenarios with the characteristics which are considered relevant by experts for the risk assessment. Scenario 1 is the base case with a closed caprock as described in Section 15.2.1. In Scenario 2, the caprock ceases to cover the CO_2 reservoir at 5 km distance from the injection. In Scenario 3, a vertically oriented fault zone of higher permeability ($K_{fault}=10^{-12}$ m^2) exists at the same horizontal distance. The aquitard zones along the fault zone are less permeable ($K= 5.5\cdot 10^{-15}$ m^2) in Scenario 4 compared to Scenario 3. Scenarios 5 and 6 consider worst-case (in terms of brine migration) modifications of Scenario 3, where low permeable barriers are introduced: in Scenario 5 only in the reservoir, in Scenario 6 over the whole depth. Such a list of scenarios is, of course, never complete, but should contain a representative selection of possible features. It could be extended, for example, by scenarios with inclined caprock, however not with this radially-symmetric approach.

Due to the grid resolution, only very simplified fault zones extending some hundred meters (in this study: 370 m) are investigated. Such broad fault zones are nevertheless reasonable since larger zones of different rocks are often found in geological formations, for example as a result of subrosion.

Results   Figure 15.5 shows the total brine mass discharge over time into the water aquifer for each scenario. Figure 15.6 presents the brine mass discharge only from the region, where

the fault zone is (S3-S6) or would be (S1, S2). Figure 15.7 illustrates the pressure profile directly above the caprock in a horizontal direction for selected cases.
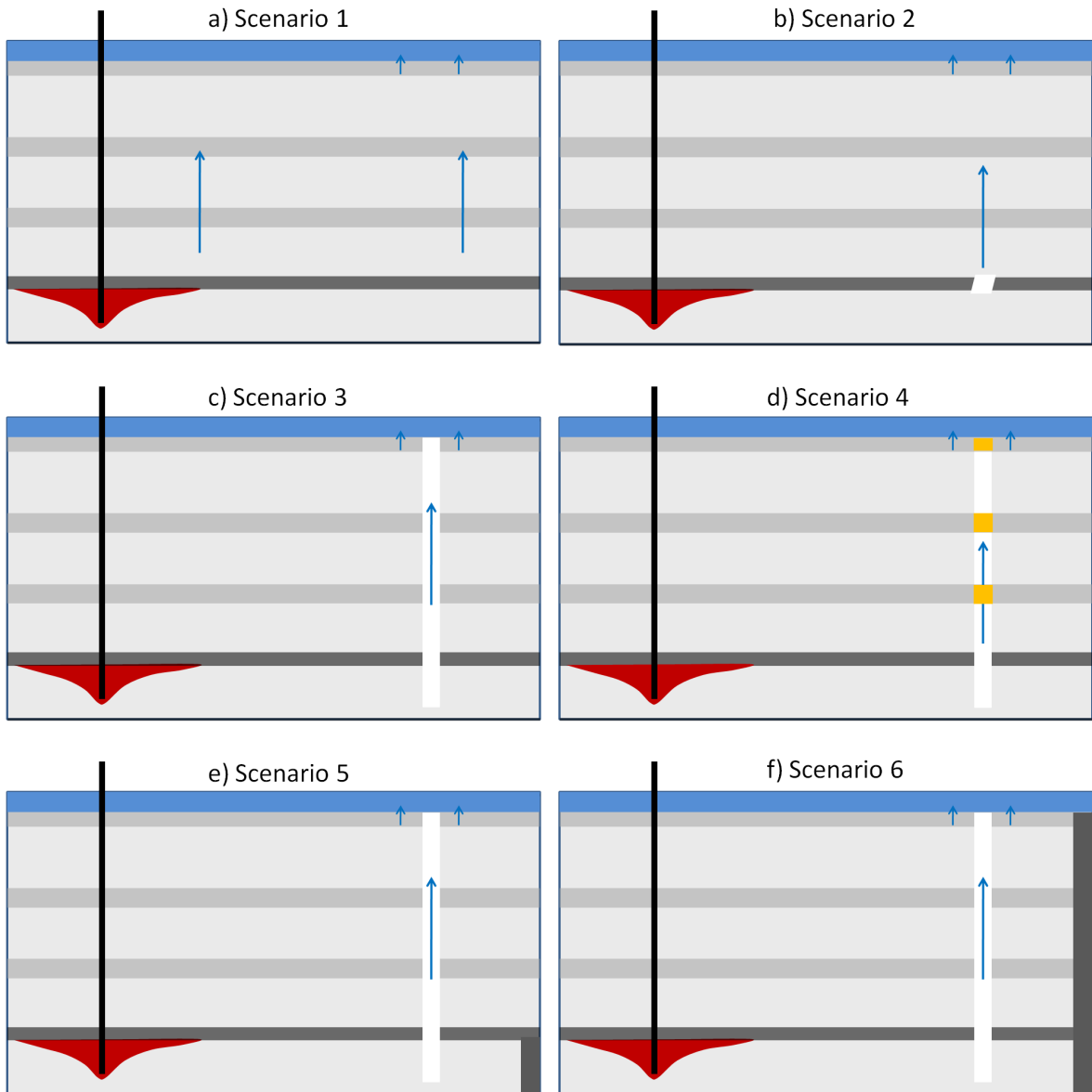
Figure 15.4.: Schematic view of the different geological scenarios (Walter et al., 2012). Scenario 1: reference scenario with closed caprock, see Figure 15.2; Scenario 2: gap in the caprock at about 5 km distance; Scenario 3: fault zone at about 5 km distance ; Scenario 4: lower permeability of the fault zone in the aquitard layers; Scenario 5: barrier of low permeability within the reservoir (at 9500 m); Scenario 6: barrier of low permeability over the whole domain (at 9500 m);
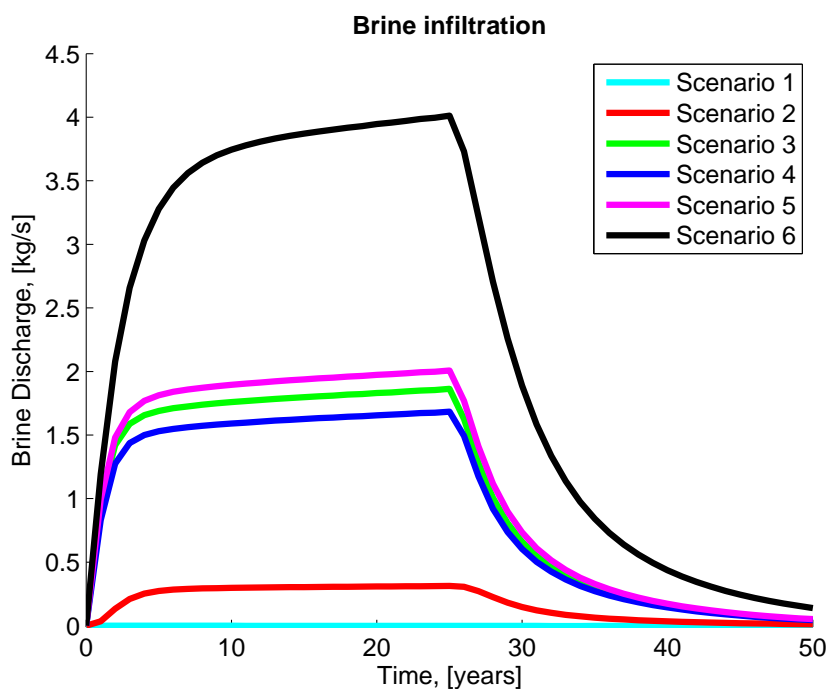
Figure 15.5.: Total brine discharge into the freshwater aquifer (over the whole domain) for the different scenarios.

In Scenario 1, i.e. the reference case, no brine discharge into the aquifer is observed (see Figure 15.5 and 15.6) because the caprock has no leaking fault and has a permeability of $K_{Caprock} = 10^{-20}$ m^2. The brine is displaced in a horizontal direction and leaves the model domain laterally through the open (constant-head) boundary at the end of the domain. The pressure profile above the caprock in Figure 15.7 shows no increase. This implies that brine is not displaced through the caprock. A simple calculation shows how much brine would actually be displaced if incompressibility was assumed for the fluids and the rock. The given CO_2 injection of 2.64 kg/s then corresponds to a displaced brine discharge of 4.53 kg/s.

In Scenario 2 (see Figure 15.4b), the caprock ceases at about 5 km distance and the overlying sediment has a high permeability. The gap in the caprock leads to a maximum discharge into the aquifer of 0.31 kg/s and the pressure above the caprock increases strongly (see Figure 15.5 and 15.7). However, the brine discharge through the gap, along the region, where the fault zone would be, is almost zero (see Figure 15.6), and the highly concentrated salt water from the reservoir does not reach the freshwater aquifer directly. Rather, the increase in pressure leads to a flow of brine from the layer directly below into the aquifer. The difference in salt concentration is not modeled because the salinity is constant in the model.
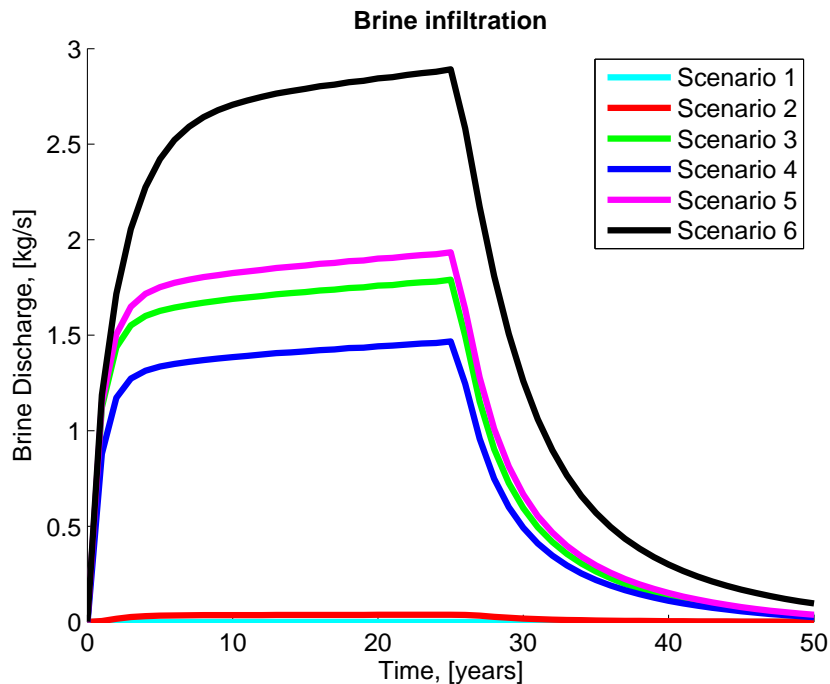
Figure 15.6.: Brine discharge into the freshwater aquifer (along the fault zone) for the different scenarios.

A clear pressure peak directly at the gap zone of the caprock can be identified. The aquitards in between the reservoir and the aquifer apparently do not act as completely impermeable layers like the caprock above the reservoir. However, the intermediate aquitards cause a partial retention of the brine and the highly concentrated salt water is held back in lower layers.

The importance of the low permeability aquitards above the caprock becomes clear when comparing Scenario 2 with Scenario 3, which has a fault zone at about 5 km distance throughout the vertical extent of the domain (see Figure 15.4c). Here, all aquitards are crossed by a highly permeable zone. A much higher discharge into the aquifer with a maximum of 1.86 kg/s occurs since the brine can be displaced more easily through the highly permeable fault zone. Comparing the brine discharge over the whole layer with the discharge along the fault zone alone shows that almost all the brine is displaced along the fault zone (97%). The pressure above the caprock is much lower than in Scenario 2, since the pressure can propagate more easily through the highly permeable fault zone.

Scenario 4 is a variation of Scenario 3: the fault zone is less permeable ($K=5.5 \cdot 10^{-15}$ m^2) where it passes through the aquitards (see Figure 15.4d). These barriers reduce the brine
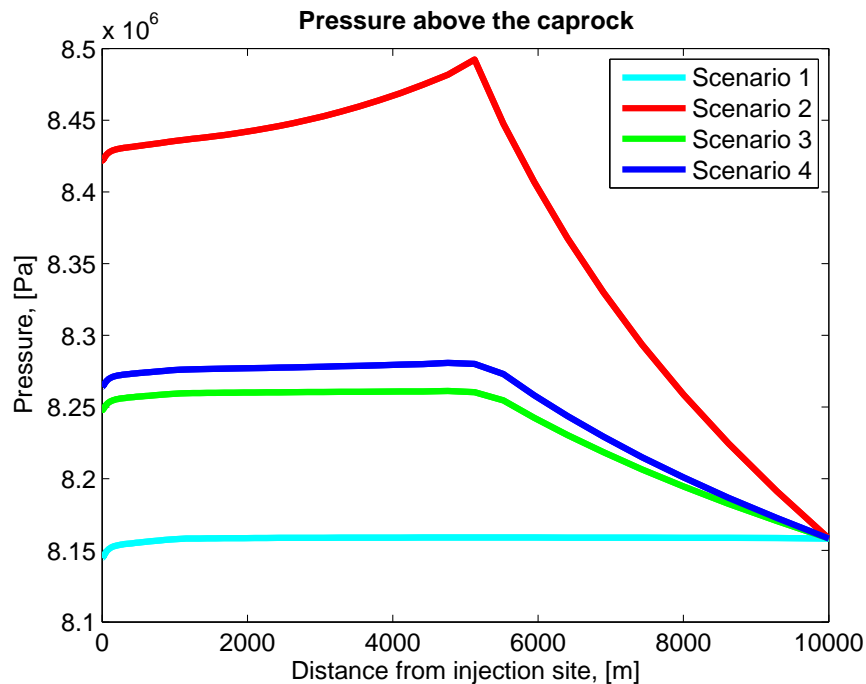
Figure 15.7.: Pressure above the caprock along the model domain ($p_{initial}= 8.1582 \cdot 10^6$Pa).

discharge and the ratio of discharge through the fault is slightly changed (91% discharge through the fault zone). The pressure peak above the caprock is higher than in Scenario 3 but still much lower than in Scenario 2, since the pressure can still propagate through the fault zone.

Scenario 5 and 6 are very unfavourable scenarios, where there is a fault zone as in Scenario 3 and there are barriers of low permeability at about 9500 m distance from the injection. In Scenario 5, the barrier of low permeability is only in the reservoir (see Figure 15.4e), whereas in Scenario 6, the barrier reaches from the reservoir through the whole vertical profile up to the aquifer (see Figure 15.4f). The discharge into the aquifer is increased in both scenarios in comparison to Scenario 3 since, in the latter case, more brine can be displaced horizontally from the reservoir (see pink and black lines in Figure 15.5). Particularly in Scenario 6, the observed discharge is more than double and reaches a value of 4.01 kg/s. Nearly all the displaced brine infiltrates the aquifer. An interesting point is that 97% of the brine still discharges through the fault zone in Scenario 5 but only 72% is displaced through the fault zone in Scenario 6. In Scenario 6, the pressure in the layer directly below the aquifer is higher due to the barrier; therefore, the displaced brine is distributed over a larger area and cannot be displaced in the horizontal direction. It infiltrates the aquifer due to the higher

pressure near the barrier. For Scenario 5, the pressure above the caprock is the same as in Scenario 3, and for Scenario 6, the pressure increases only slightly.

In most of the scenarios, the brine discharge has ceased after 50 years (25 years after CO_2 injection stops). A small amount of brine is still infiltrating the upper aquifer only in the worst of the selected cases (Scenario 6). An important issue is the different brine discharge areas of the scenarios. In Scenario 2, a discharge can be measured over the whole layer underneath the aquifer, but nearly no brine is displaced directly along the region, where the fault would be as in Scenario 3. This means that the discharged brine is water which originates from directly below the freshwater aquifer and is expected to have a significantly lower salt content than water from greater depth. This is not modeled in these simulations since a salt component of variable concentration is not explicitly considered and the salinity is constant (0.1 kg salt/kg solution). For the highly likely situation that salinity decreases at shallower depth, the obtained result is a conservative estimate. For the scenarios with a fault zone (scenarios 3-6), it is more probable than in Scenario 2 that highly concentrated brine reaches the aquifer, since it is displaced directly along the fault zone. The values for brine discharge along the fault zone in Scenarios 3-6 are more realistic since here, water with high salt content can migrate directly from the reservoir to the aquifer. However, these values are still overestimated since the rising, highly concentrated brine would mix with the surrounding, less concentrated water. These effects of dilution and mixing are a topic of ongoing work and not considered in the model presented here.

15.3.2. Impact of statistical uncertainty

Many parameters like reservoir permeability, porosity, anisotropy are uncertain. Often these uncertainties can be described statistically, but to do so requires adequate distributions. In the best case, measurement data are available for the uncertain parameters to provide an appropriate distribution. In the following, Scenario 4, which has a fault zone with lower permeability in the aquitard zones, is chosen as an example to investigate statistical uncertainties. The uncertain upscaled/averaged parameters are the permeability in the storage reservoir, the anisotropy for all layers except the fault zone, and the aquitard permeability in the fault zone. For the distribution of the reservoir permeability, the Ketzin data (see (Norden et al., 2010)) are used. The other distributions for anisotropy and fault permeability are selected based on a reasonable range of values. Anisotropy is chosen between 0 and almost 1 with a higher probability for smaller values. The data for aquitard permeability in the fault zone
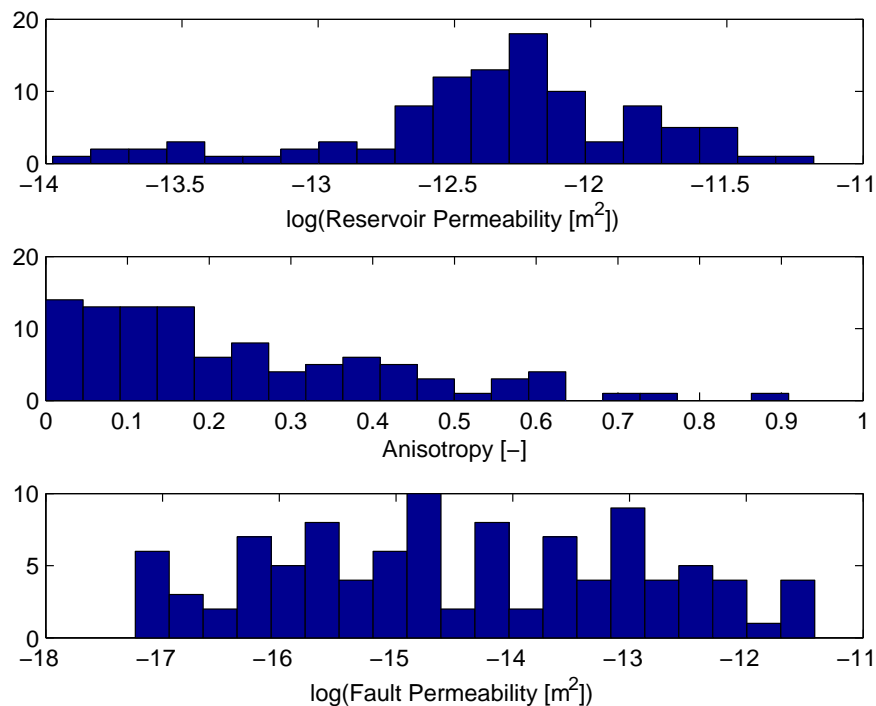
Figure 15.8.: Distributions for the uncertain parameters: Permeability of the reservoir, anisotropy, and fault permeability in the aquitard layers.

are varied between 10^{-17}m^2 and 10^{-12}m^2. For both parameters, a beta distribution is taken as the basis. Figure 15.8 shows the resulting distributions of the three parameters.

Third-order polynomials are used to project the complex model onto a response surface. Third-order is a sound trade-off between accuracy and computational effort ((Oladyshkin et al., 2011b)). Three uncertain parameters and third-order polynomials result in $P = 20$ unknown coefficients in the expansion, see Eq. (15.4), and as a consequence require 20 simulation runs (snapshots). The four collocation points chosen for each parameter are listed in Table 15.1.

Results With set-ups of 20 different combinations of the collocation points, the original model is projected onto the polynomials. For a detailed interpretation of the dependencies, Figure 15.9 shows the change in brine discharge for each uncertain parameter while the other parameters are held constant. The points shown represent a selection of snapshot runs with the same values for the remaining parameters, and the polynomials are constructed to

Reservoir permeability [m^2]	Anisotropy [-]	Fault permeability [m^2]
$1.7826 \cdot 10^{-14}$	0.0435	$1.2278 \cdot 10^{-17}$
$1.3915 \cdot 10^{-13}$	0.2291	$4.4076 \cdot 10^{-16}$
$6.9912 \cdot 10^{-13}$	0.5294	$4.1062 \cdot 10^{-14}$
$3.4914 \cdot 10^{-12}$	0.8573	$1.3027 \cdot 10^{-12}$

Table 15.1.: Collocation points for snapshot simulations.

fit through these points. In the following, the dependence of the brine discharge on each parameter and the accuracy of the polynomials are discussed.

For anisotropy, the polynomial reflects the continuously increasing behavior very well (see Figure 15.9a). Higher anisotropy increases the brine discharge into the aquifer. Increasing the anisotropy means a lower flow resistance in vertical direction. The anisotropy in all layers except for the fault zone is increased. Thus, more brine is displaced vertically and the brine discharge into the aquifer becomes greater.

An increase of the fault permeability has a similar effect as increasing anisotropy (see Figure 15.9b). More brine infiltrates the aquifer with higher fault permeability, since the resistance of the flow path along the fault is decreased. This continuous increase diminishes for high-value fault permeabilities and the brine discharge remains almost constant (see the two snapshot runs with higher permeability in Figure 15.9b) because the resistance is already low enough to let the displaced brine flow. The behavior of the brine discharge depending on the fault permeability is reflected well by the polynomial in regions of lower permeability. In the region of higher permeabilities, the highly nonlinear behavior cannot be reflected adequately by the polynomials. The extrapolation for higher permeabilities outside the range of snapshot simulations yields the wrong characteristics, and one has to be careful while using the polynomials for risk assessment. The limitations of the collocation method will be discussed in more detail in Section 15.5.

While both anisotropy and fault permeability lead to a more or less monotonic response of the brine discharge, this is not the case for reservoir permeability. At the beginning of injection, the brine discharge first increases and then decreases again with increasing permeabilities (see Figure 15.9c after 2 years of injection). If the permeability is very low, the pressure during CO_2 injection will be higher, due to higher flow resistance. The higher pressure increases the effect of compressibility of the fluids and the rock. Thus, less brine needs to be displaced. Higher permeability of the reservoir results in smaller pressure peaks. Thus,
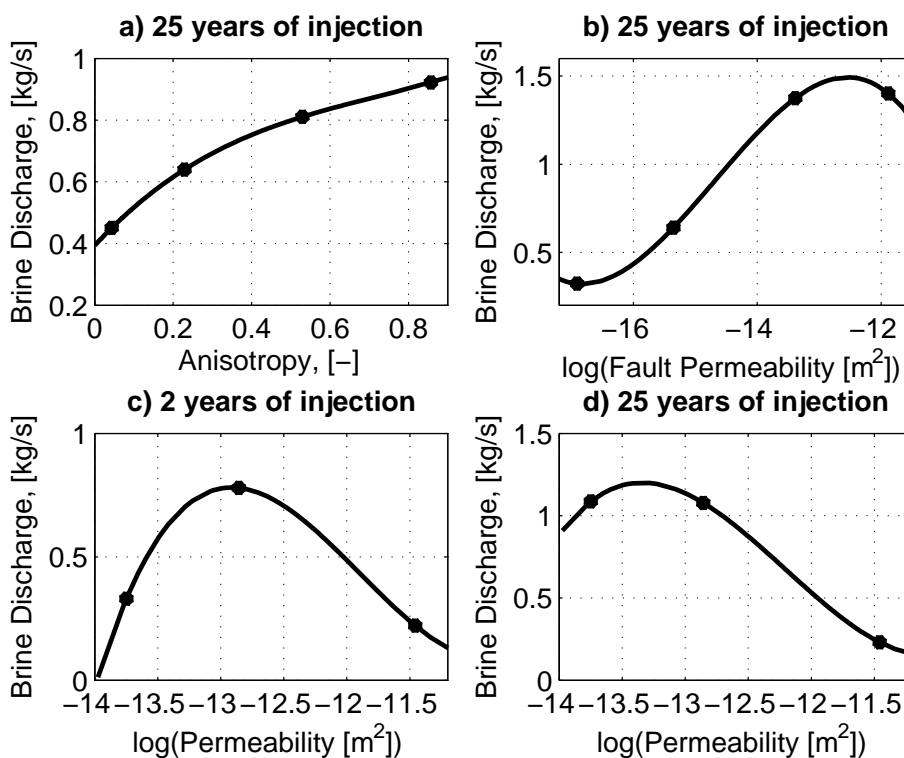
Figure 15.9.: Brine discharge versus the uncertain parameters: Snapshot runs (bullets) and fitted polynomials a) for anisotropy after 25 years of injection, b) for fault permeability after 25 years of injection, c) for reservoir permeability after 2 years of injection, and d) for reservoir permeability after 25 years of injection

at lower pressures, the compressibility effect is reduced. Consequently, more brine has to be displaced to free the required pore space, and more brine infiltrates the aquifer. At a certain value of the reservoir permeability, the brine discharge decreases again. Here, the permeability of the reservoir is similar to the permeability of the fault zone. In this situation, more brine can be displaced in horizontal direction from the reservoir instead of vertically against gravity. At this early stage of injection, the polynomials resemble the behavior well.

As the injection proceeds, the situation changes and the effect of very small permeabilities in the reservoir on the reduction of brine discharge diminishes (see Figure 15.9d: Permeability after 25 years of injection). The two snapshot runs for smallest permeabilities result in almost the same brine discharge. At later stages, the compressibility is less important since the pressure at the injection well does not rise further and the displacement becomes more relevant. Thus, the effect that the brine discharge is lower for smaller permeabilities cannot be

observed. Comparing the snapshot runs at the collocation points with the fitted polynomial at later times shows that the behavior is described well by the polynomial for higher permeabilities but a discrepancy can be identified for smaller values. Between the two snapshot runs on the left, the polynomial overestimates the brine discharge because of overshoot in the third-order polynomial. This overestimation is acceptable for the risk calculations since it gives a conservative estimate. However, the steep decrease for further decreasing permeabilities is not seen in further simulations and should not be considered in the risk estimation.

15.4. Risk assessment

15.4.1. Quantification of risk

The approach to quantify risk, followed in this work is often found in the literature, and defines risk as ((Kaplan, 1997))

$$\text{Risk} = \text{Probability} \times \text{Damage} . \tag{15.6}$$

This means that the risk is the expected loss incurred by an accident, which is defined as damage multiplied by the probability of an accident occurring. In (Kopp et al., 2010), risk is used as a statistical leakage probability that could occur at a well and is calculated by multiplying the likelihood of failure times the consequence of failure. In their work, only the cases were considered when CO_2 leaks. In this study, no explicit failure event occurs for the considered Scenario 4, because in any case brine is displaced and some damage always occurs. Risk, as it is defined here, is the probability-weighted expectation of damage.

The mathematical definition of damage has to be chosen carefully for each risk assessment individually depending on the hazard which is evaluated. It is necessary that damage can be derived from the model output of the numerical simulations. In this study, the damage is defined as the amount of infiltrated brine per time (brine discharge) into the shallower freshwater aquifer:

$$\text{Damage} = \frac{\text{Mass of brine[kg]}}{\text{Time[s]}} . \tag{15.7}$$

To determine the probability of a certain damage, a stochastic approach is required. As explained in Section 15.2.2, we reduce the complex model via the arbitrary polynomial chaos expansion (aPC) combined with the probabilistic collocation method (PCM). Mean values

and variances of the model output can be easily computed with the polynomials. To calculate the probabilities, Monte Carlo calculations on the polynomials can be performed, which is much faster than using the complex model for hundreds or thousands of runs. This can provide a probability density function (PDF) and cumulative distribution functions (CDF) for the damage, where the evidence probability can be determined for a certain damage (e.g. high pressures or brine discharge events).

Scenario uncertainty For the category of scenario uncertainty, there is no statistically representative information on the features of a selected scenario available. Thus, it is not possible to determine probabilities on the basis of the aPC/PCM approach. To this end, we calculate only damage values for the scenarios and compare them. The brine discharge calculated and discussed in Section 15.3.1 corresponds to the damage as defined in this work. With the results from the previous section, the damage in the scenarios can be ranked (S6>S5>S3>S4>S2>S1). Scenario 6 with a fault zone and a barrier of low permeability results in the biggest damage, and is thus the most unfavourable geological structure within the selected scenarios. It is followed by Scenario 5, which results in a smaller damage due to the fact that the brine can also flow in horizontal direction as soon as it leaves the reservoir. Scenario 3 with the laterally open reservoir reduces the damage further, which is also the case when there is a lower permeability in the aquitard zones of the fault (Scenario 4). A substantial reduction of damage is predicted for Scenario 2, and Scenario 1 is the safest one with practically no damage at all. It can be concluded that the caprock integrity and faultless overlying aquitards are the key geological features to keep brine migration to shallower aquifers at a minimum. Other geological features also have significant impact on brine migration, but not to the same extent as faults in the caprock.

Statistical uncertainties When uncertainty is statistical, both damage and probability can be calculated, since the distribution of parameters like permeability or anisotropy can be quantified. Here, a cumulative distribution function (CDF) and a probability density function (PDF) are used, and risk is estimated for Scenario 4. Figure 15.10a shows a CDF for the brine discharge after 25 years of injection for the given uncertain parameters, in this case permeability in the reservoir, fault permeability, and anisotropy. If a certain threshold value for a maximum brine discharge is given, the probability that this value is not exceeded can be determined by the CDF. Example: For a given threshold of 1.84 kg/s, the probability that this value is not exceeded is 95%. Figure 15.10b shows the PDF of the damage (brine discharge). Overall it says that damage values smaller than 1 kg/s, for example, are more probable than
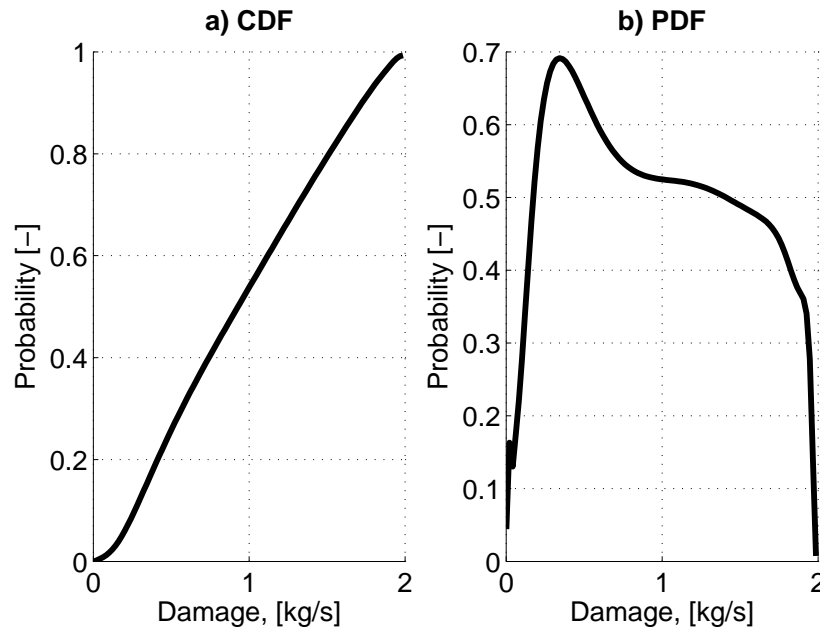
Figure 15.10.: Cumulative distribution function and probability density function for brine discharge into the water aquifer after 25 years of injection.

damages above this value. However, the risk calculation would shift the significance to larger values of damage since damage and probability are multiplied. An overall risk value for the given Scenario 4 can be estimated with the mean value for risk, i.e. the expected value of risk:

$$E[\text{Risk}]_i = \frac{\sum_i^N \text{Damage}_i \text{Probability}_i}{N} \tag{15.8}$$

where N is the number the sample size of the model parameters. For Scenario 4, the overall risk is then 0.4722 kg/s. We interpret this value as the probability-weighted expectation of damage for this particular scenario.

15.4.2. Estimates for salt concentration and the impact on the freshwater aquifers

Using estimates of damage and the risk of brine discharge into the freshwater aquifer from the methods, the question of the impact on the freshwater aquifer, i.e. the drinking-water resources can be addressed. For potable water, salt concentrations, or total dissolved solids (TDS) must not exceed a certain value. The World Health Organization (WHO) does not regulate the amount of TDS; however several countries define a critical value for TDS, e.g.

Denmark defines a critical value of 1500 mg/l, and Australia does not allow TDS higher than 500 mg/l ((Rygaard et al., 2011)). Germany allows a maximum concentration of 250 mg/l for chloride ions and 200 mg/l for sodium but no cumulative value for TDS. The gap between the calculated risk of brine discharge, obtained from numerical methods, on the one hand and salt concentration in drinking water wells on the other hand is bridged with some simple analytical considerations.

Fully mixed approach   Let us assume that the freshwater aquifer can be considered a fully mixed reactor. A steady-state salt-mass balance can then be formulated as:

$$q_{out} A_{out} c_{out} = q_{in} A_{in} c_{in} + SD_{brine} , \qquad (15.9)$$

where q [m/s] and c [kg/m^3] are the groundwater flow and concentration into and out of the control volume, A [m] is the area of the cross section, S [-] the salinity used in the numerical simulations, and D_{brine} [kg/s] the mass discharge of brine as given by Equation (15.7). It is assumed that the incoming groundwater flow contains no salt ($c_{in} = 0$), yielding the following equation for the salt concentration in the aquifer:

$$c_{out} = \frac{12 S D_{brine}}{q_{out} w_f b_{aq}} . \qquad (15.10)$$

$A_{out} = w_f b_{aq}$, w_f is the width of the diluting volume and b_{aq} is the thickness of the aquifer (here 22.5 m). The brine discharge is multiplied by 12 since the results of the numerical model are only valid for a twelfth of the whole area and we consider the entire circular domain here. The salinity of the brine discharge into the freshwater aquifer is $S = 0.1$ (as before). For the groundwater flow q_{out}, 1m/d=1/86400m/s is assumed. Figure 15.11 shows the top view of the model domain with the fault circle and the segment for the numerical simulation. The assumption of a circle-shaped fault is not very realistic, but it simplifies calculations, thus fulfilling its purpose in this principal study. Two different approximations are now discussed. In the first, the brine discharge over the whole layer (the whole circle) is used. Here, the width of the diluting volume is twice the width of the full model domain ($w_{f1} = 2 \times 10000m$). For the second approximation, only the brine discharge through the fault zone is considered and the width of the diluting volume is twice the distance from the injection site to the fault zone ($w_{f2} = 2 \times 5500\ m$).

Table 15.2 lists the maximum brine discharge obtained from the numerical simulations and the corresponding concentrations calculated by Equation (15.10) for the two different approximations and for each scenario. The concentration from the approach that considers
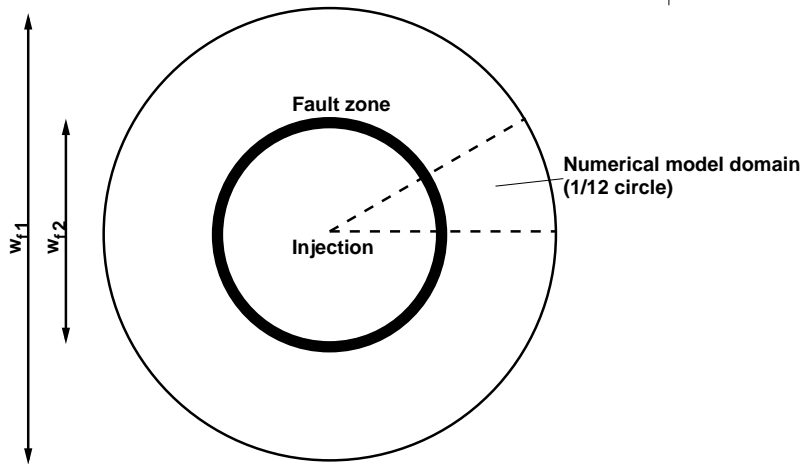
Figure 15.11.: View of the model domain from above (Walter et al., 2012).

Scenario	Brine discharge whole domain		Brine discharge fault only	
	D_{brine} [kg/s]	c_{out} mg/l	D_{brine} [kg/s]	c_{out} mg/l
1	0	0	-	-
2	0.3135	72.0	-	-
3	1.8631	429.2	1.8063	756.7
4	1.6236	387.9	1.4803	620.1
5	2.0079	462.6	1.9466	815.5
6	4.012	924.4	2.892	1211.5

Table 15.2.: Maximum brine discharge (from numerical simulations) and estimated salt concentration for the different scenarios and for the two different approximations: (i) brine discharge over the whole domain, and (ii) brine discharge through the fault zone

mixing over the whole domain is lower despite the larger brine discharge due to the larger diluting volume. For Scenarios 1 and 2, no salt concentrations from infiltration via the fault zone are calculated, since the fault zone is not connected to the freshwater aquifer. The concentrations c_{out} of each scenario are compared with the drinking water criteria, given in (Rygaard et al., 2011). The guideline value of Denmark (1500mg/l) is not exceeded in any of the scenarios if the brine discharge is assumed to occur over the whole domain. The most critical one is Scenario 6, which exceeds the TDS criterion of Australia but not that of Denmark. For the case where the infiltration occurs only through the fault zone, the drinking-water criterion in Australia is exceeded in all four scenarios, while the Danish criteria would

still tolerate these concentrations. The validity of the approximation can be tested using the mixing length L_{mix} [m] (see (Binning and Celia, 2008)):

$$L_{mix} = \frac{b_{aq}^2}{2\,\alpha_T}. \tag{15.11}$$

Here, $\alpha_T = 0.018$ m is a reasonable value for the transverse dispersivity ((Binning and Celia, 2008)). L_{mix} defines how far downstream the aquifer can be assumed to be fully mixed. For the water aquifer above, the mixing length is 14.06 km, which means that the calculated values (Table 15.2) are valid for a water production well at a distance of at least 14.06 km downstream of the discharge zone. Closer to the zone of discharge, concentrations can be higher.

These calculations prove that it is just as important to calculate the diluting volume (given by $w_f \times b_{aq}$) as it is to calculate the brine discharge D_{brine}. Depending on the diluting volume, the fully-mixed concentration in the drinking water aquifer changes strongly. As already indicated in Section 15.3.1, it appears more realistic to consider the infiltration associated with the fault zone since, in Scenarios 3-6, highly concentrated salt water can be displaced through the fault zone directly from the reservoir up to the water aquifer. The assumption of a fully-mixed system mitigates this unfavorable situation because of the large diluting volume, and it has to be stated again that this situation is only valid far downstream from the discharge zone.

Both approximations show high values for the salt concentrations compared with the drinking water criteria, which at least in part results from the fact that the salt content is not explicitly modeled in the numerical simulation. The initial salinity is set to 0.1 over the whole domain, which means that, even directly underneath the water aquifer, the model calculates with this high salinity. Furthermore, mixing effects cannot be addressed with the current numerical model. Therefore, the concentrations shown are rather conservative estimates. Explicitly accounting for the salt concentration and a variable salinity over depth is ongoing work and not within the scope of this chapter.

Analytical approximation of salt-water upconing   Near the zone of brine infiltration, i.e. for distances less than L_{mix}, the fully-mixed assumption cannot be made. A conservative approach is then to assume that concentrated brine is at the bottom of the aquifer and has impacts on the water produced by upconing at the drinking water wells. It is interesting to determine the operating conditions under which the concentrated salt water will be pumped in the well. As soon as upconing salt water reaches well screen, it will affect the

quality of the water produced. Simplified estimates with an analytical approximation allow us to develop an operational strategy for water suppliers.

(Schmorak and Mercado, 1969) and (Todd, 1980) presented an analytical approach to investigate the upconing of sea water caused by a drinking-water well. This approach can be adapted to the brine infiltration problem here. A few assumptions are necessary. First a sharp interface between the salt and the freshwater is assumed. Furthermore, only the region directly beneath the drinking water well ($r = 0$) is considered and it is assumed that the system is at steady state ($t \rightarrow \infty$). This assumption leads to a conservative estimation of upconing and thus is appropriate for risk assessment. The situation under consideration is shown in Figure 15.12. According to (Schmorak and Mercado, 1969) and with these assumptions, the upconing of the salt water can be defined as:

$$Z(r = 0, t \rightarrow \infty) = \frac{Q}{2\Pi d \frac{\Delta \rho}{\rho_f} k_x}, \qquad (15.12)$$

where $Z[m]$ is the rise of the interface above the initial state, Q [m^3/s] the pumping rate in the well, d [m] the distance between the salt-water layer and the drinking water well, $\frac{\Delta \rho}{\rho_f}$ the dimensionless difference between salt and freshwater density ρ_f and k_x $[m/s]$ the horizontal permeability. (Schmorak and Mercado, 1969) stated that this relation is limited to a critical rise Z_{crit}. When Z_{crit} is reached, a spontaneous breakthrough of the salt water occurs. In their experiments, they found that sudden breakthrough occurs for upconing higher than $Z_{crit} = d/3$.

For the water aquifer already investigated in the numerical simulations, Table 15.3 shows the values of model parameters at the depth of 500 m. Let us assume that the brine forms a thin layer at the bottom of the aquifer. Then the distance d between the salt layer and the well

$Q[m^3/s]$	0.004
$d[m]$	22.5
$T[K]$	298
$p[bar]$	50.16
$\rho_{brine}[kg/m^3]$	1068.5
$\rho_f[kg/m^3]$	998.2
$K_x[m^2]$	10^{-12}
$k_x[m/s]$	$1.0978 \cdot 10^{-05}$

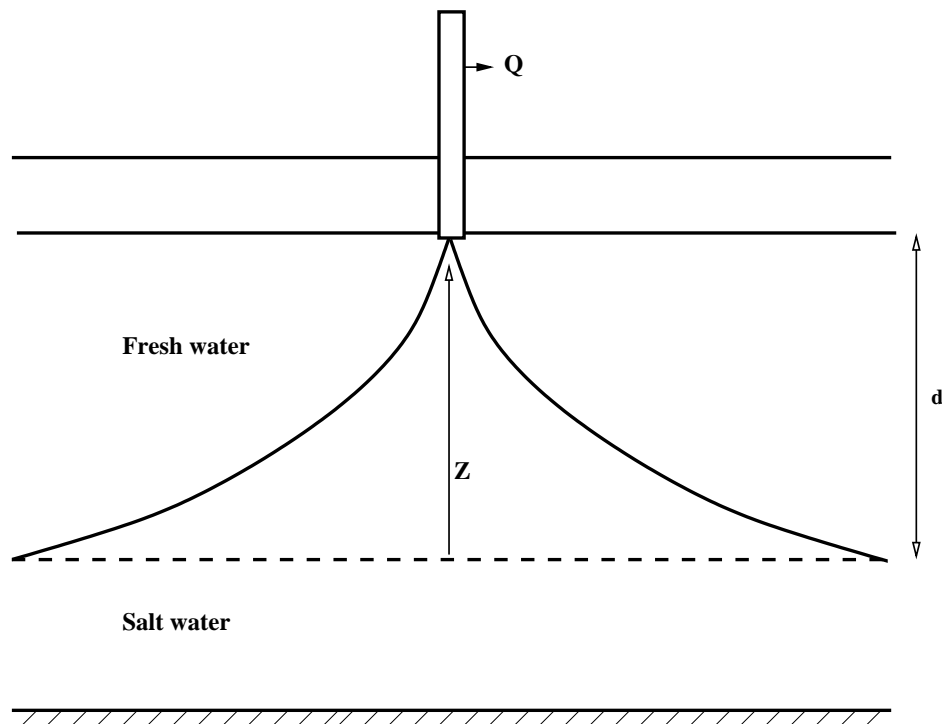Table 15.3.: Physical values for the given problem.

Figure 15.12.: Upconing of salt water to a pumping well after (Schmorak and Mercado, 1969).

screen in Equation 15.12 is the aquifer thickness (22.5 m).

Figure 15.13 shows the approximation of the salt-water rise for different vertical distances between the salt-water layer and the drinking-water well. A constant pumping rate of 0.004 m^3/s in the well is assumed. It is obvious that a smaller vertical distance between the well and the salt-water layer results in higher upconing. Without considering the effect of a sudden rise, the salt water would affect the well for distances smaller than ca. 29 m. For larger distances d, the well is not reached by saline water. However, to ensure that no salt is pumped into the drinking water well, the upconing needs to fall below the critical rise of $Z_{crit} = d/3$. This linear relation is also shown in Figure 15.13. For the situation to the left of the intersection of the two curves ($d < 50$ m), the upconing has exceeded Z_{crit}, and the well is affected by saline water. For larger distances, Z_{crit} is not exceeded and freshwater can be pumped at the given rate.

Equation (15.12) further allows an estimation of how strongly the drinking-water well is affected by the saline water. For example, the green point in Figure 15.13 represents the situation where the salt-water layer is 22.5 m below the well. In this case, salt-water upconing is greater than 37 m. This corresponds to a height of about 14.5 m salt-water infiltration into
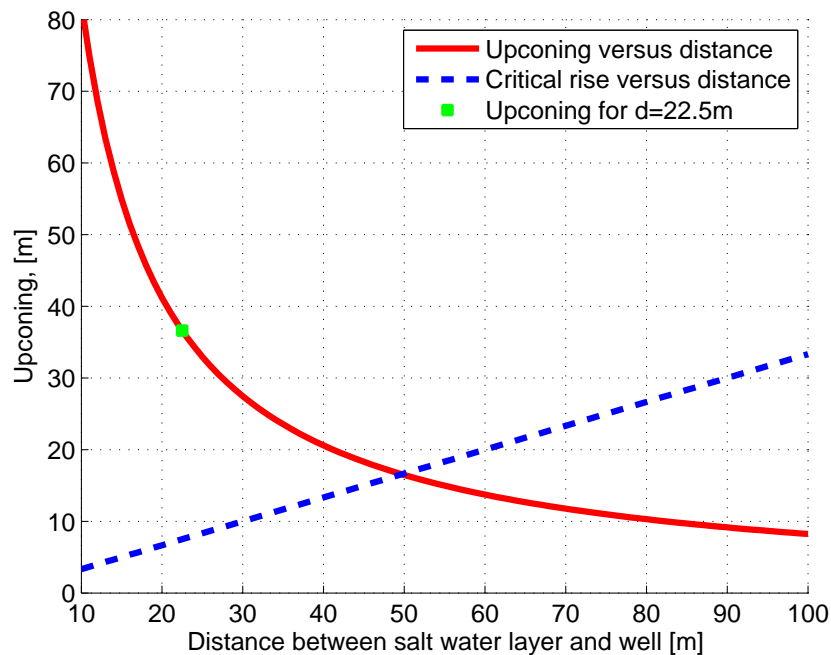
Figure 15.13.: The rise of the brine interface versus distance from brine layer to the pumping well (using $Z_{crit} = d/3$).

the well.

The pumping rate of an operational strategy is obviously a design parameter. According to Equation (15.12), the upconing depends linearly on the pumping rate. Without considering the critical rise and the abrupt jump of the brine layer, the well would be reached by salt water with a pumping rate of 0.0025 m^3/s. However, considering the critical rise, the maximum permissible pumping rate is only $8.194 \cdot 10^{-4}$ m^3/s.

An interesting point is that the upconing is inversely proportional to the density difference between salt and freshwater. A higher salt content will result in a low estimate for upconing. However, the greatest impact has upconing occuring at the highest concentration. Figure 15.14 shows how the upconing decreases with increasing density difference.

The section concludes by examining how the analytical model can be linked to the brine discharge obtained from the numerical simulation. A first approach is to check at which salt concentrations upconing of brine would actually occur. As discussed above, upconing is inversely proportional to the density difference between salt water and freshwater. The maximum concentration at which upconing occurs can be calculated. If the concentrations in the water aquifer are above this critical concentration, then upconing is not an issue. For
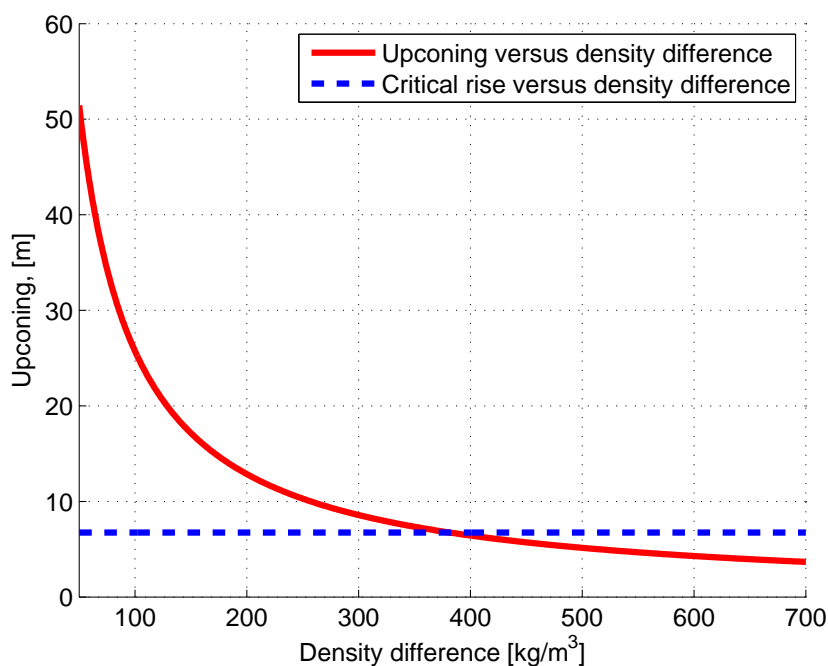
Figure 15.14.: The rise of the brine interface versus the density difference.

this example, the values in Table 15.3 are used with the reservoir thickness of $d = 22.5$ m and a pumping rate of 0.004 m^2/s. The maximum density difference is 343.03 kg/m^3 and the corresponding salinity equals 0.4233. The concentration is calculated by multiplying the salinity with the brine density of 1341.3 kg/m^3 resulting in $c_{crit} = 567.78$ kg/m^3. Such a high salinity has, of course, no physical meaning since the water will be fully saturated with salt for a salinity of about 0.35. Thus, it must be concluded that in the presented example with the small aquifer thickness of $d = 22.5$ m, upconing will always occur. This means that within a radius of 14.06 km from the brine discharge zone, the drinking water could be affected by the salt concentration.

15.5. Conclusions

The study has shown that it is important to address uncertainties on different levels for a comprehensive assessment of risk related to CO_2 storage. In this approach, a distinction is made between scenario uncertainty and statistical uncertainty. Scenario uncertainty requires expert opinion to identify relevant geological scenarios; statistical uncertainty requires information on parameter distributions as well as a concept for testing the parameter space,

for example with Monte Carlo methods. The concept applied here uses Monte Carlo testing on computationally very cheap polynomials which represent the response surface of the full-complexity models on the basis of the arbitrary polynomial chaos expansion combined with the probabilistic collocation method. The ultimate aim is to determine a reliable order of magnitude for occurring damages and corresponding probabilities; in this study, damage represents brine discharge and, eventually, the concentration of salt in a drinking water well.

The suggested practical work flow of risk assessment includes different steps.

(i) It is necessary to identify those scenarios which are both realistic and lead to high damage. Expert opinion and geological knowledge is inevitable. This was shown in Section 15.3.1. The total volume of displaced brine depends on the volume of injected CO_2, although compressibility effects can lead to a time-shift of brine displacement. The displaced brine has several possible escape routes: (a) towards the outer regions of the injection horizon if the formation is large enough and not sealed at its edges, (b) through the caprock, or (c) through faults. The study gives clear evidence that fault structures are the most important features in these scenarios and brine discharge to freshwater aquifers crucially depends on the connectivity between the storage reservoir and the freshwater aquifer. The likelihood of large brine discharge into freshwater aquifers is far smaller if faults do not connect the two layers, even if the caprock is not faultless.

(ii) For a selected scenario with given geological features (gap in the caprock, fault zone, aquifers, aquitards, etc.), the statistical uncertainties can be included, i.e. the variability of parameters such as permeability, porosity, or anisotropy. Section 15.3.2 addressed this issue for Scenario 4. Knowing the expected range of the tested parameter space and the corresponding distribution functions, it is possible to quantify probabilities of certain damage events. The aPC method allows, for example, the calculation of response surfaces and cumulative distribution functions. For the particular geological set-up of Scenario 4, a value could be assigned to the risk of brine discharge into the freshwater aquifer (here: 0.4722 kg/s).

(iii) The risk of brine discharge needs to be transformed into the risk of exceeding drinking-water criteria in water-production wells. Using simple analytical approaches, it could be shown that this can be done on the basis of a few assumptions. For example, a fully mixed state of the aquifer appears realistic if the production well is far enough from the discharge zone. If the salt intrusion into the freshwater aquifer occurs rather close to a production well, fully-mixed conditions are not achieved and salt water can be assumed to form a layer at the bottom of the reservoir due to gravity. In such a case, upconing of salt water is the relevant mechanism that could lead to salt concentrations in the water production well (see Section 15.4.2).

A conservative assumption in the presented study is that of a constant salinity of all water below the freshwater horizon. This overestimates the concentration arriving in the aquifer, and future work will consider this by using spatially coupled models with a multiphase model in the region where the CO_2 phase spreads and a single-phase compositional model in the major part of the model domain.

This study suggests that salt-water intrusion into freshwater aquifers due to CO_2 storage is probably a locally confined problem, at least in terms of salt concentration. It occurs not over large areas if storage sites with multiple barriers are chosen. To which extent geochemistry is changed by intruding brine remains an issue to be investigated.

This study showed that the probabilistic collocation method combined with the arbitrary polynomial chaos expansion could not reproduce all the non-linearities occurring in the model. Over- and undershoots between the collocation points can occur with even more problems beyond the collocation points. However, according to PCM the modeler can extract a lot of information in the main range of the parameter distribution, which is sufficient to understand the principal behavior. In that way, the aPC via PCM framework helps to keep the compromise between the computational effort and degree of expansion. Higher-order polynomials could probably improve this; however, only at higher computational costs for determining the polynomials.

In the simulations of this work, the calculated overall brine discharge varied from almost zero to about 4 kg/s for the different scenarios, while the statistical variability considered in Scenario 4 lead to a brine flux variation between zero and almost 2 kg/s. Therefore, it appears as if scenario uncertainty is a little higher here, but still of the same order of magnitude as statistical uncertainty. More knowledge decreases both scenario and statistical uncertainty.

Finally, we claim that the presented set of methods is a reasonable trade-off between computational effort and desired accuracy for combining quantitative and qualitative uncertainty assessment in modelling CO_2 storage.

16. Sensitivity analysis and risk assessment for large-scale CO$_2$ storage

Bibliographic Note: The content of this chapter is based on the following original article: Ashraf M., Oladyshkin S., Nowak W., Geological storage of CO$_2$: global sensitivity analysis and risk assessment using arbitrary polynomial chaos expansion. International Journal of Greenhouse Gas Control, Elsevier, V. 19, P. 704-719, 2013.

Geological storage of CO$_2$ is a proposed interim solution for mitigating the climate change. Modeling CO$_2$ storage is accompanied by huge geological uncertainties and excessive computational demands. However, the considerable costs and potential hazards of the technique require feasibility studies to assess all possible risks. This makes computationally efficient methods for sensitivity analysis, uncertainty quantification and probabilistic risk assessment indispensable.

Our goal is to demonstrate the application and feasibility of the arbitrary polynomial chaos expansion (aPC) for these tasks under realistic conditions. We model a typical CO$_2$ injection scenario in realistic geological realizations of a shallow marine deposit. Our scenario features uncertain parameters that control the structure of geological heterogeneities, including the density of barriers, the aggradation angle, fault transmissibility and regional groundwater effects. The aPC approximates the models by a polynomial-based response surface to speed up the involved statistical analysis of an otherwise expensive simulation tool.

We demonstrate how such an analysis can guide further exploration and the design process of finding suitable injection rates. Our case study demonstrates clearly that the aPC is an efficient, feasible and hence valuable approach in this context, and we strongly encourage its future use. A key advantage of the aPC over more conventional polynomial chaos methods is the flexibility to work with arbitrary probability distributions of uncertain parameters. From

our featured parameters, we found the aggradation angle to be the most and the regional groundwater effect to be the least influential one. To the best of our knowledge, this is the first analysis of structural parameters for geological heterogeneities in the CO$_2$ context and within a probabilistic setting.

16.1. Introduction

In the context of climate change mitigation, geological storage of CO$_2$ has been proposed as interim solution. The idea has been challenged during the last decades for its costs and potential hazards (Lenzen, 2011; Viebahn et al., 2007). A large number of studies has been performed in the industry and research communities to evaluate the safety and feasibility of CO$_2$ storage, addressing issues such as the status and barriers of CO$_2$ storage (Bachu, 2008), screening and ranking of geological storage sites (Bachu, 2003), large-scale impacts of CO$_2$ injection in deep saline aquifers (Birkholzer et al., 2009), new solution methodologies for CO$_2$ leakage (Nordbotten et al., 2005b), the capture project (Thomas, 2005), and leakage estimates (Celia et al., 2004). Furthermore, many pilot projects have been installed, like In Salah (Riddiford et al., 2004), Ketzin (Förster et al., 2006), and Johansen (Eigestad et al., 2009). A discussion on the experiences from the existing pilot projects is reported in (Michael et al., 2010).

Yet, there is a big demand for studies which demonstrate the appropriateness of the storage operation. Transparent scientific results are required to communicate the facts and evidences about feasibility and possible risks within public and industry. The large involved time and space scales, however, cause substantial computational issues in such studies (e.g., (Class et al., 2009)), and the modeling procedure is accompanied by a huge extent of geological uncertainties (e.g., (Walton et al., 2004; Brennan et al., 2010; Wilson et al., 2003; Hansson and Bryngelsson, 2009b)).

In an approach to quantify the impact of geological heterogeneity on model predictions of multiphase flow in geological formations, a large number of shallow marine depositional realizations has been generated and used in the sensitivity analysis of the impact of geological uncertainties on production forecasting (SAIGUP), see (Howell et al., 2008; Manzocchi et al., 2008a; Matthews et al., 2008) . There, the impact of variable geological parameters has been quantified for oil recovery in different field development scenarios. The main general conclusion of that study is that realistic features of geological uncertainty in modeling

(other than typical hydrological parameters) can lead to considerable uncertainties in prediction. The papers (Ashraf et al., 2010a,b) used a number of SAIGUP realizations to study the impact of geological heterogeneity on the injection and early migration of CO_2 in a shallow-marine aquifer with a complex, heterogeneous geological structure. That study transferred the significance of some of the geological structural features to the case of CO_2 injection.

In practice, modeling complicated physical phenomena in the subsurface requires stochastic approaches. Uncertainty can exist in different levels, from the formulation of dependency rules in the model to uncertainty about appropriate values for the model input parameters. Uncertainty coming from any source in the modeling procedure propagates through the model to the predicted responses. Ranking the important model parameters based on their influence on the model responses can support a better understanding of the system, and it can result in a better design of subsequent studies on the stochastic nature of the process. Hence, identifying and evaluating the sensitivities and uncertainties of model parameters and their impact on prediction uncertainties and projected risks is a significant task. Sensitivity analysis is known to be the right approach to identify the significance of uncertainty sources within the modeling process (Oladyshkin and Nowak, 2012a) and to improve the understanding of model behavior (Sobol, 2001). For example, the European Commission and the United States Environment Protection Agency recommend using sensitivity analysis in the context of extended compact assessment for policy making (Commission, 2002).

Uncertainty sources within the CO_2 storage problem can be classified in different types as geological, physical and operational uncertainties. This work is devoted to geological uncertainties. However the same procedure can be applied to extend the work for other types as well. Here, we use a set of SAIGUP realizations to perform a sensitivity analysis and to assess the risks caused by uncertainties in a choice of parameters that govern the geological structure of the featured shallow-marine deposit.

The goal of this study is to test and demonstrate the applicability of a recent set of methods to a realistic scenario. We choose a stochastic response surface method to project the model response to parameter changes onto high-dimensional polynomials via the arbitrary polynomial chaos expansion (aPC) (Oladyshkin and Nowak, 2012a; Oladyshkin et al., 2011a). Highly similar ideas to the aPC have also been proposed in other scientific areas (Witteveen et al., 2007; Witteveen and Bijl, 2006; Ghanem and Doostan, 2006; Soize and Ghanem, 2004). As we review in Section 16.2, the involved orthogonal polynomial basis can be constructed for arbitrary probability distributions of the uncertain parameters. This data-driven approach provides fast convergence (Oladyshkin and Nowak, 2012a) in comparison

to the classical polynomial chaos expansion (e.g., (Wiener, 1938; Ghanem and Spanos, 1991; Le Maître and Knio, 2010)). Moreover, it avoids the subjectivity of data treatment that would arise when being forced to fall back onto a limited number of theoretical distributions that can be tolerated with previous generalized versions of polynomial chaos expansions (Wan and Karniadakis, 2007; Xiu and Karniadakis, 2002b). The reduced model represented by the response surface is significantly faster than the original complex one, and thus provides a promising starting point for global sensitivity analysis, uncertainty quantification, and probabilistic risk assessment.

In the current work, we use global sensitivity analysis rather than a local one, because local analysis fails to cover the non- linear variation of model responses over the entire range of probability distributions of the input parameters. A practical approach in global sensitivity analysis is to work with the impact of uncertain parameters on prediction variances, because this shows a good success in nonlinear problems (Reuter and Liebscher, 2008). In the current study, we use Sobol indices (Sobol, 2001) for sensitivity analysis, which are indeed working with variances. The fact that the aPC based response surface is based on orthonormal polynomials with exploitable known properties (Oladyshkin et al., 2011d) substantially simplifies this analysis.

Finally, we perform risk analysis by applying a Monte-Carlo procedure to the response surface. The approximating polynomial is fast enough to be used for a large number of Monte-Carlo realizations. This makes it possible to cover the entire range of variations in the model input described by the assigned probability distributions, and thus provides accurate estimates for the risk in the system. We conclude with a discussion of the results.

The global sensitivity analysis and uncertainty quantification studies for CO$_2$ storage existing in the literature are concerned with classic hydrological uncertain parameters like porosity, pore volume and permeability as global constants (see for example (Brennan et al., 2010; Kovscek and Wang, 2005; Oladyshkin et al., 2011a)). To the best of our knowledge, the current study is the first one that implements the proposed mathematical analysis tools on realistic geological structural parameters at reservoir scale. The parameters we consider are the level of barriers presence, aggradation angle, fault transmissibility, and regional groundwater effects. The considered features are the structural and depositional features that dictate the distribution of hydrological parameters such as permeability and porosity, both in terms of value and spatial distribution. These are among the most uncertain geological parameters identified with the SAIGUP study (except the regional groundwater effect, which is specific to this study).

16.2. Response surface via arbitrary polynomial chaos expansion

Working with uncertain parameters in complex, non-linear and dynamic systems puts a high demand on stochastic tools to analyze the system and to propagate uncertainties through the system. Conceptually straightforward numerical Monte Carlo (MC) techniques are computationally demanding since the statistical accuracy of their predictions depends on the number of realizations used. The Monte-Carlo estimation error (measured as standard deviation) for output statistics typically decreases only with the square root of the number of realizations used. Using a stochastic response surface is a promising approach in this respect.

Obviously, a response surface can be constructed in different ways, e.g. it can be constructed directly on a dense Cartesian grid of input parameters at extremely high computational efforts. In the current study, we apply an alternative methodology which demands only a minimum number of model evaluations to construct the response surface. This approach is based on the theory of polynomial chaos expansion (PCE) introduced in (Wiener, 1938). Generally, all PCE techniques can be viewed as an efficient approximation to full-blown stochastic modeling (e.g., exhaustive MC). The basic idea is to represent the response of a model to changes in variables through a response surface that is defined with the help of an orthonormal polynomial basis in the parameter space. In simple words, the dependence of model output on all relevant input parameters is approximated by a high-dimensional polynomial. The resulting polynomials are functions of the model parameters. This projection can be interpreted as an advanced approach to statistical regression.

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

Unfortunately, the original PCE concept (Wiener, 1938) is optimal only for Gaussian distributed input parameters. To accommodate for a wide range of data distributions, a recent generalization of PCE is the arbitrary polynomial chaos (aPC (Oladyshkin et al., 2011a)).

Compared to earlier PCE techniques, the aPC adapts to arbitrary probability distribution shapes of input parameters and, in addition, can even work with unknown distribution shapes when only a few statistical moments can be inferred from limited data or from expert elicitation. The arbitrary distributions for the framework can be either discrete, continuous, or discretized continuous. They can be specified either analytically (as probability density/cumulative distribution functions), numerically as histogram or as raw data sets. This goes beyond the generalization of PCE in methods such as the generalized polynomial chaos (gPC) or the multi-element gPC (ME-gPC) (Wan and Karniadakis, 2007; Xiu and Karniadakis, 2002b). The aPC approach provides improved convergence in comparison to classical PCE techniques, when applied to input distributions that fall outside the range of classical PCE. A more specific discussion and review of involved techniques will follow in Sections 16.2.1 to 16.2.3.

With an introduction to response methods via the aPC, we describe here the theoretical background that we use in our modeling procedure. The related techniques for sensitivity and risk analysis used in this work are explained in Sections 16.4 and 16.5.

16.2.1. Definitions and polynomial chaos expansion

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

$$\Gamma \approx \Upsilon(\Theta). \tag{16.1}$$

Like all PCE methods, the aPC is a stochastic approach to approximate the response surface. Considering the uncertainty in the input variables, the aPC constructs a set of polynomial basis function and expands the solution in this basis. Thus, the response vector Γ in Eq. (16.1) can be approximated by (Oladyshkin and Nowak, 2012a):

$$\Gamma \approx \sum_{i=1}^{n_c} c_i \Pi_i(\Theta). \tag{16.2}$$

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

The number n_c of unknown coefficients c_i results from the number of possible polynomials with total degree equal to or less than d. This number depends on the degree d of the

approximating polynomial, and the number of considered parameters n:

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

(16.3)

16.2.2. Data-driven orthonormal basis

All polynomials Π_i in expansion (16.2) are orthogonal, i.e., they fulfill the following condition:

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

(16.4)

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

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

16.2.3. Non-intrusive determination of the coefficients

The next task is to compute the coefficients c_i in Eq. 16.2. Generally, all PCE techniques can be sub-divided into intrusive (Ghanem and Spanos, 1993; Matthies and Keese., 2005; Xiu and Karniadakis, 2003) and non-intrusive (Keese and Matthies, 2003; Isukapalli et al., 1998; Li and Zhang, 2007; Oladyshkin et al., 2011b) approaches, i.e., methods that require or do not require modifications in the system of governing equations and corresponding changes in simulation codes. The challenge in choosing between the methods is to find a compromise

between computational effort for model evaluations and a reasonable approximation of the physical processes by the interpolation.

For our study, we prefer the probabilistic collocation method (PCM: see (Oladyshkin et al., 2011a; Li and Zhang, 2007; Oladyshkin et al., 2011b)) from the group of non-intrusive approaches like sparse quadrature (Babuska et al., 2007; Xiu and Hesthaven, 2005; Gerstner and Griebel, 2003; Barthelmann et al., 2000). In a simple sense, PCM can be interpreted as a smart (mathematically optimal) interpolation and extrapolation rule of model output between and beyond different input parameter sets. It is based on a minimal and optimally chosen set of model evaluations, each with a defined set of model parameters (called collocation points). For this reason, the collocation approach became more popular in the last years. Also, the collocation formulation does not require any knowledge of the initial model structure. It only requires knowledge on how to obtain the model output for a given set of input parameters, which allows treating the model like a "black-box". The distinctive feature of non-intrusive approaches is that any simulation model can be considered a "black-box", i.e. commercial software can be used without any modifications required.

According to (Villadsen and Michelsen, 1978), the optimal choice of collocation points corresponds to the roots of the polynomial of one degree higher $(d + 1)$ than the order used in the chaos expansion (d). This choice adapts the position of collocation points to the involved distribution shape, and is based on the theory of Gaussian integration (e.g., (Abramowitz and Stegun, 1965)). For one-dimensional problems (i.e., when analyzing only one uncertain model parameter), it allows exact numerical integrations of order $2d$ given $d + 1$ values of the function to be integrated.

For multi-parameter analysis, the number of available points from the corresponding Gaussian integration rule is $(d + 1)^n$, which is larger than the necessary number M of collocation points. The minimum value of M is equal to the number of coefficients n_c in Expansion (16.2), according to Eq. (16.3). The full tensor grid can be used only for low-order (1st, 2nd) analysis of few parameters. For higher-order analysis of many parameters, the tensor grid suffers from the curse of dimensionality (a full tensor grid in n dimensions requires $(d + 1)^n$ points, which rises exponentially in n)(Nobile et al., 2008). In that case, a smart choice of a sparse subset from the tensor grid becomes necessary. Then, PCM chooses the minimum required number of collocation points, equal to the number of coefficients n_c, from the full tensor grid according to their probability weight, i.e. according to their importance as specified by the available probability distribution of Θ. This simply means to select the collocation points from the most probable regions of the input parameter distribution (see

(Oladyshkin et al., 2011b)).

The weighted-residual method in the random space is defined as (Li and Zhang, 2007):

$$\int (\Gamma - \sum_{i=1}^{n_c} c_i \Pi_i(\Theta)) w(\Theta) p(\Theta) d\tau = 0, \qquad (16.5)$$

where $w(\Theta)$ is the weighting function and $p(\Theta)$ is the joint probability density function of Θ. Please note that choosing $w_i = \Pi_i$ in Eq. 16.5 results in the method discussed by (Ghanem and Spanos, 1991) and (Le Maître and Knio, 2010). In PCM, the weighting function is chosen as the delta function:

$$w(\Theta) = \delta(\Theta - \Theta_c). \qquad (16.6)$$

Θ_c is the set of collocation points. Substituting from Eq. (16.6) into Eq. (16.5) gives the following:

$$\Gamma_c - \sum_{i=1}^{n_c} c_i \Pi(\Theta_c) = 0, \qquad (16.7)$$

where Γ_c are the response values corresponding to the collocation values Θ_c. We solve Eq. (16.7) to find the coefficients c_i.

Hence, in total, n_c detailed runs are required to determine the n_c unknown coefficients. The roots of the data-driven polynomial basis (see Section 16.2.2) define the positions of the collocation points specific to the distribution of input parameters at hand and, thus, indicate the optimal parameter sets for model evaluation, using all available information about the input parameters. In our study, we have $n = 4$ uncertain parameters and we use a polynomial of degree $d = 2$. This means that only $n_c = 15$ detailed runs are necessary to obtain the expansion coefficients and approximate the response surface.

16.3. CO$_2$ storage problem

Here, we describe the injection scenario for which we analyze sensitivities, uncertainties, and risks in Sections 16.4 and 16.5. The same flow responses are studied here as in (Ashraf et al., 2010a,b). These are CO$_2$ pressure, CO$_2$ mobile and residual volumes and leakage risk as described below. Then, we describe the uncertain parameters considered in the study followed by a discussion on the uncertain structural aspects of the considered geological settings.

16.3.1. Modeling scenario

A typical scenario of CO$_2$ injection is defined in which a volume of 40×10^6 m^3 is injected via one well during an injection period of 30 years. This volume corresponds to 20% of the total aquifer pore volume. After stopping injection, simulation continues for 70 years to study the early migration of the CO$_2$ plume. For brevity, we omit the detailed model equations here and refer the interested reader to (Oladyshkin and Nowak, 2012a; Oladyshkin et al., 2011a).

In our scenario, we feature an aquifer system that is formed by shallow-marine deposits. There is one closed boundary on the top side of the model and the other sides are assumed to be open (Figure 16.1). All the open boundaries are modeled as Dirichlet boundaries, two of which with hydrostatic pressure distribution (the right and bottom boundaries in Figure 16.1). The remaining left boundary is also hydrostatic, but modified in order to account for the regional groundwater effect (see below).

The cells on the faces of the open boundaries are equipped with a very large pore volume multiplier, such that they numerically represent a much larger volume and effectively enlarge the domain. This helps to minimize the boundary effects of a computational domain that would otherwise be relatively small compared to the injected CO$_2$ volume (about 20% of the total pore volume, see above). The pore volume multiplier technique allows for a physically reasonable pressure build-up close to the boundary. Moreover, this allows the CO$_2$ that has left the domain to re-enter by gravity segregation after the injection has stopped.

A summary of the used parameter values is given in Table 16.1. The hydrological parame-
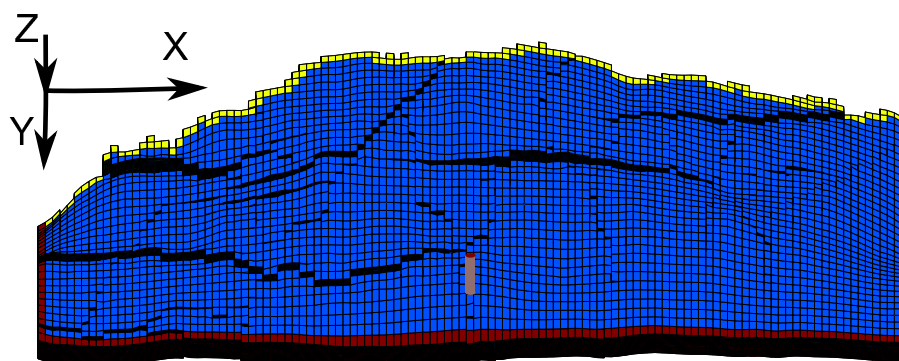

Figure 16.1.: Boundary conditions and the well location in the designed injection scenario (Ashraf et al., 2013). Red color corresponds to the open boundaries and yellow color shows the closed side on the crest.

ters like permeability and porosity vary within individual realizations due to the considered geological structure (see Figure 16.2 for the histograms of porosity and permeability in one selected realization). They also differ between the different realizations, as they are changed to represent different geological features. Although the geological realizations of this model vary in some geological features, but the same total pore volume, grid, and fault geometry is considered. The injection well is screened in the lower part of the model.

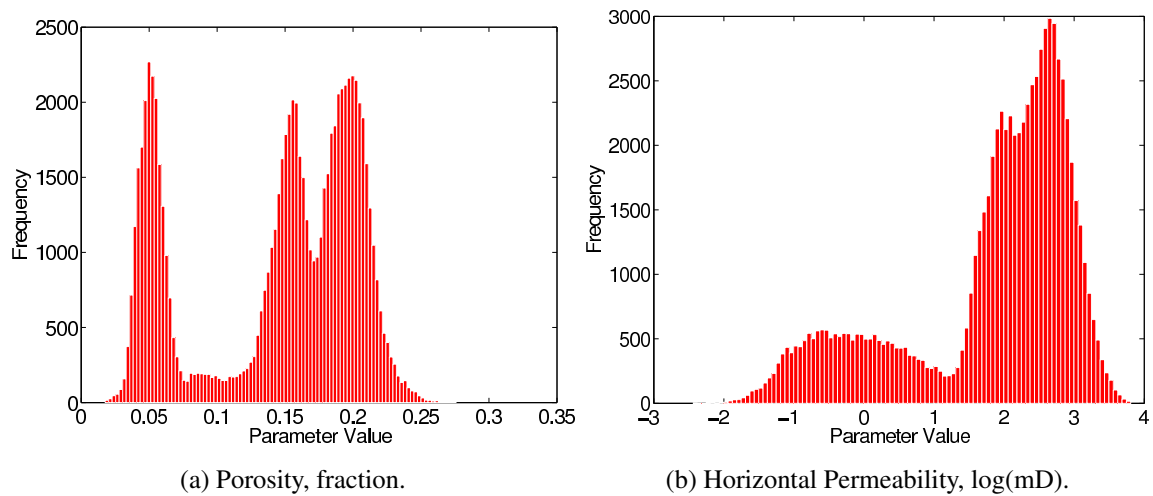(a) Porosity, fraction.

(b) Horizontal Permeability, log(mD).

Figure 16.2.: The histograms of hydrological parameters shown for a realization with low levels of heterogeneity. The vertical permeabilities are approximately one order of magnitude lower than the horizontal permeabilities.

Table 16.1.: Aquifer model information.

Parameter	Value	Unit
Number of active cells in the model	78720	-
Resolution X,Y,Z	$40 \times 120 \times 20$	-
Scale X,Y,Z	$3000 \times 9000 \times 80$	m
Injection rate	3650	m^3/day
Initial pressure	266.5	bar
Critical CO_2 and water saturations	0.2	-
CO_2 viscosity	0.04	cp
Water viscosity	0.4	cp
Rock compressibility	0.3e-6	1/bar

16.3.2. Analyzed model predictions

We seek to maximize the CO_2 storage volume and minimize the risk of leakage. These quantities are measured by various simulation outputs that are described in Table 16.2 and discussed in the following.

CO_2 pressure is considered as the spatial average of the pressure distribution in the entire domain, weighted by the CO_2-filled pore volume in each model cell. Monitoring or predicting the pressure response within the CO_2 plume is important to avoid over-pressurized injection operations.

Residual CO_2 volume is the volume of trapped CO_2 that is left in the small pores in an imbibition process. This volume is crucial for the long-term storage capacity of reservoirs.

Mobile CO_2 volume is the volume of CO_2 that can move in a continuous phase in the medium. It is considered as one of the important flow responses, because only mobile CO_2 volumes can lead to leakage through any failure in the sealing cap-rock or ill-plugged well.

Finally, we consider **leakage risk** through cap-rock failure. Cap-rock integrity is a major concern for the safety of CO_2 storage operations. An over-pressurized injection can lead to fractures that may extend up to the cap-rock, penetrate through the cap-rock, or activate pre-existing faults and fractures, and finally lead to CO_2 leakage. In addition, the capillary barrier effect of the cap-rock can be overcome by a local pressure build-up. Thus, the probability of cap-rock failure can depend on the geomechanical properties of the cap-rock and of the medium, on the topography of the cap-rock, and on the pressure build-up resulting from the CO_2 injection and migration. More details about failure mechanisms and failure criteria can be found in the literature (e.g., (Zweigel and Heill, 2003; Aker et al., 2013; Rohmer and Seyedi, 2010)). However, geomechanical modeling and knowledge about pre-existing

Table 16.2.: Important model responses and their brief description. For more information, see (Ashraf et al., 2010a,b).

Response	Description
Average CO_2 pressure	Volume average of pressure, weighted by CO_2 volume
Mobile CO_2	Volume of CO_2 in places with saturation above critical value
Residual CO_2	Volume of CO_2 in places with saturation below critical value
Leakage risk	A risk value for the leakage through the cap-rock.

features that can be activated during injection would be required to take these processes into account.

Here, we demonstrate how cap-rock integrity can be considered in the workflow of sensitivity analysis and uncertainty assessment in a simplified manner. To avoid detailed studies of multiphase flow coupled with geomechanical simulations and fracture mechanics, we follow a pragmatic approach. The idea is to assign a spatial probability distribution of cap-rock failure over the area of the cap-rock layer, such that each point of the cap-rock has its own failure probability. In principle, this probability could be assigned in correspondence with the current pressure distribution and with geological features such as varying cap-rock thickness, material properties, faults and fractures. For the means of demonstration, we simply assign a spatial Gaussian function as a scenario assumption to provide the cap-rock failure probability for each point of the cap-rock (see Figure 16.3). Leakage risk is defined as the probability of leakage (due to cap-rock failure) times the amount of escaping CO$_2$ in case of leakage. Thus, we spatially integrate the product between cap-rock failure probability and the volume of mobile CO$_2$ below each point of the cap-rock over the entire area of the cap-rock.
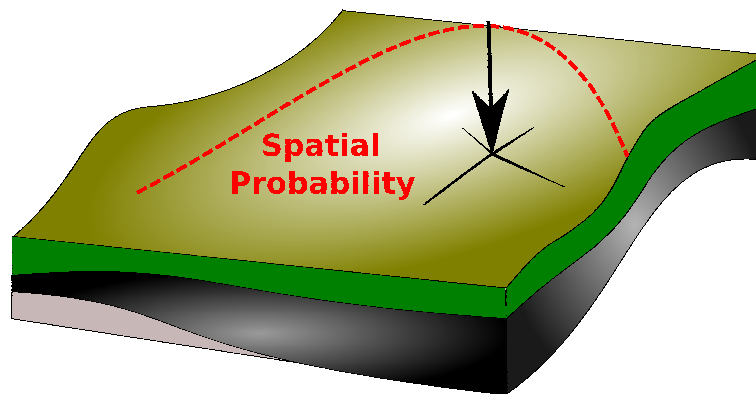

Figure 16.3.: CO$_2$ leakage risk is computed as the product of a cap-rock failure probability and the amount of mobile CO$_2$ beneath the cap-rock, integrated over the entire surface area of the cap-rock (Ashraf et al., 2013). Here, we use a Gaussian function as simple scenario assumption for the cap-rock failure probability (indicated schematically by the color shading and the dashed red line with the black coordinate system).

16.3.3. Uncertain parameters

The most apparent uncertainty in CO$_2$ storage is the lack of geological knowledge. Large geological scales and diversity of rock properties make it impossible to obtain the whole descriptive picture for a study. A geological study will therefore be accompanied by huge levels of uncertainty. Many studies have shown the significance of geological heterogeneity on underground flow performance (e.g., (Dutton et al., 2003; Eaton, 2006)). To obtain a descriptive image of a feature, like faults and depositional structure, such that uncertainty can be reduced, we must provide adequate data. The process of data collection from underground layers is very costly, therefore it is important to know the ranking of influence each feature has on the flow in order to optimize the cost of data acquisition in modeling.

From the geological parameters that are relevant for shallow-marine deposits used in (Ashraf et al., 2010a,b), we pick three parameters: the degree to which barriers may block horizontal and vertical flow, aggradation angle, and fault transmissibility. In addition to these, we consider the regional groundwater effect as an uncertain parameter in our study. Here, we give a brief description on each one, followed by the probabilities assigned to these parameters.

Barriers: During the formation of shallow-marine deposits, periodic floods result in a sheet of sandstone that dips, thins, and fines in a seaward direction. In the lower front, thin sheets of sandstone are inter-bedded with the mud-stones deposited from suspension. These mud-draped surfaces are potential significant barriers to both horizontal and vertical flow. In the SAIGUP realizations, these barriers were modeled by transmissibility multipliers in specific layers of the formation. The position of the barriers is generated by creating an elliptic cone-
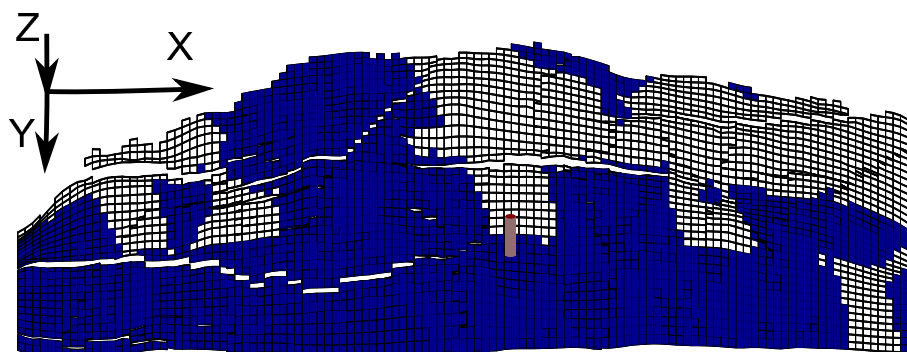

Figure 16.4.: The figure shows 50% of zero transmissibility multipliers in a specific model layer representing a medium level of barriers (Ashraf et al., 2013). One layer of the model is shown in the figure.

shaped surface that follows the plan-view shoreline shape of the facies, characterized from real world data (Howell et al., 2008). We define the degree of barrier presence by the areal percentage of zero-valued transmissibility multipliers. Figure 16.4 shows a medium level of barriers.

Aggradation angle: in shallow-marine systems, two main factors control the shape of the
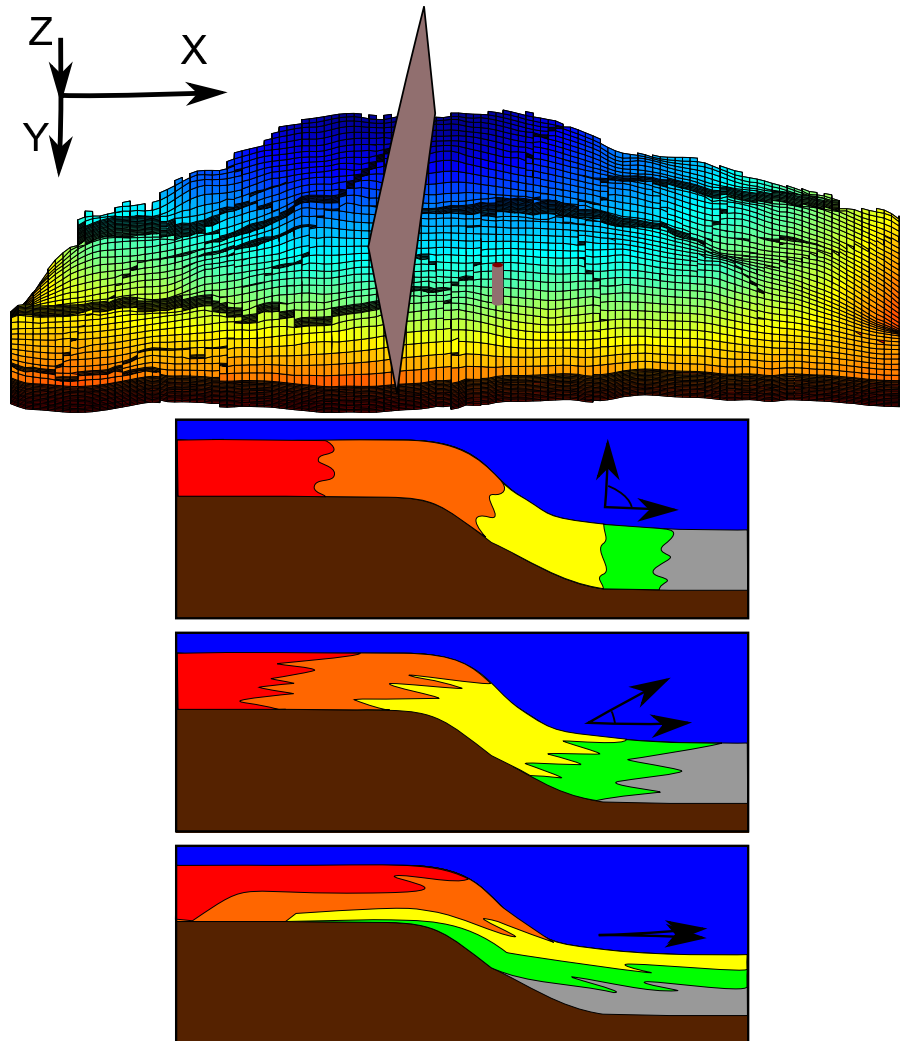

Figure 16.5.: The river flows from left to right toward the sea on the model vertical section shown here (upper figure). Aggradation angle is demonstrated in three levels (bottom figure); from top: low, medium and high aggradation angle (Ashraf et al., 2013). Between deposition and now, the entire system was rotated by tectonic effects such that the original river flow direction is oriented upward, not downward.

transition zone between river and basin: the amount of deposition supplied by the river and the accommodation space that the sea provides for these depositional masses. Deposition happens in a spectrum from larger grains depositing earlier on the land side, to fine deposits happening in the deep basin. If the river flux or sea level fluctuates, equilibrium changes into a new bedding shape based on the balance of these factors. In the SAIGUP study, progradational cases are considered, in which river flux increases and shifts the whole depositional system into the sea. The angle at which transitional deposits are stacked on each other because of this shifting is called the aggradation angle. Three levels of aggradation are shown in Figure 16.5: low, medium and high. The study reported in (Ashraf et al., 2010a,b) showed that aggradation can have a dramatic influence on the injection and migration process.

Fault transmissibility: Huge uncertainties can be involved when modeling the presence of faults. Faults are discrete objects that are modeled by changing the geometry of the simulation grid. The transmissibility for flow across faults changes during the process of faulting. This causes a spectrum of transmissibilities, from a sealing fault with no flow across it, to a fault that has not produced any barriers to the flow within its opening space.

Within a simulation grid, the influence of faults on the local and global flow behavior depends on a number of parameters including fault length, orientation, intensity and transmissibility. The well location with respect to the faults can change the overall behavior of injected CO$_2$ plume significantly. In the SAIGUP models, different levels of fault orientations, transmissibility, areal intensity, and well patterns are considered. For this study, we consider all fault
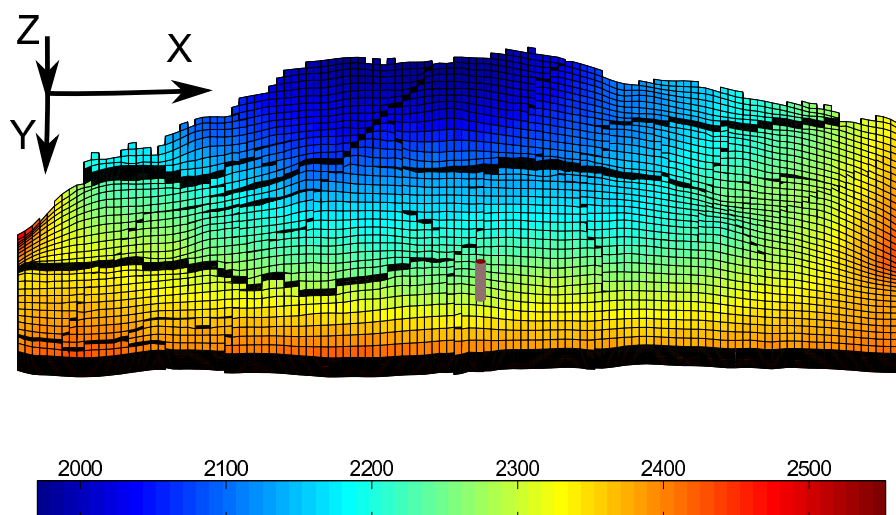

Figure 16.6.: Fault orientation and intensity of the model used in the study (Ashraf et al., 2013). Depth in meter is shown by color on the grid.

modeling parameters at their medium level and consider to vary only the fault transmissibility. These variations, however, do not affect the definition of the no-flow boundary, which is motivated by the presence of an impermeable fault.

The used geology realizations contain compartmentalized fault systems comprising approximately equal densities of strike-parallel and strike-perpendicular faults based on a portion of the Gullfaks field (Manzocchi et al., 2008b; Howell et al., 2008). Figure 16.6 shows the fault pattern and location of the injector considered for the study.

It is shown in (Manzocchi et al., 1999) that the transmissibility multiplier provides a numerically more robust representation of faults within reservoir simulation than conventional permeability multipliers. We consider the fault transmissibility multipliers to range between zero and one. A multiplier value of one corresponds to a fault permeability equal to the harmonic average of cell permeabilities across the fault, i.e., to a fault without any influence on flow (Manzocchi et al., 1999).

Regional groundwater effect: Geological modeling always comes with the uncertainty of how large the aquifer is and how it is connected to other underground aquifers. This is a direct consequence of the need to define boundary conditions to limit the computational domain, which cannot always coincide with meaningful physical boundaries in large-scale systems. However, connections to active external aquifers can be accounted for by adapting the values for the boundary conditions accordingly. Some connections might even change throughout the year, depending on rainfall. The flux across model boundaries might influence the CO$_2$ plume dynamics during and after injection. To simulate such effects, we changed the left boundary pressure by adding an uncertain additional pressure value Δp that varies between 0 and 100 bars.

As a scenario assumption, this pressure value is added at the start of injection, i.e., the pressure distribution is not at a steady state when the simulation starts, and this triggers a corresponding transient brine flow. We do so in order to analyze the effect of transient groundwater effects on the system. This may seem an arbitrary choice, but assuming a steady-state would also be arbitrary to some extent.

The overall process for sensitivity analysis, uncertainty propagation, and risk assessment starts by specifying probability information for the uncertain parameters. Next, one has to design and choose the simulation cases required to obtain the expansion coefficients in the approximating polynomial. However, in our study, we had access to the set of SAIGUP geological realizations and simulation results that had been designed without the considerations

possible with the aPC. The computing time for each SAIGUP realization was about 2 hours on a 2.4GHz Intel Xeon CPU, and we decided to recycle these highly expensive simulations in our study. The large computing times are a key motivation to build a cheaper surrogate model for further analysis. Hence, we assume the histograms of uncertain parameters such that they result in collocation points that coincide with the SAIGUP designed values. Therefore, the histograms used in this study are almost uniform, as shown in Figure 16.7. In fact, these input distributions could also be handled with the gPC method already mentioned in the introduction, and would correspond to the use of Legendre polynomials. In our case, we use the aPC to avoid the step of modeling the input distributions as exactly uniform. Consequently, the polynomials resulting from the aPC approach are very close to Legendre polynomials. The aPC, however, could be used for any type of histograms and so provides the freedom in other studies to adapt to arbitrary input statistics.

The main concern here is not a unique probability description of the input geological parameters, but rather we perform an uncertainty analysis practice, relying on a scenario assumption of probability distributions. Thus, no general geological conclusion is expected from this study, and results might change by feeding the work-flow with a different probability description.

16.4. Sensitivity analysis

In this section, we tackle global sensitivity analysis with Sobol indices based on the aPC technique, following the line of work on aPC by (Oladyshkin and Nowak, 2012a; Oladyshkin et al., 2011a,d). The big advantage of global aPC-based sensitivity analysis is that one can obtain global sensitivity information at computational costs that are hardly larger than those for local analysis. The reason is the following: local methods use infinitesimally small spacing between parameter sets for model evaluation to get numerical derivatives evaluated at a single point. The aPC-based method places the parameter sets for model evaluation at an optimized spacing in parameter space. This can be interpreted as fitting secants (or polynomials for non-linear analysis) to the model response. These secants (polynomials) approximate the model over the entire parameter space in a weighted least-square sense (compare with the best unbiased ensemble linearization approach described by (Nowak, 2009b)). This is more beneficial compared to computing a tangent or local second derivatives (compare FORM, SORM methods, e.g., (Jang et al., 1994)) that approximate the model well just around one point in the parameter space.

(a) Barriers, percentage.

(b) Aggradation angle, degree.

(c) Fault transmissibility multiplier,-.

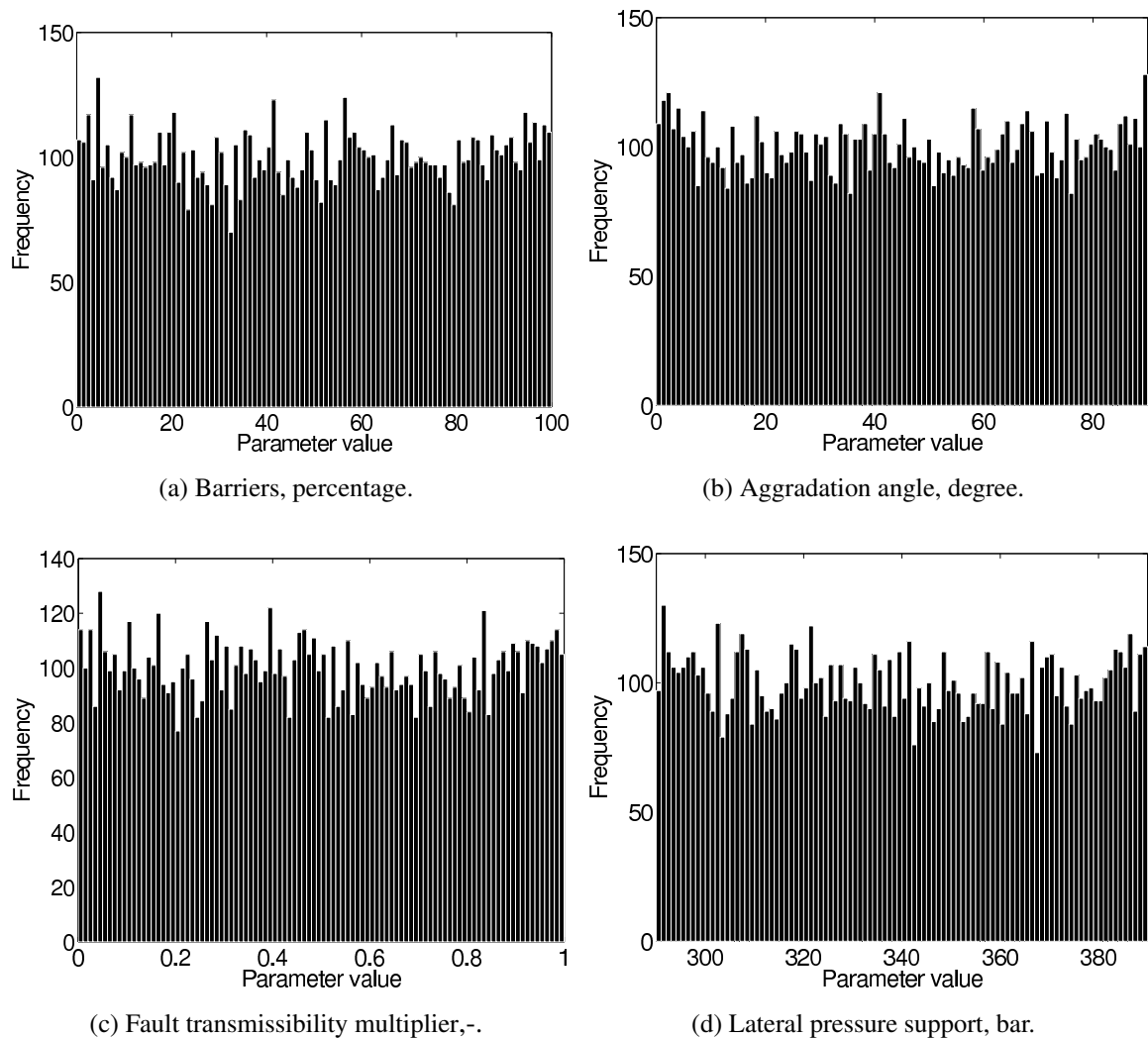(d) Lateral pressure support, bar.

Figure 16.7.: The histograms of geological variables used in this study are sampled from uniform distributions.

The system featured here is non-linear due to two reasons: First, the involved multi-phase flow equations (Oladyshkin et al., 2011a) form a coupled system of non-linear partial differential equations, and second, these equations are non-linear in their coefficients. The latter is even more significant if parameters are spatially heterogeneous.

In the following, we briefly summarize the Sobol sensitivity indices technique for quantifying the relative importance of each individual input parameter in the final prediction. Then, we implement the method for our geological CO_2 storage problem, based on the aPC response surface.

The model responses featured here for global sensitivity analysis (this section) and for the probabilistic risk analysis (see Section 16.5) are listed in Table 2 and have been discussed in Section 16.3.1. In the sense of global sensitivity analysis (Saltelli et al., 2008), not only should the analysis technique be global, but also should the analyzed quantities be global. In the latter, global refers to the fact that they are relevant for the engineer, are crucial in decision processes, etc. For example, an overall leakage risk is more informative in final decisions than the leakage rate at a specific point, and a total stored volume of CO_2 is more informative for volumetric efficiency considerations of the reservoir than the CO_2 saturation at individual points.

16.4.1. Sobol sensitivity indices

The method is well described in the literature (Sobol, 2001; Saltelli et al., 2008; Saltelli, 2004; Reuter and Liebscher, 2008). More recent works are concerned about expediting calculation pace by computing Sobol indices analytically from polynomial chaos expansions (Crestaux et al., 2009; Oladyshkin et al., 2011a,d; Le Maître and Knio, 2010; Sudret, 2008). The idea behind the combination of PCE techniques with Sobol indices is to replace the analyzed system with an approximating function which leads to mathematical and numerical benefits in the sensitivity analysis.

Using polynomials for this approximation is convenient, because it is easy to analytically obtain the output variances from the statistics of the input variables of the polynomials. In our case, the solution is approximated by orthogonal polynomials with ascending polynomial degree. We expand the variance of model output into individual components originating from all possible combinations of input parameters. Assume that we break the system output into components as follows:

$$\Gamma = \Gamma_0 + \sum_i \Gamma_i + \sum_i \sum_{j>i} \Gamma_{ij} + ... \tag{16.8}$$

A single index (here: i) shows dependency to a specific input variable. More than one index (e.g.: i and j) shows interaction of two or more input variables. If we consider the input vector Θ to have n components θ_i for $i = 1,..,n$, then $\Gamma_i = f_i(\theta_i)$ and $\Gamma_{ij} = f_{ij}(\theta_i, \theta_j)$. In practice, we stop at a finite number of terms in Eq. (16.8). The first order sensitivity index, the so called Sobol index, is defined statistically as follows (Saltelli et al., 2008):

$$S_i = \frac{\mathrm{V}[\mathrm{E}(\Gamma \mid \theta_i)]}{\mathrm{V}(\Gamma)}, \quad (16.9)$$

where $\mathrm{E}(\Gamma \mid \theta_i)$ is the conditional expectation of output Γ for a given value of θ_i and V is the variance operator. In plain words, S_i is the fraction of total variance $V(\Gamma)$ that can be explained by the parameter θ_i. Since θ_i can be fixed at any value in its uncertainty interval, each of those values produces a distinct expectation. In Eq. (16.9), the variance of those expectations is divided by the unconditional variance of output (i.e., with no input variable fixed). For more than one index, a higher-order Sobol index can be defined as:

$$S_{ij} = \frac{\mathrm{V}[\mathrm{E}(\Gamma \mid \theta_i, \theta_j)] - \mathrm{V}[\mathrm{E}(\Gamma \mid \theta_i)] - \mathrm{V}[\mathrm{E}(\Gamma \mid \theta_j)]}{\mathrm{V}(\Gamma)}. \quad (16.10)$$

Here, $\mathrm{V}[\mathrm{E}(\Gamma \mid \theta_i, \theta_j)]$ is the variance of output expectations after fixing θ_i and θ_j. This index represents the significance of variation in output generated from the joint uncertainty in several input variables, i.e., from the interaction of uncertain parameters. If we add all indices that contain a given variable θ_i, the sum is called the total Sobol index:

$$S_{Ti} = S_i + \sum_{j \neq i} S_{ij} + \sum_{j \neq i} \sum_{k \neq i} S_{ijk} + \dots \quad (16.11)$$

The total Sobol index is a sensitivity measure to rank parameters according to their influence on the model results. When this index is close to zero, the corresponding parameter has a negligible role in the variation of the system response. In that case, the uncertainty in that parameter does not introduce a considerable uncertainty in the response, and the parameter could be omitted from further analyses.

In practice, we evaluate the Sobol indices analytically from the expansion coefficients of the aPC as described by (Oladyshkin et al., 2011d).

16.4.2. Sensitivity analysis

We calculate the total Sobol indices for the geological CO_2 storage problem that is described earlier. The results are based on an aPC expansion of order two that is obtained by fifteen detailed simulations. The choice of order two is supported by the results of (Oladyshkin et al., 2011b), where the authors found in a similar CO_2 storage problem that second order may be the cheapest non-linear expansion, but still sufficiently accurate for this type of purpose. Recently, (Oladyshkin et al., 2011d) provided the results of a numerical convergence analysis

for aPC-based Sobol analysis. They report that increasing the expansion order beyond 2 introduces only small changes to the sensitivity values for their considered system, and does not change the ranking of the analyzed parameters anymore. A study similar to the current study without aPC needed one hundred and sixty runs to perform a sensitivity analysis with a different method (Ashraf et al., 2010b). The pattern of sensitivity reported here is similar to what is produced in that study, but at dramatically reduced costs.

16.4.3. Results

The flow behavior in the domain is influenced by the type and intensity of different heterogeneities. This influence can be traced in the CO$_2$ pressure and saturation distributions over time. During injection, viscous forces imposed by the injector dominate the force balance. Viscous forces act in the form of spatial pressure gradients in all directions. After 30 years, the injection stops, and gravity starts playing the major role in the flow dynamics, acting in the vertical direction (Ashraf et al., 2010a,b).

Barriers and aggradation angle have different impacts on the flow during each flow regime, i.e., injection or after injection. Low fault transmissibility hinders the flow and keeps the pressure in compartments. Geometry distortion in the geological layers because of the faulting processes plays a considerable role in the splitting of CO$_2$ plumes within the domain. Water flux from lateral boundaries due to the regional groundwater effect enhances the spread of CO$_2$ and leads the mass of CO$_2$ toward the other open boundaries.

Figure 16.8 shows the sensitivity of different responses to the uncertain parameters. Total Sobol indices are plotted at specific times. End of simulation refers to the year 100, i.e., 70 years after injection stops. This time duration is long enough for the flow to stabilize at a stationary condition for the majority of the model runs.

As already observed in (Ashraf et al., 2010a,b), the aggradation angle plays a significant role in the flow behavior. In cases with low aggradation angle, the stratigraphy of rock types is a pattern of parallel layering. For higher aggradation angles, rock-types are distributed between more modeling layers. The effective vertical permeability changes from the harmonic average (in Figure 16.9a) toward the arithmetic average (in Figure 16.9c), as the aggradation angle increases from 0 to 90 degrees. The harmonic average might be much smaller than the arithmetic average, in particular when there are vertically impermeable rock-types in the medium. The shallow marine depositional system contains some rock-types with almost zero transmissibility in the vertical direction. Therefore, a low aggradation angle can hinder the
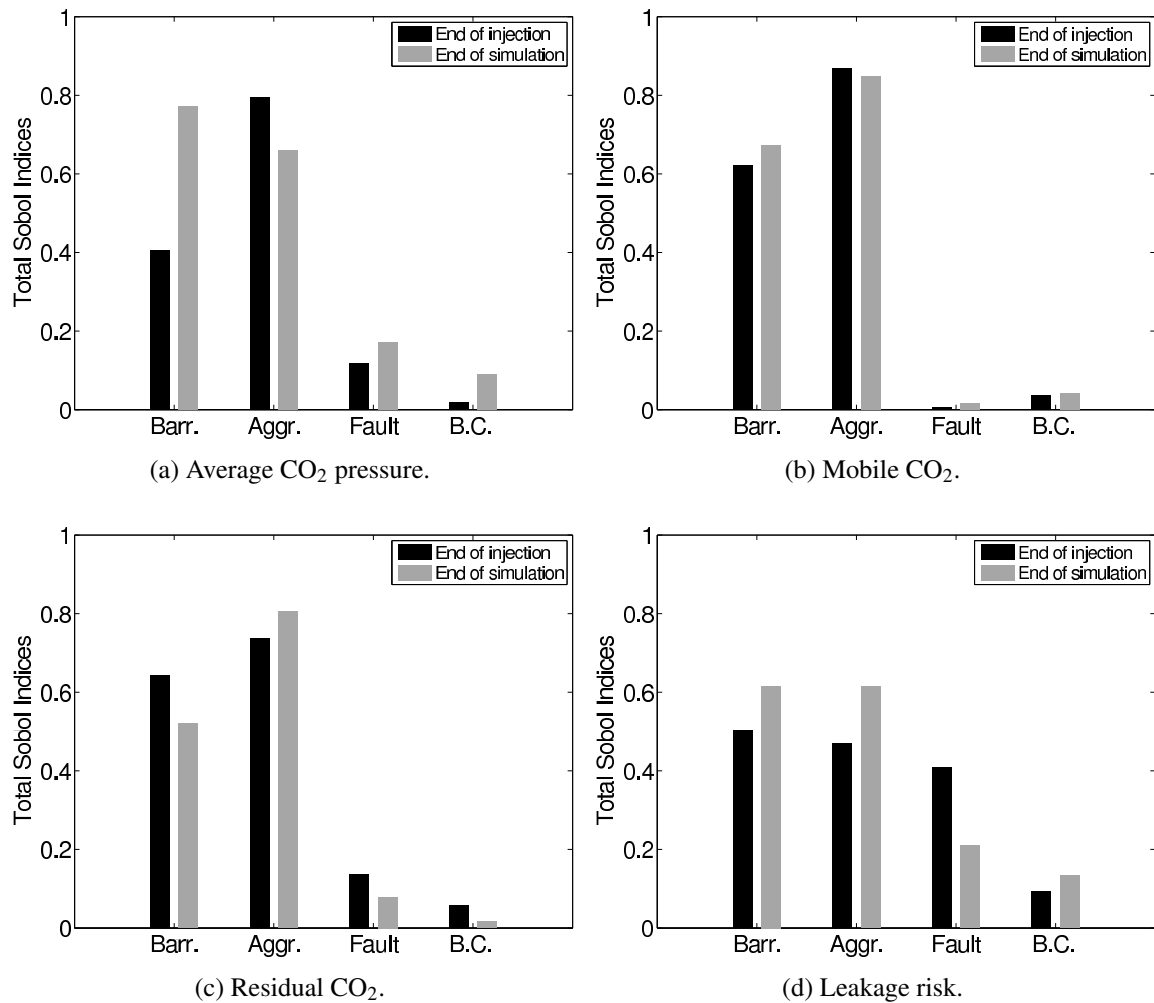
(a) Average CO_2 pressure.

(b) Mobile CO_2.

(c) Residual CO_2.

(d) Leakage risk.

Figure 16.8.: Sensitivity analysis for different responses (a: average CO_2 pressure, b: mobile CO_2, c: residual CO_2, and d: leakage risk) with respect to the uncertain parameters. In the figures above, Barr. is for barriers, Aggr. for aggradation angle, Fault for fault transmissibility, and B.C. for regional groundwater effect.

flow from traveling upward across layers in the domain and force it to stay trapped in some lower layers, as seen for many of the low aggradation angle realizations in our study. The relatively large sensitivities to the level of barrier presence is based on the same effects.

Our results show a relatively weak sensitivity of responses with respect to the water influx from one side of the model. This sensitivity is in particular low during injection, when the high pressure imposed from the well dominates the dynamics of flow in the medium (Figure 16.8a). The sensitivity patterns for the mobile and residual CO_2 volume are similar in Figures 16.8b and 16.8c, because the mobile and residual CO_2 volume add up to the total injected

(a) Low aggradation angle.   (b) Medium aggradation angle.
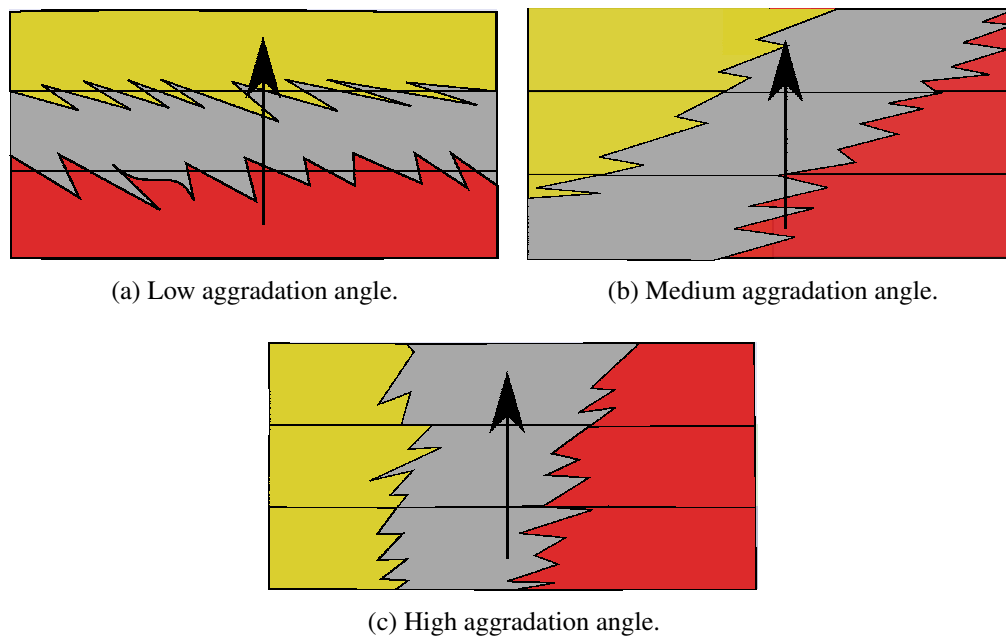

(c) High aggradation angle.

Figure 16.9.: Illustration of how the aggradation angle affects the effective vertical conductivity (Ashraf et al., 2013).

CO_2 volume, with the exception of the CO_2 volume that has left the domain. Hence, they are highly dependent on each other.

More detailed results are shown in Figures 16.10a to 16.10d. Total Sobol indices are plotted for each response during the entire time interval. When the flow regime switches from injection to a gravity-dominated system, we observe a jump or sharp drop in some of the sensitivity plots (Figures 16.10a and 16.10c at 30 years).

The sensitivity of the CO_2 pressure with respect to the presence of barriers jumps up, right after stopping the injection. This happens because barriers slow down the pressure release through open boundaries, resulting in local pressure build-ups.

The sensitivity of the residual CO_2 volume with respect to barriers presence drops soon after injection. This is reasonable since the residual trapped volume is more a function of lateral flow in the medium, compared to the vertical flow in the relatively small thickness of the aquifer.

(a) CO_2 pressure.

(b) Mobile CO_2 volume.
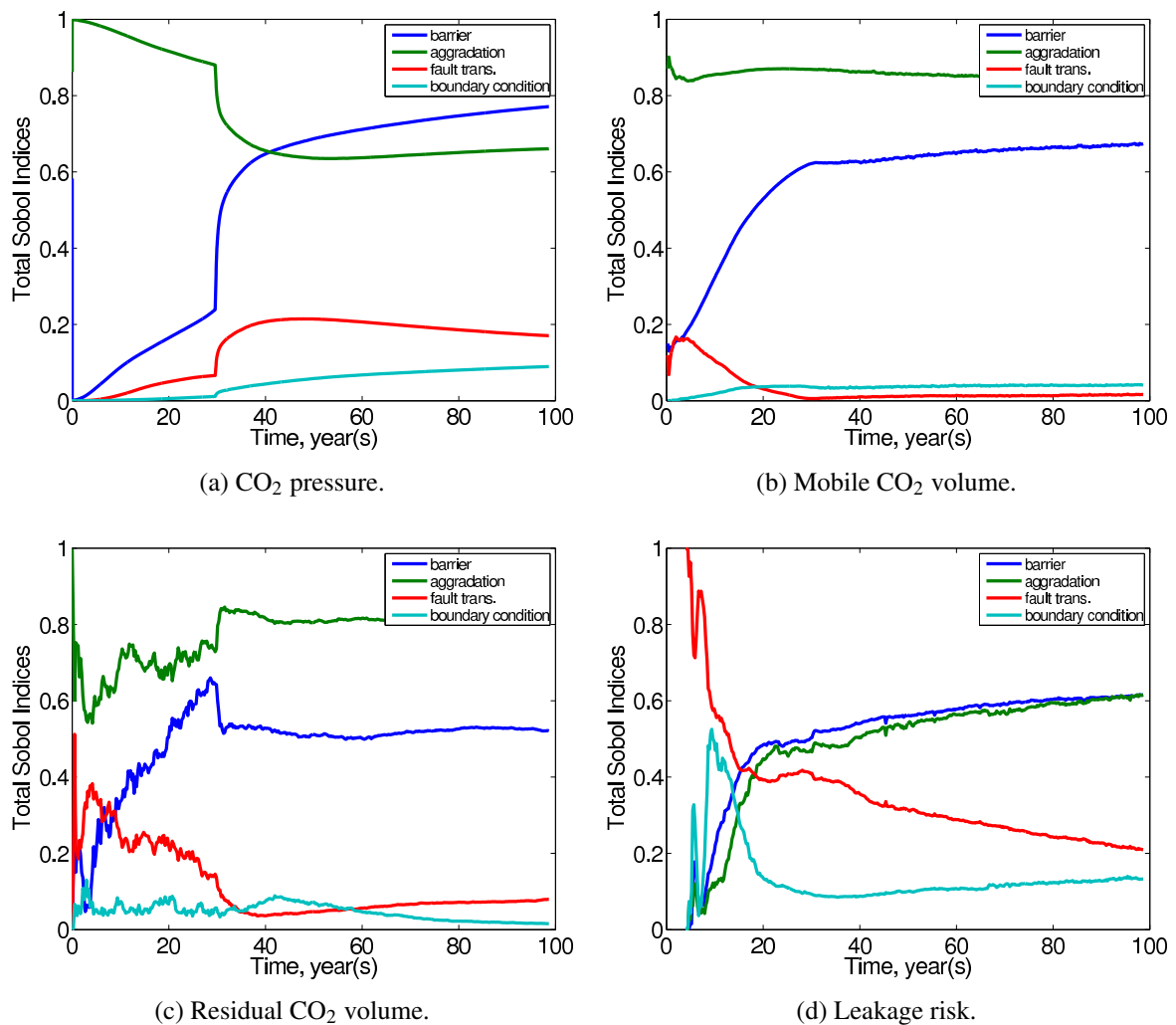
(c) Residual CO_2 volume.

(d) Leakage risk.

Figure 16.10.: Sensitivities (expressed by total Sobol indices) plotted versus time for different responses.

16.5. Risk analysis

The risk R of a process is quantitatively defined as the extent of consequence C caused by the process, multiplied by the probability P of that consequence to happen:

$$R = P \times C. \tag{16.12}$$

The consequence can be defined by direct measures in the simulation responses, or it can be related to consequences caused in the environment outside the considered system. For example, in the case of CO_2 injection into deep aquifers, the amount of CO_2 which stays mobile and undissolved in the medium for a time after injection can be considered as a consequence,

bearing the potential of leakage up to the surface if exposed to a geological leakage point. The consequence could also be defined by a criterion for external consequences, like the rate of climate change (either locally or globally) due to CO_2 leakage, the costs of pumping CO_2 that does not remain in the subsurface, or via the related costs for CO_2 emission certificates.

The other part is the probability of these consequences to happen. This depends on the stochastic behavior of the process which results in the respective outcomes.

We use the polynomial-based reduced model for risk analysis, because it is fast enough to perform a Monte-Carlo analysis with a large number (here: 10000) of realizations on the polynomials. Thanks to the higher-order approximation via the aPC, the principal non-linear physical behavior of CO_2 storage is included in the analysis, and detailed probabilistic risk assessment becomes feasible. We analyze here the same quantities as in Section 16.4, i.e., average CO_2 pressure, the volume of mobile or immobile CO_2, and leakage risk. For definitions, see Section 16.3.1.

16.5.1. Quantification of expected values in CO_2 storage

Average response values can be calculated analytically from the polynomial (e.g., (Oladyshkin et al., 2011a)) or via the Monte-Carlo post-process as mentioned above. Figures 16.11a to 16.11d show some of the calculated expectations as functions of time. In Figure 16.11a, the mobile CO_2 volume increases linearly in the medium because of the constant injection rate during the first thirty years. After injection, the mobile volume of CO_2 is reduced due to the trapped volume in residual form and the migration of CO_2 across open boundaries.

Figure 16.11b shows the expected values for the volume of residually trapped CO_2 as a function of time. The plot shows the significance of imbibition during the plume migration period, when water replaces CO_2 that is moving upward because of gravity segregation. During injection, CO_2 invades the aquifer and drainage is dominant. Therefore, the expected residual CO_2 plot shows a smaller slope during injection than what it shows later in time.

When injection starts, a pressure pulse travels through the medium at a finite velocity because of the slight compressibility of brine. The initially built-up pressure releases through open boundaries over time and the average pressure drops in the aquifer (Figure 16.11c). The large pressure build-up in the very early time steps occurs because large pressure values have to be exceeded in the injection cell before CO2 becomes mobile at saturations above the
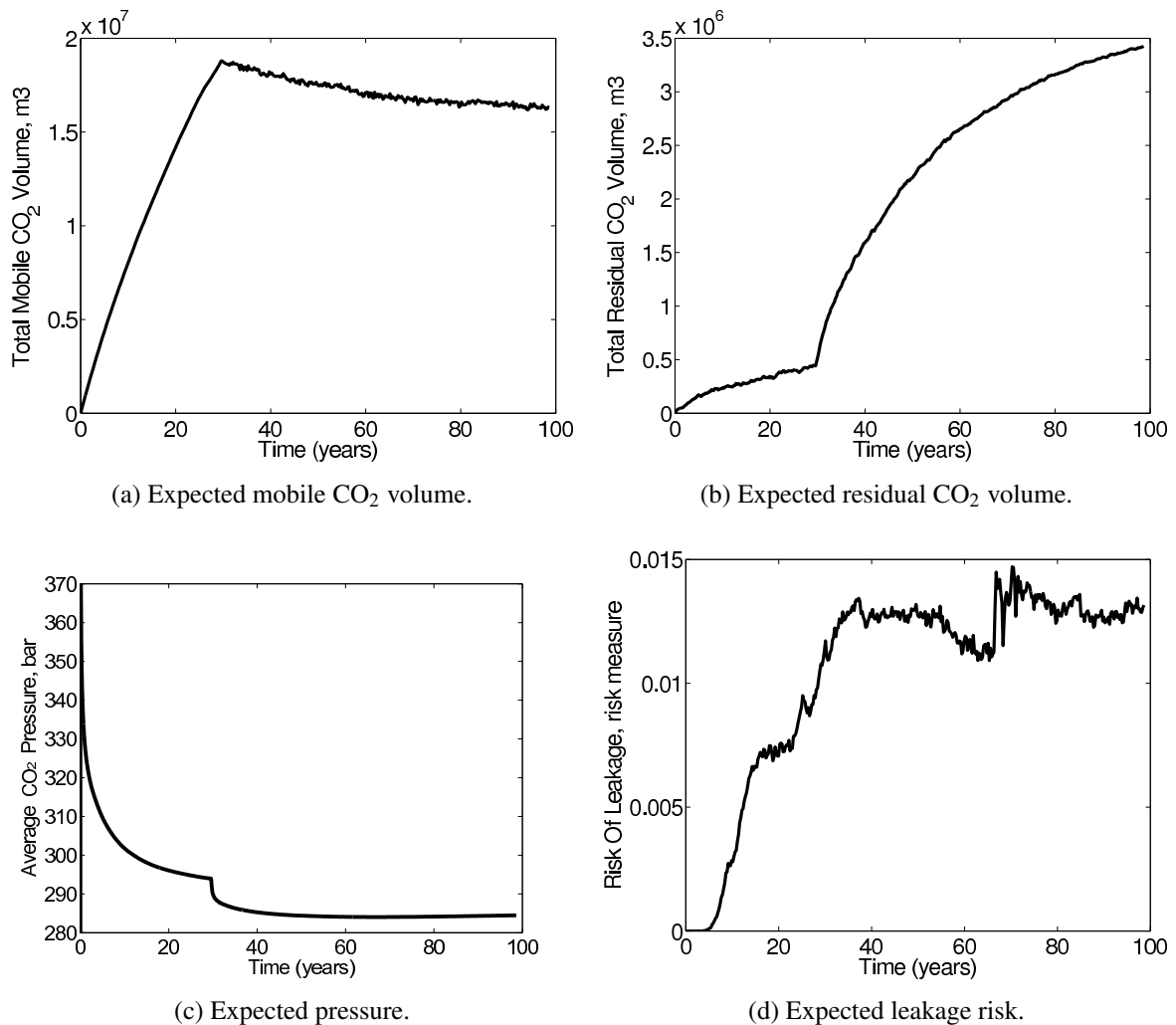
(a) Expected mobile CO_2 volume.

(b) Expected residual CO_2 volume.

(c) Expected pressure.

(d) Expected leakage risk.

Figure 16.11.: Expectation for response values versus time. The presssure value for initial time step in Figure c goes up to 670 bars.

residual value. During this period, the CO2 pressure is defined almost only by the pressure in the injection cell (compare the definition of CO2 pressure in Section 3.2). Under realistic injection settings, a pressure rise of up to 400 bars (from 270 to 670 bars in the first simulation time step, not visible in Figure 16.11c) would be very unrealistic and would not be allowed to occur. At the end of Section 16.5.2, we will investigate this issue in more detail.

Also, during early injection time, the pressure is larger than at the end of injection. There are a few realizations where the contributions from the external aquifer support, a dense barrier system close to 100% areal coverage, an adverse aggradation angle of the formation and extremely low fault transmissibilities interact to effectively block the CO_2 flow close to the

well. This has strong effects on pressure when the rock at the injector position happens to be poorly permeable, leading to a very poor injectivity. An adapted CO$_2$ injection strategy would react by lowering the injection rate, by choosing a different injection position, or by even abandoning the entire site.

Based on the results of the current study, it is possible to identify such adverse combinations and guide site investigation strategies to pay attention to such situations. In a follow-up study (ready for submission), we are currently investigating an active injection strategy controlled by an upper allowable pressure limit.

However, the initial sharp pressure increase is released very quickly. This happens, when first parts of the CO$_2$ plume have found flow pathways into regions with better rock properties, providing the possibility to relax the pressure build-up, and also to let the CO$_2$ escape towards the boundaries.

The expected leakage risk is plotted in Figure 16.11d, and increases in value as the injected CO$_2$ travels upward and accumulates beneath the sealing cap-rock.

16.5.2. Results of CO$_2$ storage risk assessment

In this section, the probability distributions (rather than expected values) of system responses during and after injection are studied. Results from the MC analysis of the response surface are given as histograms of output values and also as cumulative distribution functions (CDF) for probabilities (Figures 16.12 and 16.13).

Figures 16.12a to 16.12c show the histograms of responses obtained from the Monte-Carlo process at the end of injection. A long tail is observed for lower residual and mobile CO$_2$ values in Figures 16.12b and 16.12c. The long tail means a large range of possible low values. Pressure shows a long tail for higher values. This means that even high critical values still have substantial probabilities to be exceeded, indicating that the possibility of geomechanical damage to sealing layers will have to receive a large attention. We observe an issue of mass conservation in Figure 16.12b, where a few realizations show more mobile CO$_2$ in the domain than the total injected volume (which is about 40×10^6 m^2). This is a typical issue for a large class of statistical methods that interpolate or extrapolate simulation results in the parameter space, because their setup is not based on the mass conservation equation. In this specific case, the mass conservation issue is caused by approximating the response surface via polynomials, with vanishing residuals only at the collocation points.

(a) Average CO_2 pressure.

(b) Mobile CO_2 volume.

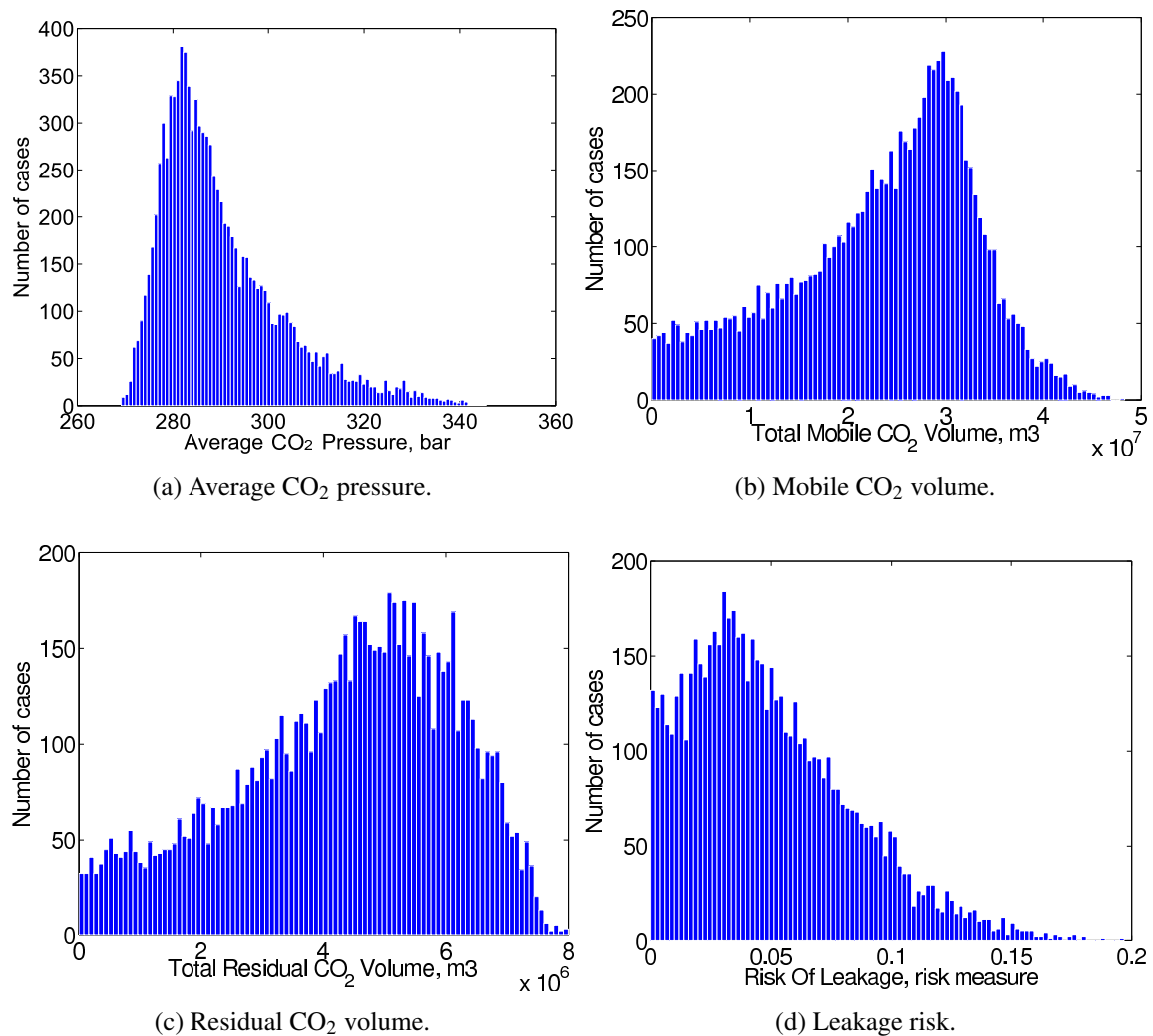(c) Residual CO_2 volume.

(d) Leakage risk.

Figure 16.12.: Histograms of selected response values at end of injection.

The polynomials are evaluated at many randomly chosen parameter sets drawn from the histograms shown in Figure 16.7, which do not coincide with the collocation points.

Finally, we report how the corresponding probabilities change over time in Figures 16.13a to 16.13c. High pressure buildup is considerable during the early injection time, and it is negligible after injection during plume migration (Figure 16.13a). An over-pressurized injection can induce fracturing in the medium, extending to the sealing layers. Any fractures caused in the structural traps can expose the mobile CO_2 to leakage paths. Therefore, higher pressure values can be interpreted as high risk in early time.

The presented framework for risk assessment indicates that the pressure in the reservoir is unacceptably large (see Section 16.5.1 and Figure 16.11c) and can be too high. In the

(a) Average CO$_2$ pressure.

(b) Mobile CO$_2$ volume.

(c) Residual CO$_2$ volume.


(d) Leakage risk over time.
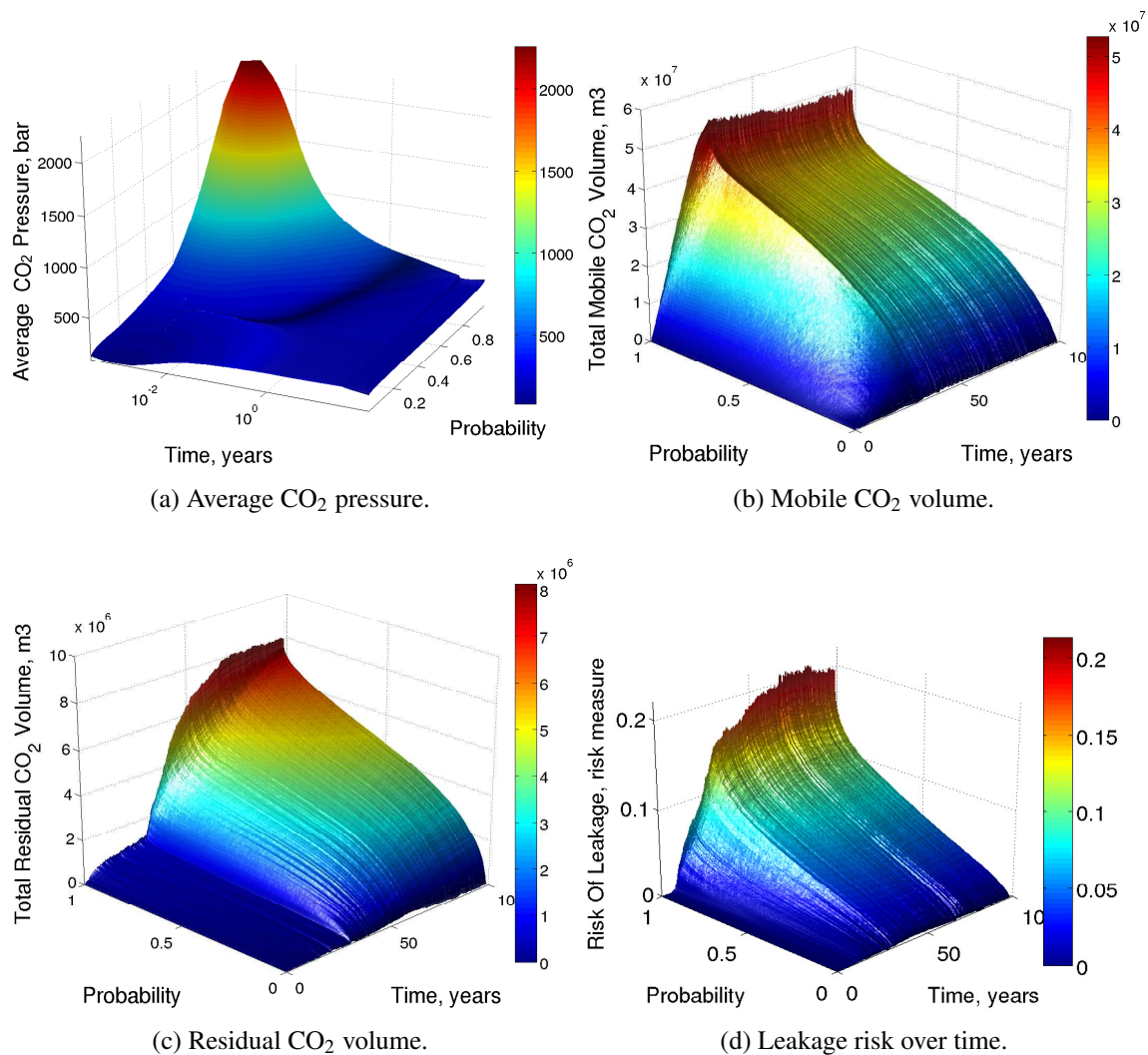
Figure 16.13.: Evolution of the cumulative distribution function of different response values over time.

following, we use our method to investigate this critical issue. Figures 16.14a and 16.14b show the predicted time evolution of field-average CO$_2$ pressure for a collocation point with adverse and well-suitable values of the aggradation angle, respectively. It becomes apparent, that the unacceptable pressure values arise only under extreme values of the aggradation angle. Reacting to this insight, we see and discuss two possible options in the following.

The first option is to lower the injection rate, so that we keep the pressure values in a safe region, even under the probability that the reservoir might have an adverse aggradation angle. Figure 16.15a shows the injection rate for a safe scenario: the injection rate ramps up over a year from zero to one fifth of the level used before. The corresponding pressure behavior for
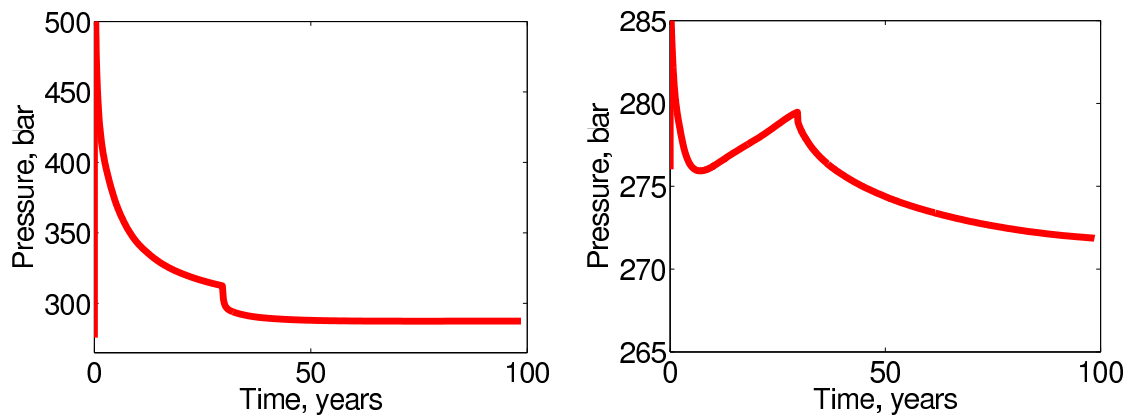
(a) Average CO_2 pressure for an extreme aggradation case in the original (high) injection rate scenario.

(b) Average CO_2 pressure for an average case in the original (high) injection rate scenario.
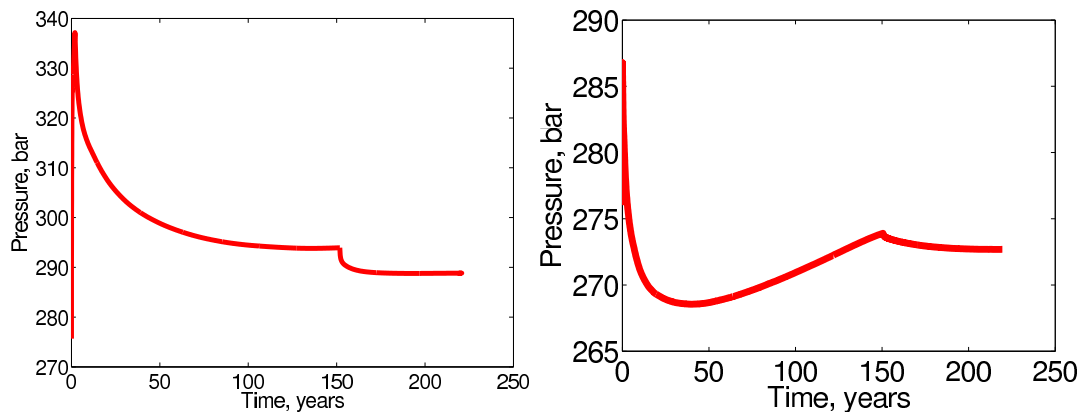
(c) Average CO_2 pressure for the extreme aggradation case in an adapted (lower) injection rate scenario.

(d) Average CO_2 pressure for an average case in the adapted (lower) injection rate scenario.

Figure 16.14.: Expectation for response values versus time. The presssure value for initial time step in Figure c goes up to 670 bars.

the adverse and the well-suitable cases is shown in Figures 16.14c and 16.14d. Please note that the case with the less extreme aggradation angle shows the typical rise in CO_2 pressure up to the end of injection (Figure 16.14b and 16.14d), following just after the initial pressure peak due to first entry.

The second option is to improve our understanding about the properties of the analyzed storage site. In particular, some additional exploration actions could help to reduce the uncertainty in the aggradation angle. In the previous analysis we considered that all values of aggradation angle between zero and ninety degrees are equiprobable, which is a very con-
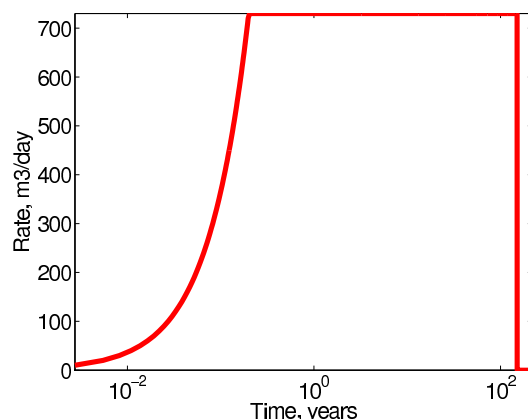
Figure 16.15.: Adapted injection rate scenario for safer conditions.

servative assumption on the initial state of knowledge. As a scenario variation, we will now assume that further exploration decreased the probability of the extreme aggradation values. Figure 16.16a shows the modified assumption on the aggradation distribution, where the extreme values have low probability values in comparison to the initial assumption. The present aPC framework allows estimating the influence of such an uncertainty reduction onto the model output without expensive computational costs. Technically, the Monte-Carlo process can be performed on the response surface under the new assumption on uncertainty. Figures 16.16b to 16.16c show the new expected field-average CO_2 pressure and the histogram of average CO_2 pressure at the end of injection, respectively. The new pressure statistics indicate a feasible reservoir operation, even with the original (large) injection rate.

16.6. Conclusions

In this chapter, we used the arbitrary polynomial chaos expansion (aPC) method in a sensitivity analysis and risk assessment process. The goal was to demonstrate the application and feasibility of aPC-based methods in the context of realistic CO_2 injection scenarios. We implemented this method for a typical CO_2 storage problem. Four uncertain parameters with assumed uncertainty distributions are considered. Injection and early migration of CO_2 is studied. The flow sensitivity to geological heterogeneity is evaluated and quantified using Sobol indices. Risk analysis is performed on the defined problem. Flow dynamics are discussed and corresponding interpretations and explanations of the sensitivity and risk results are provided.

The performance of the aPC method has been satisfactory. It is very fast, compared to other
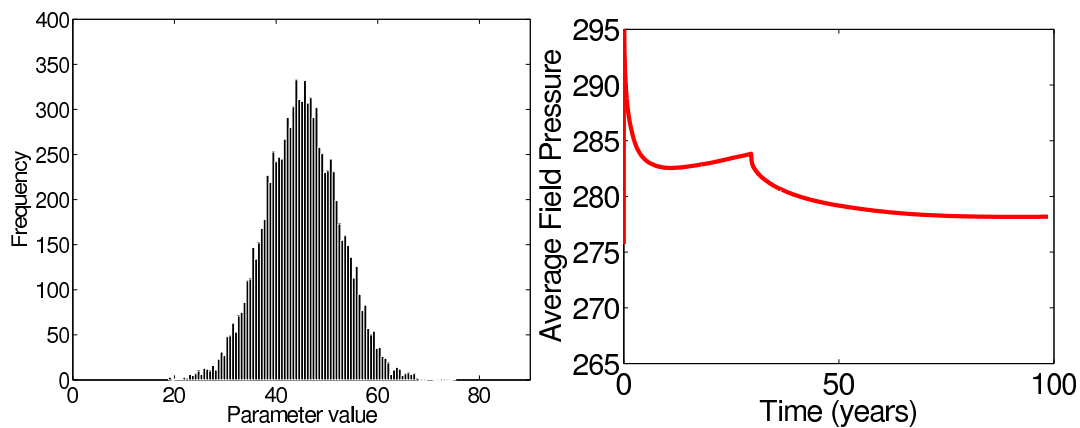
(a) Histogram of aggradation angle filtered out for extreme values.

(b) Expected average CO_2 pressure.
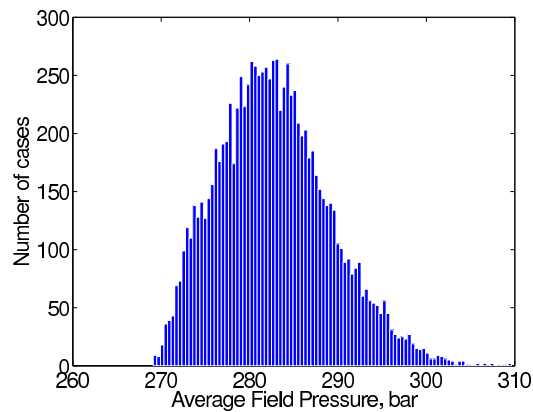
(c) Histogram of average CO_2 pressures.

Figure 16.16.: Extreme aggradation angle values can result in impractical injection operations. Filtering out the extreme aggradation cases (e.g., by geophysical screening) leads to more favourable conditions. (a): more narrow distribution of aggradation. (b): New expected value for CO_2 field-average pressure under more narrow aggradation range. (c): New distribution of CO_2 field-average pressure at end of injection.

stochastic methods for low-parametric systems, and this speed-up allows us to perform an extensive Monte-Carlo process on the aPC-based response surface to calculate the probability of response values throughout simulation time. This study was a first-time application of the aPC to study a realistically complex type of geological structural uncertainty. Based on our assessment of aPC feasibility, we can strongly encourage the use of aPC for sensitivity and risk analysis in complex situations.

The results have shown that the most influential parameter for most of the responses is the aggradation angle of deposition layers of the considered shallow-marine aquifer. The least relevant parameter is the regional groundwater effect, especially during injection time. We re-iterate that the aim of this study was to demonstrate a practice of using arbitrary polynomial chaos expansion for the sensitivity and risk analysis of a typical CO$_2$ storage problem. Since, in general, the levels of involved input uncertainty are not unique, the physical and geological conclusions of this study are restricted to the probability assumptions taken here and should not be generalized to systems that are very different.

Part V.

Summary and Conclusions

The proposed thesis presented research in the field of environmental modeling in the face of complexity and uncertainty. The thesis demonstrated achievements for the construction of reliable and feasible models that can adequately describe physical concepts and, at the same time, account for uncertainty. The achievements were subdivided into three principal parts.

The first part focused on complex physical concepts and their mathematical description. It offered several possibilities to accelerate the modeling process. Chapter 7 presented an effective framework based on the streamline technique for multiphase multi-component flows in large-scale heterogeneous petroleum reservoirs. Chapter 6 showed an efficient handling of thermodynamic properties of a compressible compositional two-phase flow in porous media which demands an extremely short time of simulation. Chapter 8 offered a calibrated analytical form of hydrogen solubility in water to accelerate modeling of hydrogen migration around a storage site of radioactive waste. Additionally, the first part of the thesis emphasized the challenge of finding a healthy and reasonable compromise between numerical techniques, conceptual physical formulations, estimation of model parameters and their uncertainties, computational effort, etc. For this purpose, Chapter 5 briefly discussed different sources of possible error within an entire modeling challenge and illustrated their possible impact using an application to carbon dioxide storage.

The second part was dedicated to the development of efficient model reduction methodologies for uncertainty quantification. Chapter 9 focused on the development of arbitrary polynomial chaos that can adapt to arbitrary probability distribution shapes of input parameters and provides improved convergence in comparison to classical chaos expansion techniques. Chapter 10 provided a flexible and efficient framework for global sensitivity analysis for quantification of the effects of modeling parameters on the overall model uncertainty. Chapter 11 proposed an advanced framework for model calibration and history matching based on the arbitrary polynomial chaos expansion and strict Bayesian principles. This combination offered a statistical method for history matching that is accurate, yet has a computational speed that is more than sufficient for real-time application.

The third part demonstrated the potential of uncertainty quantification in the overall modeling procedure using an application to the storage of energy relevant gases in geological formations and discussed related challenges. Chapter 12 discussed an integrative approach to robust design under uncertainty and probabilistic risk assessment for gas storage in geological formations. It showed that the probability of failure and negative impacts is an interplay between controllable engineering aspects and uncertain aspects of the system. Chapter 13 analyzed the pressure increasing in the storage which might result in caprock failure or

brine displacement. Chapter 14 presented a minimally subjective approach for uncertainty quantification based on a new and purely data-driven version of polynomial chaos expansion. Chapter 15 distinguished different kinds of uncertainties associated with the applied models and estimated the risk of brine discharge into freshwater aquifers due to CO_2 injection into geological formations. At the end of the third part in Chapter 16 the influence of uncertain parameters controlling structural heterogeneities on flow of CO_2 in a realistic largeŰscale storage was investigated.

To conclude, it should be mentioned that every specific environmental system is unique and requires an individual and careful treatment. However, the overall modeling process of environmental systems often demands tasks such as prediction of possible scenarios, design of planned engineering actions, estimation of risks, calibration of the model using available data etc. Moreover, the modeling process in the field of environmental engineering is a chain of many tasks. Like any chain, it is only as strong as its weakest link. Ignoring or neglecting, for the sake of computational efficiency, of some of the components or processes in the system could be very dangerous. Thus, an adequate combination of a reliable physical model with advanced stochastic tools offers a powerful base for the forecasting of environmental behavior.

Bibliography

Aanonsen, S., Naevdal, G., Oliver, D., Reynolds, A., and Vallès, B. The ensemble Kalman filter in reservoir engineering – a review. *SPE Journal*, 14(3):393–412, 2009.

Aarnes, J. On the use of a mixed multiscale finite element method for greaterflexibility and increased speed or improved accuracy in reservoir simulation. *Multiscale Modeling & Simulation*, 2(3):421–439, 2004.

Abramowitz, M. and Stegun, I., A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1965.

Aker, E., Skurtveit, E., Grande, L., Cuisiat, F., Johnsen, Ø., Soldal, M., and Bohloli, B. Experimental methods for characterization of cap rock properties for CO_2 storage. *Multiphysical Testing of Soils and Shales*, pages 303–308, 2013.

Anderson, M. and Burt, T. P. Modelling strategies. *in Anderson, M.G. and Burt, T.P. (eds.), Hydrological Forecasting*, pages 1–13, 1985.

Andra. *Dossier Argile 2005: Les recherches de l'Andra sur le stockage géologique des déchets radioactifs à haute activité et à vie longue*. Collection les Rapports, Andra, Châtenay-Malabry, France, 2005.

Ashraf, M., Lie, K., Nilsen, H., Nordbotten, J., and Skorstad, A. Impact of geological heterogeneity on early-stage CO_2 plume migration. In *CMWR*, 2010a.

Ashraf, M., Lie, K.-A., Nilsen, H. M., and Skorstad, A. Impact of geological heterogeneity on early-stage CO_2 plume migration: sensitivity study. In *Proceedings of ECMOR XII*, 2010b.

Ashraf, M., Oladyshkin, S., and Nowak, W. Geological storage of CO2: global sensitivity analysis and risk assessment using arbitrary polynomial chaos expansion. *International Journal of Greenhouse Gas Control*, 19:704–719, 2013. doi: 10.1016/j.ijggc.2013.03.023.

Askey, R. and Wilson, J. *Some basic hypergeometric polynomials that generalize Jacobi polynomials*. Memoirs of the American Mathematical Society, AMS, Providence, 1985.

Augustin, F., Gilg, A., Paffrath, M., Rentrop, P., and Wever, U. Polynomial chaos for the approximation of uncertainties: Chances and limits. *Euro. J. of Applied Mathematics*, 19: 149–190, 2008.

Aziz, K. and Settari, A. *Petroleum reservoir simulation*, volume 476. Applied Science Publishers London, 1979.

Babuska, I., Nobile, F., and Tempone, R. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45 (3):1005–1034, 2007.

Bachu, S. Screening and ranking of sedimentary basins for sequestration of CO_2 in geological media in response to climate change. *Environmental Geology*, 44(3):277–289, 2003. ISSN 0943-0105.

Bachu, S. CO_2 storage in geological media: Role, means, status and barriers to deployment. *Progress in Energy and Combustion Science*, 34(2):254–273, 2008.

Bangerth, W., Klie, H., Wheeler, M., Stoffa, P., and Sen, M. On optimization algorithms for the reservoir oil well placement problem. *Computational Geosciences*, 10(3):303–319, 2006.

Barthelmann, V., Novak, E., and Ritter, K. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, 12(4):273–288, 2000.

Bastian, P., Blatt, M., Dedner, A., Engwer, C., Klöfkorn, R., Kornhuber, R., Ohlberger, M., and Sander, O. A generic grid interface for parallel and adaptive scientific computing. part ii: Implementation and tests in dune. *Computing*, 82(2):121–138, 2008.

Batalin, O., A.I., B., and Zaharov, M. Y. *Phase equilibrium in systems of natural hydrocarbons*, volume 272. Nedra, Moskow, 1992.

Batycky, R. *A three-dimensional two-phase field scale streamline simulator*. PhD thesis, stanford university, 1997.

Batycky, R., Blunt, M., and Thiele, M. a 3d multi-phase streamline simulator with gravity and changing well conditions. *Proceedings of the 17th International Energy Agency*

Collaborative Project on Enhanced Oil Recovery, Sydney, Australia (September 1996), 1996.

Bellin, A., Salandin, P., and Rinaldo, A. Simulation of dispersion in heterogeneous porous formations: Statistics, first-order theories, convergence of computations. *Water Resour. Res.*, 28(9):2211–2227, 1992.

Benekos, I., Shoemaker, C., and Stedinger, J. Probabilistic risk and uncertainty analysis for bioremediation of four chlorinated ethenes in groundwater. *Stochastic Environmental Research and Risk Assessment*, 21(4):375–390, 2007. ISSN 1436-3240.

Benson, S., Hepple, M. R., Apps, J., Tsang, C., and Lippmann., M. Lessons learned from natural and industrial analogues for storage of carbon dioxide in deep geological formations. *LBNL-51170. Lawrence Berkeley National Laboratory, Berkeley, CA.*, 2002.

Bielinski, A. *Numerical simulation of CO_2 sequestration in geological formations*. PhD thesis, Institut für Wasserbau, Universität Stuttgart, 2006.

Binning, P. J. and Celia, M. A. Pseudokinetics arising from the upscaling of geochemical equilibrium. *Water Resour. Res.*, 44(7):W07410–, July 2008. ISSN 0043-1397.

Birkholzer, J. and Zhou, Q. Basin-scale hydrogeologic impacts of co< sub> 2</sub> storage: Capacity and regulatory implications. *International Journal of Greenhouse Gas Control*, 3(6):745–756, 2009.

Birkholzer, J. T., Zhou, Q., and Tsang, C.-F. Large-scale impact of CO_2 storage in deep saline aquifers: A sensitivity study on pressure response in stratified systems. *Int. J. of Greenhouse Gas Control*, 3:181–194, 2009.

Blatman, G. and Sudret, B. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering and System Safety*, 95:1216–1229, 2010.

Bliznyuk, N., Ruppert, N., Shoemaker, C., A., Regis, R., G., Wild, S., M., and Mugunthan, P. Bayesian calibration of computationally expensive models using optimization and radial basis function approximation. *Journal of Computational and Graphical Statistics*, 17(2): 1–25, 2008. doi: 10.1198/106186008X320681.

Brennan, S., Burruss, R., Merrill, M., Freeman, P., and Ruppert, L. A probabilistic assessment methodology for the evaluation of geologic carbon dioxide storage. *US Geological Survey Open-File Report*, 1127:31, 2010.

Buzzard, G. Global sensitivity analysis using sparse grid interpolation and polynomial chaos. *Reliability Engineering & System Safety*, 2011.

Caflisch, R. Monte carlo and quasi-monte carlo methods. *Acta numerica*, 1998:1–49, 1998.

Cameron, R. and Martin, W. The orthogonal development of nonlinear functionals in series of fourier-hermite functionals. *Ann. Math.*, 48(2):385–392, 1947.

Celia, M., Bachu, S., Nordbotten, J., Gasda, S., and Dahle, H. Quantitative estimation of CO_2 leakage from geological storage: Analytical models, numerical models and data needs. In *Proceedings of 7th International Conference on Greenhouse Gas Control Technologies.(GHGT-7)*, 2004.

Chen, J.-S., Wang, L., Hu, H.-Y., and Chi, S.-W. Subdomain radial basis collocation method for heterogeneous media. *Int. J. Numer. Meth. Engng*, 80(2):163–190, 2009.

Chopra, A. K. and D., C. R. Proof of the two-phase steady-state theory for flow through porous media. *SPE Formation Evolution*, December:603–608, 1986.

Christie, M. and Clifford, P. A fast procedure for upscaling in compositional simulation. In *SPE Reservoir Simulation Symposium*, 1997.

Class, H., Ebigbo, A., Helmig, R., Dahle, H., Nordbotten, J. N., Celia, M. A., Audigane, P., Darcis, M., Ennis-King, J., Fan, Y., Flemisch, B., Gasda, S., Jin, M., Krug, S., Labregere, D., Naderi, A., Pawar, R. J., Sbai, A., Sunil, G. T., Trenty, L., and Wei, L. A benchmark-study on problems related to CO_2 storage in geologic formations. *Computational Geosciences*, 13:451–467, 2009.

Coats, K. An equation of state compositional model. *Old SPE Journal*, 20(5):363–376, 1980.

Cominelli, A., Ferdinandi, F., de Montleau, P., and Rossi, R. Using gradients to refine parameterization in field-case history-matching projects. *SPE Reserv. Evalu. Eng.*, 10(3): 233–240, 2007.

Commission, E. *European Commission's Communication on Extended Impact Assessment, Brussels*. Office for Official Publications of the European Communities, 2002.

Comunian, A. and Renard, P. Introducing wwhypda: a world-wide collaborative hydrogeological parameters database. *Hydrogeology Journal*, 17(2):481–489, 2009.

Cortis, A., Oldenburg, C., and Benson, S. M. The role of optimality in characterizing CO_2 seepage from geologic carbon sequestration sites. *Int. J. of Greenhouse Gas Control*, 2: 640–652, 2008.

Crestaux, T., Le Maitre, O., and Martinez, J.-M. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering and System Safety*, 94(7):1161–1172, 2009.

Dagan, G. Theory of solute transport by groundwater. *Annual Review of Fluid Mechanics*, 19:183–215, 1987.

Dagan, G. Time-dependent macrodispersion for solute transport in anisotropic heterogeneous aquifers. *Water Resour. Res.*, 24(9):1491–1500, 1988.

Dagan, G. *Flow and Transport in Porous Formations*. Springer Verlag, Berlin, 1989.

Danesh, A. *PVT and phase behaviour of petroleum reservoir fluids*, volume 47. Elsevier Science, 1998.

Daoud, A. M. I. *Automatic History Matching in Bayesian Framework for Field-Scale Applications*. Texas A&M University, 2004.

Darcis, M., Class, H., and Flemisch, B. Coupling models of different complexity for the simulation of co_2 storage in saline aquifers. *Energy Procedia*, 1(1):1767–1774, 2009.

Darcis, M., Class, H., Flemisch, B., and Helmig, R. Sequential model coupling for feasibility studies of co2 storage in deep saline aquifers. *Oil & Gas Science and Technology–Revue dIFP Energies nouvelles*, 66(1):93–103, 2011.

Dargahi-Noubary, G. and Razzaghi, M. Bootstrap construction of the upper confidence limit for unreliability. *Reliability Engineering & System Safety*, 37(1):1–6, 1992.

de Barros, F. P. J. and Rubin, Y. A risk-driven approach for subsurface site characterization. *Water Resources Research*, 44(1):W01414, 2008. ISSN 0043-1397.

de Barros, F. P. J., Rubin, Y., and Maxwell, R. The concept of comparative information yield curves and its application to risk-based site characterization. *Water Resources Research*, 45(6):W06401, 2009. ISSN 0043-1397.

de Barros, F. P. J., Ezzedine, S., and Rubin, Y. Impact of hydrogeological data on measures of uncertainty, site characterization and environmental performance metrics. *Advances in Water Resources*, 2011. doi: doi:10.1016/j.advwatres.2011.05.004.

Der Kiureghian, A. and Liu, P. Structural reliability under incomplete probability information. *Journal of Engineering Mechanics*, 112(1):85–104, 1986.

Dinariev, O. Y. About stability of stationary gas-condensate filtrational flow of mixture. *Appl. Math. Mech.*, 60:768–777, 1996.

Ditlevsen, O. and Madsen, H. *Structural reliability methods*, volume 384. John Wiley & Sons, New York, 1992.

Dutton, S., Flanders, W., and Barton, M. Reservoir characterization of a Permian deep-water sandstone, East Ford field, Delaware basin, Texas. *AAPG bulletin*, 87(4):609–628, 2003. ISSN 0149-1423.

Eaton, T. On the importance of geological heterogeneity for flow simulation. *Sedimentary Geology*, 184(3-4):187–201, 2006. ISSN 0037-0738.

Ebigbo, A., Class, H., and Helmig, R. CO_2 leakage through an abandoned well: problem-oriented benchmarks. *Computational Geosciences*, 11(2):103–115, 2007.

Efron, B. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial Mathematics, 1987.

Efron, B. and Tibshirani, R. J. *An Introduction to the Bootstrap (Monographs on Statistics & Applied Probability)*. Chapman & Hall, 2010.

Eigestad, G., Dahle, H., Hellevang, B., Riis, F., Johansen, W., and Øian, E. Geological modeling and simulation of CO_2 injection in the Johansen formation. *Computational Geosciences*, 13(4):435–450, 2009.

Er, G. A method for multi-parameter pdf estimation of random variables. *Structural safety*, 20(1):25–36, 1998.

Ernst, O., Mugler, A., Starkloff, H., and Ullmann, E. On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(02):317–339, 2012.

Evans, M., Hastings, N., and Peacock, B. Statistical distributions. *Statistical Distributions. Wiley-Interscience, 3 edition*, page 221, 200.

Evensen, G. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2006.

Ewing, R., Pilant, M., Wade, G., and Watson, A. Estimating parameters in scientific computation. *IEEE Comput. Sci. Eng.*, 1(3):19–31, 1994.

Fajraoui, N., Ramasomanana, F., Younes, A., Mara, T., Ackerer, P., and Guadagnini, A. Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media. *Water Resources Research*, 47(2):W02521, 2011.

Feraille, M. and Marrel, A. Prediction under uncertainty on a mature field. *Oil & Gas Science and Technology–Revue dIFP Energies nouvelles*, 67(2):193–206, 2012.

Fetel, E. *Quantification des incertitudes liées aux simulation d'écoulement dans un réservoir pétrolier à l'aide de surfaces de réponse non liéaires.* PhD thesis, Institut National Polytechnique de Lorraine, 2007.

Firoozabadi, A. *Thermodynamics of hydrocarbon reservoirs*, volume 2. McGraw-Hill New York, 1999.

Flemisch, B., J., F., Helmig, R., Niessner, J., and Wohlmuth, B. DuMuX: a multi-scale multi-physics toolbox for flow and transport processes in porous media. *In A. Ibrahimbegovic and F. Dias, editors*, ECCOMAS Thematic Conference on Multi-scale Computational Methods for Solids and Fluids, Cachan, France, November 28–30, 2007, 2007.

Flemisch, B., Darcis, M., Erbertseder, K., Faigle, B., Lauser, A., Mosthaf, K., Müthing, S., Nuske, P., Tatomir, A., Wolff, M., and Helmig, R. Dumux: Dune for multi-phase, component, scale, physics, ... flow and transport in porous media. *Advances in Water Resources*, 34 (9):1102–1112, 2011. doi: DOI:10.1016/j.advwatres.2011.03.007.

Foglia, L., Mehl, S., W., Hill, M., C., Perona, P., and Burlando, P. Testing alternative ground water models using cross validation and other methods. *Ground Wate*, 45(5):627–641, 2007.

Foo, J. and Karniadakis, G. Multi-element probabilistic collocation method in high dimensions. *Journal of Computational Physics*, 229(5):1536–1557, 2010. ISSN 0021-9991.

Förster, A., Norden, B., Zinck-Jørgensen, K., Frykman, P., Kulenkampff, J., Spangenberg, E., Erzinger, J., Zimmer, M., Kopp, J., Borm, G., et al. Baseline characterization of the CO2SINK geological storage site at Ketzin, Germany. *Environmental Geosciences*, 13(3): 145–161, 2006.

Gao, G. and Reynolds, A. An improved implementation of the LBFGS algorithm for automatic history matching. *SPE Journal*, 11(1):5–17, 2006.

Gao, G. and Reynolds, A. A stochastic optimization algorithm for automatic history matching. *SPE Journal*, 12(2):196–208, 2007.

Gavalas, G., Shah, P., and Seinfeld, J. Reservoir history matching by Bayesian estimation. *SPE Journal*, 16(6):337–350, 1976.

Gerstner, T. and Griebel, M. Dimension–adaptive tensor–product quadrature. *Computing*, 71(1):65–87, 2003.

Ghanem, R. and Doostan, A. On the construction and analysis of stochastic models: Characterization and propagation of the errors associated with limited data. *Journal of Computational Physics*, 217:63–81, 2006.

Ghanem, R. and Spanos, P. A stochastic Galerkin expansion for nonlinear random vibration analysis. *Probabilistic Engineering Mechanics*, 8:255–264, 1993.

Ghanem, R. and Spanos, P. D. Polynomial chaos in stochastic finite elements. *Journal of Applied Mechanics*, 57:197–202, 1990.

Ghanem, R. G. and Spanos, P. D. *Stochastic finite elements: A spectral approach*. Springer-Verlag, New York, 1991.

Ghomian, Y., Pope, G., and Sepehrnoori, K. Development of a response surface based model for minimum miscibility pressure (mmp) correlation of co2 flooding. In *SPE Annual Technical Conference and Exhibition*, 2008.

Gilks, W., Richardson, S., and Spiegelhalter, D. *Markov chain Monte Carlo in practice*. Chapman & Hall, 1996.

Grigoriu, M. *Stochastic Calculus: Applications in Science and Engineering*. Birkhauser, Boston, 2002.

Gu, Y. and Oliver, D. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE J.*, 12(4):438–446, 2007.

Gunn, R., Chueh, P., and Prausnitz, J. Inversion temperatures and pressures for cryogenic gases and their mixtures. *Cryogenics*, 6(6):324–329, 1966.

Hansson, A. and Bryngelsson, M. Expert opinions on carbon dioxide capture and storage: A framing of uncertainties and possibilities. *Energy Policy*, 37:2273–2282, 2009a.

Hansson, A. and Bryngelsson, M. Expert opinions on carbon dioxide capture and storage–a framing of uncertainties and possibilities. *Energy Policy*, 37(6):2273–2282, 2009b.

Haro Sandoval, E., Anstett-Collin, F., and Basset, M. Sensitivity study of dynamic systems using polynomial chaos. *Reliability Engineering & System Safety*, 104:15–26, 2012.

He, J., Sarma, P., Durlofsky, L., and Chen, W. Use of reduced-order models for improved data assimilation within an EnKF context. *presented in Reservoir Simulation Symposium, the Woodlands, Texas, USA*, SPE 141967, 2011.

Helton, J. and Davis, F. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69, 2003.

Helton, J. C. Treatment of uncertainty in performance assessments for complex systems. *Risk Analysis*, 14(4):483–511, 1994. ISSN 1539-6924. doi: 10.1111/j.1539-6924.1994. tb00266.x.

Henderson, G. D., Danesh, A., Tehrani, D., and Peden, J. M. The effect of velocity and interfacial tension on the relative permeability of gas condensate fluids in the wellbore region. In *8th European IOR Symposium, May 15-17, Vienna*, 1995.

Hendricks Franssen, H.-J. and Kinzelbach, W. Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems. *J. Hydrol.*, 365(3–4):261–274, 2009.

Hoffman, F. O. and Hammonds, J. S. Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis*, 14(5):707–712, 1994. ISSN 1539-6924. doi: 10.1111/j. 1539-6924.1994.tb00281.x.

Holtz, M. Residual gas saturation to aquifer influx: A calculation method for 3-d computer reservoir model construction. *SPE Gas Technology Symposium*, 2002.

Homma, T. and Saltelli, A. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, 52(1):1–17, 1996.

Howell, J., Skorstad, A., MacDonald, A., Fordham, A., Flint, S., Fjellvoll, B., and Manzocchi, T. Sedimentological parameterization of shallow-marine reservoirs. *Petroleum Geoscience*, 14(1):17–34, 2008. ISSN 1354-0793.

Huang, S., Mahadevan, S., and Rebba, R. Collocation-based stochastic finite element analysis for random field problems. *Probabilistic engineering mechanics*, 22(2):194–205, 2007.

IPCC. *Special report on carbon dioxide capture and storage, Technical report, Intergovernmental Panel on Climate Change (IPCC), prepared by Working Group III*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2005.

Isukapalli, S., S., Roy, A., and Georgopoulos, P., G. Stochastic response surface methods (srsms) for uncertainty propagation: Application to environmental and biological systems. *Risk Analysis*, 18(3):351–363, 1998.

Jakeman, J., D. and Roberts, S., G. Stochastic Galerkin and collocation methods for quantifying uncertainty in differential equations: a review. *Australian & New Zealand Industrial and Applied Mathematics Journal*, 50:815–830, 2008.

Jang, Y., S., Sitar, N., and Kiureghian, A., D. Reliability analysis of contaminant transport in saturated porous media. *Water Resources Research*, 30(8):2435–2448, 1994.

Jaynes, E., T. On the rationale of maximum-entropy methods. *Proc. IEEE*, 106:939–952, 1982.

Jin, B. Fast Bayesian approach for parameter estimation. *Int. J. Numer. Meth. Engng.*, 76: 230–252, 2008.

Kaplan, S. The words of risk analysis. *Risk Analysis*, 17 (4):407–417, 1997.

Karlin, S. *Total Positivity: Vol.: 1*. Stanford University Press, 1968.

Keese, A. and Matthies, H. G. Sparse quadrature as an alternative to mc for stochastic finite element techniques. *Proc. Appl. Math. Mech.*, 3:493–494, 2003.

Kitanidis, P. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Research*, 22(4):499–507, 1986. ISSN 0043-1397.

Kitanidis, P. K. Quasi-linear geostatistical theory for inversing. *Water Resources Research*, 31(10):2411–2419, 1995.

Koldoba, A. V. and Koldoba, E. V. The discontinuous solutions of the transport equations for compositional flow in porous media. *Transport in porous media*, 52(2):267–277, 2003.

Kopp, A., Class, H., and Helmig, H. Investigations on CO_2 storage capacity in saline aquifers - part 1: dimensional analysis of flow processes and reservoir characteristics. *Int. J. of Greenhouse Gas Control*, 3:263–276, 2009.

Kopp, A., Binning, P., Johannsen, K., Helmig, R., and Class, H. A contribution to risk analysis for leakage through abandoned wells in geological co_2 storage. *Advances in Water Resources*, 33(8):867–879, 2010.

Korn, G. and Korn, M. Curvilinear coordinates sysmem mathematical handbook for scientist and engineers, 1968.

Kovscek, A. and Wang, Y. Geologic storage of carbon dioxide and enhanced oil recovery. i. uncertainty quantification employing a streamline based proxy for reservoir flow simulation. *Energy Conversion and Management*, 46(11):1920–1940, 2005.

Kravaris, C. and Seinfeld, J. Identification of parameters in distributed parameter-systems by regularization. *SIAM J. Control Optim.*, 23(2):217–241, 1985.

Kumar, A., Ozah, R., Noh, M., Pope, G., Bryant, S., Sepehrnoori, K., and Lake, L. Reservoir simulation of CO_2 storage in deep saline aquifers. *SPE Journal*, 10 (3):336–348, 2005.

Lacroix, S., Vassilevski, Y., Wheeler, M., and Wheeler, J. Iterative solvers of the implicit parallel accurate reservoir simulator (ipars). *Numerical Linear Algebra with Applications*, 4:537–549, 2001.

Le Maitre, O. and Knio, O. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer, 2010.

Le Maître, O. and Knio, O. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer, 2010.

Lemieux, J.-M. Review: The potential impact of underground geological storage of carbon dioxide in deep saline aquifers on shallow groundwater resources. *Hydrogeology Journal*, 19:757–778, 2011. ISSN 1431-2174.

Lenzen, M. Global warming effect of leakage from CO_2 storage. *Critical Reviews in Environmental Science and Technology*, 41(24):2169–2185, 2011.

Leube, P., Geiges, A., and Nowak, W. Bayesian assessment of the expected data impact on prediction confidence in optimal sampling design. *Water Resources Research*, 48 W02501, 2012a. doi: 10.1029/2010WR010137. submitted manuscript.

Leube, P. C., Nowak, W., and Schneider, G. Temporal moments revisited: Why there is no better way for physically-based model reduction in time. *Water Resources Research*, 2012b. accepted.

Lewicki, J., Birkholzer, J., and Tsang, C.-F. Natural and industrial analogues for leakage of CO_2 from storage reservoirs: identification of features, events, and processes and lessons learned. *Environmental Geology*, 52:457–467, 2007. ISSN 0943-0105.

Li, H. and Zhang, D. Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods. *Water Resources Research*, 43:44–48, 2007.

Li, H., Sarma, P., and Zhang, D. A comparative study of the probabilistic collocation and experimental design methods for petroleum reservoir uncertainty quantification. *SPE Journal*, pages SPE–140738–PA–P, 2011.

Li, R., Reynolds, A., and Oliver, D. Sensitivity coefficients for three-phase flow history matching. *J. Can. Pet. Technol.*, 42(4):70–77, 2003.

Lia, O., Omre, H., Tjelmeland, H., Holden, L., and Egeland, T. Uncertainties in reservoir production forecasts. *AAPG Bull*, 81(5):775–802, 1997.

Lialin, B. E. and Silniov, K. A. Concepts of mathematical modelling of a layers systems on the base of streamline method. *Neftegazovoe delo*, 2005.

Lin, G. and Tartakovsky, A. An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow and solute transport in randomly heterogeneous porous media. *Advances in Water Resources*, 32(5):712–722, 2009.

Lindsay, B. On the determinants of moment matrices. *The Annals of Statistics*, pages 711–721, 1989.

Liou, C.-L. and Lin, C.-H. Applications of the methods of weighted residuals in system science. *International Journal of Systems Science*, 22(9):1509–1525, 1991.

Liu, N. and Oliver, D. Ensemble Kalman filter for automatic history matching of geologic facies. *Journal of Petroleum Science and Engineering*, 47:147–161, 2005.

Lumley, J. The structure of inhomogeneous turbulent flows. *In Atmospheric Turbulence and Wave Propagation, ed. A. M. Yaglom*, pages 166–178, 1967.

Makhlouf, E., Chen, W.H., Wasserman, M., and Seinfeld, J. A general history matching algorithm for three-phase, three-dimensional petroleum reservoirs. *SPE Adv. Technol.*, 1 (2):83–91, 1993.

Maltz, F. H. and Hitzl, D. L. Variance reduction in Monte Carlo computations using multi-dimensional Hermite polynomials. *Journal of Computational Physics*, 2:345–376, 1979.

Manzocchi, T., Walsh, J., Nell, P., and Yielding, G. Fault transmissibility multipliers for flow simulation models. *Petroleum Geoscience*, 5(1):53–63, 1999.

Manzocchi, T., Carter, J., Skorstad, A., Fjellvoll, B., Stephen, K., Howell, J., Matthews, J., Walsh, J., Nepveu, M., Bos, C., et al. Sensitivity of the impact of geological uncertainty on production from faulted and unfaulted shallow-marine oil reservoirs: objectives and methods. *Petroleum Geoscience*, 14(1):3–11, 2008a. ISSN 1354-0793.

Manzocchi, T., Matthews, J., Strand, J., Carter, J., Skorstad, A., Howell, J., Stephen, K., and Walsh, J. A study of the structural controls on oil recovery from shallow-marine reservoirs. *Petroleum Geoscience*, 14(1):55–70, 2008b.

Marzouk, Y., Najm, H., and Rahn, L. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comput. Phys.*, 224(2):560–586, 2007.

Matthews, J., Carter, J., Stephen, K., Zimmerman, R., Skorstad, A., Manzocchi, T., and Howell, J. Assessing the effect of geological uncertainty on recovery estimates in shallow-marine reservoirs: the application of reservoir engineering to the SAIGUP project. *Petroleum Geoscience*, 14(1):35–44, 2008. ISSN 1354-0793.

Matthies, H., G. and Keese., A. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comp. Meth. Appl. Mech. Engrg.*, 194:1295–1331, 2005.

Maul, P. R., Metcalfe, R., Pearce, J., Savage, D., and West, J. M. Performance assessments for the geological storage of carbon dioxide: Learning from the radioactive waste disposal experience. *International Journal of Greenhouse Gas Control*, 1(4):444 – 455, 2007. ISSN 1750-5836. doi: 10.1016/S1750-5836(07)00074-6.

Maxwell, R. M. and Kastenberg, W. E. Stochastic environmental risk analysis: An integrated methodology for predicting cancer risk from contaminated groundwater. *Stochastic Environmental Research and Risk Assessment*, 13(1):27–47, 1999.

Michael, K., Golab, A., Shulakova, V., Ennis-King, J., Allinson, G., Sharma, S., and Aiken, T. Geological storage of CO_2 in saline aquifers–a review of the experience from existing storage operations. *International Journal of Greenhouse Gas Control*, 4(4):659–667, 2010.

Morariu, V. I., Srinivasan, B. V., Raykar, V. C., Duraiswami, R., and Davis, L. S. Automatic online tuning for fast Gaussian summation. In *Advances in Neural Information Processing Systems (NIPS)*, 2008.

Moritz, H. Least-squares collocation. *Reviews of Geophysics and Space Physics*, 16(3): 421–430, 1978.

Morris, M., D. Morris factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

Morrison, T. and Billet, F. The solubility of non-electrolytes. part ii. the effect of variation in non-electrolyte. *Journal of the Chemical Society*, pages 3819–3822, 1952.

Naevdal, G., Johnsen, L., Aanonsen, S., and Vefring, E. Reservoir monitoring and continuous model updating using ensemble Kalman filter. *SPE Journal*, 10(1):66–74, 2005.

Naevdal, G., Hanea, R., Oliver, D. S., and Valles, B. Ensemble Kalman filter for model updating - a special issue. *Computational Geosciences*, 15(2):223–224, 2011.

Nigmatulin, R. I. Mechanics of heterogeneous media, 1987.

Nikolaevski, V. N., Bondarev, E. A., Mirkin, M., et al. Flow of hydrocarbon liquids in porous media. *Nedra, Moscow*, 1968.

Nobile, F., Tempone, R., and Webster, C. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.

Nordbotten, J., Celia, M., and Bachu, M. Injection and storage of CO_2 in deep saline aquifers: analytical solution for CO_2 plume evolution during injection. *Transport in Porous Media*, 58(3):339–360, 2005a.

Nordbotten, J., Celia, M., Bachu, S., and Dahle, H. Semianalytical solution for CO_2 leakage through an abandoned well. *Environmental Science & Technology*, 39(2):602–611, 2005b.

Nordbotten, J., Celia, M., Kavetski, D., and Bachu, S. A semi-analytical model estimating leakage associated with CO_2 storage in large-scale multi-layered geological systems with multiple leaky wells. *Environmental Science & Technology*, 43(3):743–749, 2009.

Nordbotten, J. M. and Celia, M. A. *Geological Storage of CO2: Modeling Approaches for Large-Scale Simulation*. Wiley, 2011.

Norden, B., Förster, A., Vu-Hoang, D., Marcelis, F., Springer, N., and Le Nir, I. Lithological and petrophysical core-log interpretation in CO_2sink, the european CO_2 onshore research storage and verification project. *SPE Res Eval & Eng*, 13 (2):179–192, 2010.

Nowak, W. Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. *Water Resources Research*, 45(4):W04431, 2009a.

Nowak, W. Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. *Water Resour. Res.*, 45(W04431), 2009b. doi: 10.1029/2008WR007328.

Nowak, W., de Barros, F. P. J., and Rubin, Y. Bayesian geostatistical design: Task-driven optimal site investigation when the geostatistical model is uncertain. *Water Resources Research*, 46(3):W03535, 2010. ISSN 0043-1397.

Nowak, W., Rubin, Y., and de Barros, F. A hypothesis-driven approach to optimal site investigation. *Water Resources Research*, doi:10.1029/2011WR011016, 2012.

Oladyshkin, S. and Nowak, W. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. *Reliability Engineering and System Safety*, 106:179–190, 2012a. doi: 10.1016/j.ress.2012.05.002.

Oladyshkin, S. and Nowak, W. *Polynomial response surfaces for probabilistic risk assessment and risk control via robust design. Novel Approaches and Their Applications in Risk*. InTech, 2012b. ISBN 979-953-307-765-8.

Oladyshkin, S. and Panfilov, M. Two-phase flow with phase transitions in porous media: instability of stationary solutions and a semi-stationary model. In *Third Biot Conference on Poromechanics, Norman, Oklahoma, USA*, pages 529–535, 2005.

Oladyshkin, S. and Panfilov, M. Limit thermodynamic model for compositional gas-liquid systems moving in a porous medium. *Transport in Porous Media*, 70(2):147–165, 2007a.

Oladyshkin, S. and Panfilov, M. Streamline splitting between thermodynamics and hydro-dynamics in compositional gas-liquid flow through porous media. *Comptes rendus de l'Academie des sciences Mecanique*, 335(1):7–12, 2007b.

Oladyshkin, S. and Panfilov, M. Open thermodynamic model for compressible multicomponent two-phase flow in porous media. *Journal of Petroleum Science and Engineering*, 81: 41–48, 2012.

Oladyshkin, S., Class, H., Helmig, R., and Nowak, W. Highly efficient tool for probabilistic risk assessment of ccs joint with injection design. *Computational Methods in Water Resources (CMWR), XVIII International Conference on Water Resources. 21. - 24. Juni 2010, Barcelona, Spain*, 2010.

Oladyshkin, S., Class, H., Helmig, R., and Nowak, W. A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. *Advances in Water Resources*, 34:1508–1518, 2011a. doi: 10.1016/j.advwatres.2011.08.005.

Oladyshkin, S., Class, H., Helmig, R., and Nowak, W. An integrative approach to robust design and probabilistic risk assessment for CO_2 storage in geological formations. *Computational Geosciences*, 15(3):565–577, 2011b. doi: 10.1007/s10596-011-9224-8.

Oladyshkin, S., Class, H., Helmig, R., Nowak, W., de Barros, F. P. J., and Ashraf, M. Data-driven polynomial response surfaces as efficient tool for applied tasks under uncertainty. In *SAMSI Geosciences Applications Opening Workshop 2011*, 2011c.

Oladyshkin, S., de Barros, F. P. J., and Nowak, W. Global sensitivity analysis: a flexible and efficient framework with an example from stochastic hydrogeology. *Advances in Water Resources*, 37:10–22, 2011d.

Oladyshkin, S., Class, H., and Nowak, W. Bayesian updating via Bootstrap filtering combined with data-driven polynomial chaos expansions: methodology and application to history matching for carbon dioxide storage in geological formations. *Computational Geosciences*, 17(4):671–687, 2013a. doi: 10.1007/s10596-013-9350-6.

Oladyshkin, S., Schroeder, P., Class, H., and Nowak, W. Chaos expansion based Bootstrap filter to calibrate CO2 injection models. *Energy Procedia*, (40):398–407, 2013b. doi: 10.1016/j.egypro.2013.08.046.

Oldenburg, C. Screening and ranking framework for geologic CO_2 storage site selection on the basis of health, safety, and environmental risk. *Environmental Geology*, 54:1687–1694, 2008. ISSN 0943-0105.

Oldenburg, C. M., Bryant, S. L., and Nicot, J.-P. Certification framework based on effective trapping for geologic carbon sequestration. *International Journal of Greenhouse Gas Control*, 3(4):444 – 457, 2009. ISSN 1750-5836. doi: 10.1016/j.ijggc.2009.02.009.

Oliver, D. and Chen, Y. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15:185–221, 2011.

Oliver, D., Reynolds, A., Bi, Z., and Abacioglu, Y. Integration of production data into reservoir models. *Pet. Geosci.*, 7:65–73, 2001.

Oliver, D. S., Reynolds, A. C., and Liu, N. *Inverse theory for petroleum reservoir characterization and history matching.* Cambridge Univ Press, 2008.

Orbey, H. and Sandler, S. *Modeling Vapor-Liquid Equilibria: cubic equations of state and their mixing rules.* Cambridge University Press, 1998.

P.A., S. and Smørgrav, E. Response surface methodology approach for history matching and uncertainty assessment of reservoir simulation models. *Europec/EAGE Conference and Exhibition*, SPE 113390:Rome, Italy, 9–12 June, 2008.

Paffrath, M. and Wever, U. Adapted polynomial chaos expansion for failure detection. *Journal of Computational Physics*, 226(1):263–281, 2007.

Pajonk, O., Rosic, B. V., Litvinenko, A., and Matthies, H. G. A deterministic filter for non-Gaussian bayesian estimation – Applications to dynamical system estimation with noisy measurements. *Physica D*, 241(775-788), 2012.

Panfilova, I. and Panfilov, M. Near-critical gas–liquid flow in porous media: Monovariant model, analytical solutions and convective mass exchange effects. *Transport in porous media*, 56(1):61–85, 2004.

Pappenberger, F. and Beven, K. J. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. *Water Resources Research*, 42(5):1–8, 2006.

Pearson, K. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

Peng, D. and Robinson, D. A new two-constant equation of state. *Industrial & Engineering Chemistry Fundamentals*, 15(1):59–64, 1976.

Plischke, E. An effective algorithm for computing global sensitivity indices (easi). *Reliability Engineering and System Safety*, 95(4):354–360, 2010.

Pollock, D. Semianalytical computation of path lines for finite-difference models. *Ground Water*, 26(6):743–750, 1988.

Pray, H., Schweickert, C., and Minnich, B. Solubility of hydrogen, oxygen, nitrogen, and helium in water at elevated temperatures. *Industrial & Engineering Chemistry*, 44(5): 1146–1151, 1952.

Prempraneerach, P., Hover, F., Triantafyllou, M., and Karniadakis, G. Uncertainty quantification in simulations of power systems: Multi-element polynomial chaos methods. *Reliability Engineering & System Safety*, 95(6):632–646, 2010.

Red-Horse, J. and Benjamin, A. A probabilistic approach to uncertainty quantification with limited information. *Reliability Engineering & System Safety*, 85(1):183–190, 2004.

Redlich, O. and Kwong, J. On the thermodynamics of solutions. v. an equation of state. fugacities of gaseous solutions. *Chemical Reviews*, 44(1):233–244, 1949.

Reuter, U. and Liebscher, M. Global sensitivity analysis in view of nonlinear structural behavior. In *Proceedings of the 7th LS-Dyna Forum, Bamberg*, 2008.

Riddiford, F., Wright, I., Bishop, C., Espie, T., and Tourqui, A. Monitoring geological storage the in salah gas CO_2 storage project. In *Proceedings of the 7th International Greenhouse Gas Technologies Conference, Vancouver, BC*, 2004.

Riesz, M. Sur le probleme des moments et le théoreme de parseval correspondant. *Acta Litt. Acad. Sci. Szeged.*, 1:209–225, 1923.

Riva, M. and Willmann, M. Impact of log-transmissivity variogram structure on groundwater flow and transport predictions. *Advances in Water Resources*, 32(8):1311–1322, 2009. ISSN 0309-1708.

Robert, C. P. and Casella, G. *Monte Carlo methods*. New York: Springer, 2004.

Rodrigues, J. Calculating derivatives for automatic history matching. *Computational Geosciences*, 10:119–136, 2006.

Rohmer, J. and Seyedi, D. Coupled large scale hydromechanical modelling for caprock failure risk assessment of CO_2 storage in deep saline aquifers. *Oil & Gas Science and Technology–Revue de l'Institut Français du Pétrole*, 65(3):503–517, 2010.

Rowe, W. D. Understanding uncertainty. *Risk Analysis*, 14(5):743–750, 1994. ISSN 1539-6924. doi: 10.1111/j.1539-6924.1994.tb00284.x.

Rubin, Y. *Applied Stochastic Hydrogeology*. Oxford University Press, Oxford, 2003.

Rubin, Y. and Dagan, G. Conditional estimates of solute travel time in heterogenous formations: impact of transmissivity measurements. *Water Resour. Res.*, 28(4):1033–1040, 1992.

Rubin, Y., Cushey, M. A., and Bellin, A. Modeling of transport in groundwater for environmental risk assessment. *Stochastic Hydrol. Hydraul.*, 8(1):57–77, 1994.

Rygaard, M., Arvin, E., Bath, A., and Binning, P. Designing water supplies: Optimizing drinking water composition for maximum economic benefit. *Water Research*, 45(12): 3712 – 3722, 2011. ISSN 0043-1354. doi: DOI:10.1016/j.watres.2011.04.025.

S., O. and Panfilov, M. Hydrogen penetration in water through porous medium: application to a radioactive waste storage site. *Environmental Earth Sciences*, 64(4):989–999, 2011.

S., O., Royer, J.-J., and Panfilov, M. Effective solution through the streamline technique and HT-splitting for the 3D dynamic analysis of the compositional flows in oil reservoirs. *Transport in Porous Media*, 74(3):311–329, 2008.

Saad, G. and Ghanem, R. Characterization of reservoir simulation models using a polynomial chaos-based ensemble Kalman filter. *Water Resour. Res.*, 45:W04417, 2009.

Saltelli, A. Global sensitivity analysis: an introduction. In *Proc. 4th International Conference on Sensitivity Analysis of Model Output*, pages 27–43, 2004.

Saltelli, A., Ratto, M., and Andres, T. *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, 2008.

Sambucini, V. A reference prior for the analysis of a response surface. *Journal of Statistical Planning and Inference*, 137(4):1119–1128, 2007.

Sarma, P., Durlofsky, L., Aziz, K., and Chen, W. Efficient real-time reservoir management using adjoint-based optimal control and model updating. *Computational Geosciences*, 10 (1):3–36, 2006.

Sarma, P., Durlofsky, L., and Aziz, K. Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. *Mathematical Geosciences*, 40 (1):3–32, 2008.

Schäfer, F., Walter, L., Class, H., and Müller, C. The regional pressure impact of CO_2 storage: a showcase study from the North German Basin. *Environmental Earth Sciences*, 65 (7): 2037–2049, 2011. doi: DOI:10.1007/s12665-011-1184-8.

Scheidt, C., Caers, J., Chen, Y., and Durlofsky, L. J. A multi-resolution workflow to generate high-resolution models constrained to dynamic data. *Computational Geosciences*, 15(3): 545–563, 2011.

Schmorak, S. and Mercado, A. Upconing of fresh water-sea water interface below pumping wells, field study. *Water Resources Research*, 5 (6):1290–1311, 1969.

Schoeniger, A., Nowak, W., and Hendricks Franssen, H.-J. Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography. *Water Resources Research*, 48(4):W04431, 2012.

Sedov, L. I. Mechanics of continuum [in russian], vol. 2, 1976.

Shi, L., Yang, J., Zhang, D., and Li, H. Probabilistic collocation method for unconfined flow in heterogeneous media. *Journal of Hydrology*, 365(1):4–10, 2009.

Siebert, W. M. Circuits, signals, and systems. *MIT Press: Cambridge, MA*, pages 410–411, 1986.

Siirila, E. R., Navarre-Sitchler, A. K., Maxwell, R. M., and McCray, J. E. A quantitative methodology to assess the risks to human health from CO_2s leakage into groundwater. *Advances in Water Resources*, 36(0):146 – 164, 2012. ISSN 0309-1708. doi: 10.1016/j. advwatres.2010.11.005.

Smith, A. F. M. and Gefland, A. E. Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46(2):84–88, 1992.

Sobol, I. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280, 2001.

Sobol, I. M. On sensitivity estimation for nonlinear mathematical models. *Mathem. Mod.*, 2 (1):112–118, 1990.

Soize, C. and Ghanem, R. Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput.*, 26(2):395–410, 2004.

Stieltjes, T., J. Quelques recherches sur la théorie des quadratures dites méchaniques. *Oeuvres I*, pages 377–396, 1884.

Stone, M. Applications of the theory of boolean rings to general topology. *Transactions of the American Mathematical Society*, 41(3):375–481, 1937.

Storlie, C., Swiler, L., Helton, J., and Sallaberry, C. Implementation and evaluation of non-parametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, 94(11):1735–1763, 2009.

Sudret, B. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93(7):964–979, 2008.

Sun, N. Z. *Inverse Problems in Groundwater Modeling (Theory and Applications of Transport in Porous Media)*. Springer, 1999.

Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.

Thiele, M. and Batycky, R. Discussion of spe65604-Streamline simulation: A technology update. *Journal of Petroleum Technologyl*, 53(5):26–27, 2001.

Thiele, M. and Edwards, M. Physically based higher order godunov schemes for compositional simulation. In *SPE Reservoir Simulation Symposium*, 2001.

Thiele, M., Batycky, R., Blunt, M., and Orr Jr, F. Simulating flow in heterogeneous systems using streamtubes and streamlines. *SPE Reservoir Engineering*, 11(1):5–12, 1996.

Thiele, M., Batycky, R., and Blunt, M. A streamline-based 3d field-scale compositional simulator. *paper SPE*, 38889, 1997.

Thomas, D. *Carbon Dioxide Capture for Storage in Deep Geologic Formations-Results from the CO_2 Capture Project*, volume 1. Elsevier Science Ltd, 2005.

Todd, D. K. *Groundwater Hydrology*. John Wiley and Sons, New York, 1980.

USEPA. Risk assessment guidance for superfund Volume I: Part a, human health manual. *Tech. Rep. Rep.EPA/540/1-89/002*, December 1989.

USEPA. Risk assessment guidance for superfund: Volume III – part a, process for conducting probabilistic risk assessment. *Tech. Rep. Rep.EPA 540/R-02/002,*, December 2001.

Viebahn, P., Nitsch, J., Fischedick, M., Esken, A., Schüwer, D., Supersberger, N., Zuberbühler, U., and Edenhofer, O. Comparison of carbon capture and storage with renewable energy technologies regarding structural, economic, and ecological aspects in Germany. *International Journal of Greenhouse Gas Control*, 1(1):121–133, 2007.

Villadsen, J. and Michelsen, M. L. *Solution of differential equation models by polynomial approximation*. Prentice-Hall, 1978.

Vrugt, A. J., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. H., and Higdon, D. Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences & Numerical Simulation*, 10(3):271–288, 2009.

Wackernagel, H. *Multivariate geostatistics, an introduction with applications*. Second, Springer, Berlin, 1998.

Walas Stanley, M. Phase equilibria in chemical engineering, 1984.

Walker, W., Harremoes, P., Rotmans, J., Van der Sluijs, J., Van Asselt, M., Janssen, P., and Von Krauss, M. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1):5–17, 2003.

Walter, L., Oladyshkin, S., Class, H., Darcis, M., and Helmig, R. A study on pressure evolution in a sand channel system during CO_2 injection. *Energy Procedia*, 4:3722–3729, 2011.

Walter, L., Binning, P., Oladyshkin, S., Flemisch, B., and Class, H. Brine migration resulting from CO2 injection into saline aquifers - an approach to risk estimation including various levels of uncertainty. *International Journal of Greenhouse Gas Control*, 9:495–506, 2012. doi: 10.1016/j.ijggc.2012.05.004.

Walton, F., Tait, J., LeNeveu, D., and Sheppard, M. Geological storage of CO_2: A statistical approach to assessing performance and risk. In *Proceedings of the 7th International Conference on Greenhouse Gas Control Technologies (GHGT-7)*, 2004.

Wan, X. and Karniadakis, G., E. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM Journal of Scientific Computing*, 28(3):901–928, 2006.

Wan, X. and Karniadakis, G., E. Error control in multi-element generalized polynomial chaos method for elliptic problems with random coefficients. *Communication in Computational Physics*, 5:793–820, 2009.

Wan, X. and Karniadakis, G. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM Journal on Scientific Computing*, 28(3):901–928, 2007.

Wand, M. P. and Jones, M. C. *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability (60). Boca Raton, FL, U.S.: Chapman & Hall., 1994.

Wang, Y. and Kovscek, A. Streamline approach for history matching production data. *SPE Journal*, 5(4):353–362, 2000.

Wang, Y., Li, G., and Reynolds, A. Estimation of depths of fluid contacts by history matching using iterative ensemble Kalman smoothers. *SPE Journal*, 15(2):509–529, 2010.

Webster, M., Tatang, M., and McRae, G. Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model. *Report Series No. 4, Massachusetts Institute of Technology*, 1996.

Whitson, C., Fevang, Ø., and Sævareid, A. Gas condensate relative permeability for well calculations. *Transport in porous media*, 52(2):279–311, 2003.

Wiebe, R. and Gaddy, V. The solubility of hydrogen in water at 0, 50, 75 and 100c from 25 to 1000 atmospheres. *Journal of the American Chemical Society*, 56(1):76–79, 1934.

Wiebe, R., Gaddy, V., and Heins, C. Solubility of hydrogen in water at 25c from 25 to 1000 atmospheres. *Industrial & Engineering Chemistry*, 24(7):823–825, 1932.

Wiener, N. The homogeneous chaos. *Am. J. Math*, 60:897–936, 1938.

Wildenborg, A., Leijnse, A., Kreft, E., Nepveu, M., Obdam, A., Orlic, B., Wipfler, E., Grift, B., Kesteren, W., Gaus, I., et al. *Risk assessment methodology for CO_2 storage: the scenario approach.* Elsevier, 2005.

Williams, M., Keating, J., and Barghouty, M. The stratigraphic method: a structured approach to history-matching complex simulation models. *SPE Reserv. Evalu. Eng.*, 1(2): 169–176, 1998.

Wilson, E., Johnson, T., and Keith, D. Regulating the ultimate sink: Managing the risks of geologic CO_2 storage. *Environmental Science and Technology*, 37(16):3476–3483, 2003.

Winter, C. L., Guadagnini, A., Nychka, D., and Tartakovsky, D. M. Multivariate sensitivity analysis of saturated flow through simulated highly heterogeneous groundwater aquifers. *Journal of Computational Physics*, 217(1):166–175, 2006. ISSN 0021-9991.

Winterstein, S. R. Nonlinear vibration models for extremes and fatigue. *Journal of Engineering Mechanics*, 114(10):1772–1790, 1988.

Witteveen, J. A. S. and Bijl, H. Modeling arbitrary uncertainties using Gram-Schmidt polynomial chaos. *44th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, Nevada:AIAA–2006–896, 2006.

Witteveen, J. A. S., Sarkar, S., and Bijl, H. Modeling physical uncertainties in dynamic stall induced fluid structure interaction of turbine blades using arbitrary polynomial chaos. *Computers and Structures*, 85:866–878, 2007.

Wojtkiewicz, S. F., Eldred, M. S., Field, R. V., and Urbina, A. Uncertainty quantification in large computational engineering models uncertainty quantification. *Optimization*, 14, 2001.

Woodbury, A. and Ulrych, T. Minimum relative entropy: forward probabilistic modeling. *Water Resources Research*, 29(8):2847–2860, 1993.

Xiu, D. Efficient collocational approach for parametric uncertainty analysis. *Communications in computational physics*, 2(2):293–309, 2007.

Xiu, D. and Hesthaven, J. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005.

Xiu, D. and Karniadakis, G. The wiener-askey polynomial chaos for stochastic differential equations. *SIAM Journal of Scientific Computing*, 24(2):619–644, 2002a.

Xiu, D. and Karniadakis, G. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002b.

Xiu, D. and Karniadakis, G. E. Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, 187:137–167, 2003.

Yeh, W. Review of parameter identification procedures in groundwater hydrology. *Water Resour. Res*, 22(2):95–108, 1986.

Zabalza-Mezghani, I., Manceau, E., Feraille, M., and Jourdan, A. Uncertainty management: From geological scenarios to production scheme optimization. *Journal of Petroleum Science and Engineering*, 44:11–25, 2004.

Zafari, M. and Reynolds, A. Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter. *SPE J.*, 12(3):382–391, 2007.

Zhang, D. *Stochastic Methods for Flow in Porous Media*. Academic Press, San Diego, 2002.

Zhang, D. and Lu, Z. An efficient, high-order perturbation approach for flow in random media via Karhunen-Loeve and polynomial expansions. *Journal of Computational Physics*, 194:773–794, 2004.

Zweigel, P. and Heill, L. Studies on the likelihood for caprock fracturing in the sleipner CO_2 injection case. Technical report, Sintef Petroleum Research Report, 2003.
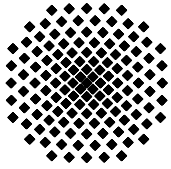
**Institut für Wasser- und
Umweltsystemmodellierung
Universität Stuttgart**

Pfaffenwaldring 61
70569 Stuttgart (Vaihingen)
Telefon (0711) 685 - 64717/64749/64752/64679
Telefax (0711) 685 - 67020 o. 64746 o. 64681
E-Mail: iws@iws.uni-stuttgart.de
http://www.iws.uni-stuttgart.de

Verzeichnis der Mitteilungshefte

5   Plate, Erich: *Beitrag zur Bestimmung der Windgeschwindigkeitsverteilung in der durch eine Wand gestörten bodennahen Luftschicht,*
und
Röhnisch, Arthur; Marotz, Günter: *Neue Baustoffe und Bauausführungen für den Schutz der Böschungen und der Sohle von Kanälen, Flüssen und Häfen; Gestehungskosten und jeweilige Vorteile*, sowie Unny, T.E.: *Schwingungsuntersuchungen am Kegelstrahlschieber*, 1967

6   Seiler, Erich: *Die Ermittlung des Anlagenwertes der bundeseigenen Binnenschiffahrtsstraßen und Talsperren und des Anteils der Binnenschiffahrt an diesem Wert*, 1967

7   *Sonderheft anläßlich des 65. Geburtstages von Prof. Arthur Röhnisch mit Beiträgen von* Benk, Dieter; Breitling, J.; Gurr, Siegfried; Haberhauer, Robert; Honekamp, Hermann; Kuz, Klaus Dieter; Marotz, Günter; Mayer-Vorfelder, Hans-Jörg; Miller, Rudolf; Plate, Erich J.; Radomski, Helge; Schwarz, Helmut; Vollmer, Ernst; Wildenhahn, Eberhard; 1967

8   Jumikis, Alfred: *Beitrag zur experimentellen Untersuchung des Wassernachschubs in einem gefrierenden Boden und die Beurteilung der Ergebnisse*, 1968

9   Marotz, Günter: *Technische Grundlagen einer Wasserspeicherung im natürlichen Untergrund*, 1968

10   Radomski, Helge: *Untersuchungen über den Einfluß der Querschnittsform wellenförmiger Spundwände auf die statischen und rammtechnischen Eigenschaften*, 1968

11   Schwarz, Helmut: *Die Grenztragfähigkeit des Baugrundes bei Einwirkung vertikal gezogener Ankerplatten als zweidimensionales Bruchproblem*, 1969

12   Erbel, Klaus: *Ein Beitrag zur Untersuchung der Metamorphose von Mittelgebirgsschneedecken unter besonderer Berücksichtigung eines Verfahrens zur Bestimmung der thermischen Schneequalität*, 1969

13   Westhaus, Karl-Heinz: *Der Strukturwandel in der Binnenschiffahrt und sein Einfluß auf den Ausbau der Binnenschiffskanäle*, 1969

14   Mayer-Vorfelder, Hans-Jörg: *Ein Beitrag zur Berechnung des Erdwiderstandes unter Ansatz der logarithmischen Spirale als Gleitflächenfunktion*, 1970

15   Schulz, Manfred: *Berechnung des räumlichen Erddruckes auf die Wandung kreiszylindrischer Körper*, 1970

16   Mobasseri, Manoutschehr: *Die Rippenstützmauer. Konstruktion und Grenzen ihrer Standsicherheit*, 1970

17   Benk, Dieter: *Ein Beitrag zum Betrieb und zur Bemessung von Hochwasserrückhaltebecken*, 1970

33　Rueff, Helge: *Untersuchung der schwingungserregenden Kräfte an zwei hintereinander angeordneten Tiefschützen unter besonderer Berücksichtigung von Kavitation*, 1974

34　Röhnisch, Arthur: *Einpreßversuche mit Zementmörtel für Spannbeton - Vergleich der Ergebnisse von Modellversuchen mit Ausführungen in Hüllwellrohren*, 1975

35　*Sonderheft anläßlich des 65. Geburtstages von Prof. Dr.-Ing. Kurt Petrikat mit Beiträgen von*:　Brombach, Hansjörg; Erbel, Klaus; Flinspach, Dieter; Fischer jr., Richard; Gàl, Attila; Gerlach, Reinhard; Giesecke, Jürgen; Haberhauer, Robert; Hafner Edzard; Hausenblas, Bernhard; Horlacher, Hans-Burkhard; Hutarew, Andreas; Knoll, Manfred; Krummet, Ralph; Marotz, Günter; Merkle, Theodor; Miller, Christoph; Minor, Hans-Erwin; Neumayer, Hans; Rao, Syamala; Rath, Paul; Rueff, Helge; Ruppert, Jürgen; Schwarz, Wolfgang; Topal-Gökceli, Mehmet; Vollmer, Ernst; Wang, Chung-su; Weber, Hans-Georg; 1975

36　Berger, Jochum: *Beitrag zur Berechnung des Spannungszustandes in rotationssymmetrisch belasteten Kugelschalen veränderlicher Wandstärke unter Gas- und Flüssigkeitsdruck durch Integration schwach singulärer Differentialgleichungen*, 1975

37　Dirksen, Wolfram: *Berechnung instationärer Abflußvorgänge in gestauten Gerinnen mittels Differenzenverfahren und die Anwendung auf Hochwasserrückhaltebecken*, 1976

38　Horlacher, Hans-Burkhard: *Berechnung instationärer Temperatur- und Wärmespannungsfelder in langen mehrschichtigen Hohlzylindern*, 1976

39　Hafner, Edzard: *Untersuchung der hydrodynamischen Kräfte auf Baukörper im Tiefwasserbereich des Meeres*, 1977, ISBN 3-921694-39-6

40　Ruppert, Jürgen: *Über den Axialwirbelkammerverstärker für den Einsatz im Wasserbau*, 1977, ISBN 3-921694-40-X

41　Hutarew, Andreas: *Beitrag zur Beeinflußbarkeit des Sauerstoffgehalts in Fließgewässern an Abstürzen und Wehren*, 1977, ISBN 3-921694-41-8, vergriffen

42　Miller, Christoph: *Ein Beitrag zur Bestimmung der schwingungserregenden Kräfte an unterströmten Wehren*, 1977, ISBN 3-921694-42-6

43　Schwarz, Wolfgang: *Druckstoßberechnung unter Berücksichtigung der Radial- und Längsverschiebungen der Rohrwandung*, 1978, ISBN 3-921694-43-4

44　Kinzelbach, Wolfgang: *Numerische Untersuchungen über den optimalen Einsatz variabler Kühlsysteme einer Kraftwerkskette am Beispiel Oberrhein*, 1978, ISBN 3-921694-44-2

45　Barczewski, Baldur: *Neue Meßmethoden für Wasser-Luftgemische und deren Anwendung auf zweiphasige Auftriebsstrahlen*, 1979, ISBN 3-921694-45-0

46 Neumayer, Hans: *Untersuchung der Strömungsvorgänge in radialen Wirbelkammerverstärkern*, 1979, ISBN 3-921694-46-9

47 Elalfy, Youssef-Elhassan: *Untersuchung der Strömungsvorgänge in Wirbelkammerdioden und -drosseln*, 1979, ISBN 3-921694-47-7

48 Brombach, Hansjörg: *Automatisierung der Bewirtschaftung von Wasserspeichern*, 1981, ISBN 3-921694-48-5

49 Geldner, Peter: *Deterministische und stochastische Methoden zur Bestimmung der Selbstdichtung von Gewässern*, 1981, ISBN 3-921694-49-3, vergriffen

50 Mehlhorn, Hans: *Temperaturveränderungen im Grundwasser durch Brauchwassereinleitungen*, 1982, ISBN 3-921694-50-7, vergriffen

51 Hafner, Edzard: *Rohrleitungen und Behälter im Meer*, 1983, ISBN 3-921694-51-5

52 Rinnert, Bernd: *Hydrodynamische Dispersion in porösen Medien: Einfluß von Dichteunterschieden auf die Vertikalvermischung in horizontaler Strömung*, 1983, ISBN 3-921694-52-3, vergriffen

53 Lindner, Wulf: *Steuerung von Grundwasserentnahmen unter Einhaltung ökologischer Kriterien*, 1983, ISBN 3-921694-53-1, vergriffen

54 Herr, Michael; Herzer, Jörg; Kinzelbach, Wolfgang; Kobus, Helmut; Rinnert, Bernd: *Methoden zur rechnerischen Erfassung und hydraulischen Sanierung von Grundwasserkontaminationen*, 1983, ISBN 3-921694-54-X

55 Schmitt, Paul: *Wege zur Automatisierung der Niederschlagsermittlung*, 1984, ISBN 3-921694-55-8, vergriffen

56 Müller, Peter: *Transport und selektive Sedimentation von Schwebstoffen bei gestautem Abfluß*, 1985, ISBN 3-921694-56-6

57 El-Qawasmeh, Fuad: *Möglichkeiten und Grenzen der Tropfbewässerung unter besonderer Berücksichtigung der Verstopfungsanfälligkeit der Tropfelemente*, 1985, ISBN 3-921694-57-4, vergriffen

58 Kirchenbaur, Klaus: *Mikroprozessorgesteuerte Erfassung instationärer Druckfelder am Beispiel seegangsbelasteter Baukörper*, 1985, ISBN 3-921694-58-2

59 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1984/85 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart), 1985, ISBN 3-921694-59-0, vergriffen

60 Spitz, Karlheinz: *Dispersion in porösen Medien: Einfluß von Inhomogenitäten und Dichteunterschieden*, 1985, ISBN 3-921694-60-4, vergriffen

61 Kobus, Helmut: *An Introduction to Air-Water Flows in Hydraulics*, 1985, ISBN 3-921694-61-2

62   Kaleris, Vassilios: *Erfassung des Austausches von Oberflächen- und Grundwasser in horizontalebenen Grundwassermodellen*, 1986, ISBN 3-921694-62-0

63   Herr, Michael: *Grundlagen der hydraulischen Sanierung verunreinigter Porengrundwasserleiter*, 1987, ISBN 3-921694-63-9

64   Marx, Walter: *Berechnung von Temperatur und Spannung in Massenbeton infolge Hydratation*, 1987, ISBN 3-921694-64-7

65   Koschitzky, Hans-Peter: *Dimensionierungskonzept für Sohlbelüfter in Schußrinnen zur Vermeidung von Kavitationsschäden*, 1987, ISBN 3-921694-65-5

66   Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1986/87 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart) 1987, ISBN 3-921694-66-3

67   Söll, Thomas: *Berechnungsverfahren zur Abschätzung anthropogener Temperaturanomalien im Grundwasser*, 1988, ISBN 3-921694-67-1

68   Dittrich, Andreas; Westrich, Bernd: *Bodenseeufererosion, Bestandsaufnahme und Bewertung*, 1988, ISBN 3-921694-68-X, vergriffen

69   Huwe, Bernd; van der Ploeg, Rienk R.: *Modelle zur Simulation des Stickstoffhaushaltes von Standorten mit unterschiedlicher landwirtschaftlicher Nutzung*, 1988, ISBN 3-921694-69-8, vergriffen

70   Stephan, Karl: *Integration elliptischer Funktionen*, 1988, ISBN 3-921694-70-1

71   Kobus, Helmut; Zilliox, Lothaire (Hrsg.): *Nitratbelastung des Grundwassers, Auswirkungen der Landwirtschaft auf die Grundwasser- und Rohwasserbeschaffenheit und Maßnahmen zum Schutz des Grundwassers*. Vorträge des deutschfranzösischen Kolloquiums am 6. Oktober 1988, Universitäten Stuttgart und Louis Pasteur Strasbourg (Vorträge in deutsch oder französisch, Kurzfassungen zweisprachig), 1988, ISBN 3-921694-71-X

72   Soyeaux, Renald: *Unterströmung von Stauanlagen auf klüftigem Untergrund unter Berücksichtigung laminarer und turbulenter Fließzustände*,1991, ISBN 3-921694-72-8

73   Kohane, Roberto: *Berechnungsmethoden für Hochwasserabfluß in Fließgewässern mit überströmten Vorländern*, 1991, ISBN 3-921694-73-6

74   Hassinger, Reinhard: *Beitrag zur Hydraulik und Bemessung von Blocksteinrampen in flexibler Bauweise*, 1991, ISBN 3-921694-74-4, vergriffen

75   Schäfer, Gerhard: *Einfluß von Schichtenstrukturen und lokalen Einlagerungen auf die Längsdispersion in Porengrundwasserleitern*, 1991, ISBN 3-921694-75-2

76   Giesecke, Jürgen: *Vorträge, Wasserwirtschaft in stark besiedelten Regionen; Umweltforschung mit Schwerpunkt Wasserwirtschaft*, 1991, ISBN 3-921694-76-0

77 Huwe, Bernd: *Deterministische und stochastische Ansätze zur Modellierung des Stickstoffhaushalts landwirtschaftlich genutzter Flächen auf unterschiedlichem Skalenniveau*, 1992, ISBN 3-921694-77-9, vergriffen

78 Rommel, Michael: *Verwendung von Kluftdaten zur realitätsnahen Generierung von Kluftnetzen mit anschließender laminar-turbulenter Strömungsberechnung*, 1993, ISBN 3-92 1694-78-7

79 Marschall, Paul: *Die Ermittlung lokaler Stofffrachten im Grundwasser mit Hilfe von Einbohrloch-Meßverfahren*, 1993, ISBN 3-921694-79-5, vergriffen

80 Ptak, Thomas: *Stofftransport in heterogenen Porenaquiferen: Felduntersuchungen und stochastische Modellierung*, 1993, ISBN 3-921694-80-9, vergriffen

81 Haakh, Frieder: *Transientes Strömungsverhalten in Wirbelkammern,* 1993, ISBN 3-921694-81-7

82 Kobus, Helmut; Cirpka, Olaf; Barczewski, Baldur; Koschitzky, Hans-Peter: *Versucheinrichtung zur Grundwasser und Altlastensanierung VEGAS, Konzeption und Programmrahmen*, 1993, ISBN 3-921694-82-5

83 Zang, Weidong: *Optimaler Echtzeit-Betrieb eines Speichers mit aktueller Abflußregenerierung*, 1994, ISBN 3-921694-83-3, vergriffen

84 Franke, Hans-Jörg: *Stochastische Modellierung eines flächenhaften Stoffeintrages und Transports in Grundwasser am Beispiel der Pflanzenschutzmittelproblematik*, 1995, ISBN 3-921694-84-1

85 Lang, Ulrich: *Simulation regionaler Strömungs- und Transportvorgänge in Karstaquiferen mit Hilfe des Doppelkontinuum-Ansatzes: Methodenentwicklung und Parameteridentifikation*, 1995, ISBN 3-921694-85-X, vergriffen

86 Helmig, Rainer: *Einführung in die Numerischen Methoden der Hydromechanik*, 1996, ISBN 3-921694-86-8, vergriffen

87 Cirpka, Olaf: *CONTRACT: A Numerical Tool for Contaminant Transport and Chemical Transformations - Theory and Program Documentation -*, 1996, ISBN 3-921694-87-6

88 Haberlandt, Uwe: *Stochastische Synthese und Regionalisierung des Niederschlages für Schmutzfrachtberechnungen*, 1996, ISBN 3-921694-88-4

89 Croisé, Jean: *Extraktion von flüchtigen Chemikalien aus natürlichen Lockergesteinen mittels erzwungener Luftströmung*, 1996, ISBN 3-921694-89-2, vergriffen

90 Jorde, Klaus: *Ökologisch begründete, dynamische Mindestwasserregelungen bei Ausleitungskraftwerken*, 1997, ISBN 3-921694-90-6, vergriffen

91 Helmig, Rainer: *Gekoppelte Strömungs- und Transportprozesse im Untergrund - Ein Beitrag zur Hydrosystemmodellierung-*, 1998, ISBN 3-921694-91-4, vergriffen

92   Emmert, Martin: *Numerische Modellierung nichtisothermer Gas-Wasser Systeme in porösen Medien,* 1997, ISBN 3-921694-92-2

93   Kern, Ulrich: *Transport von Schweb- und Schadstoffen in staugeregelten Fließgewässern am Beispiel des Neckars*, 1997, ISBN 3-921694-93-0, vergriffen

94   Förster, Georg: *Druckstoßdämpfung durch große Luftblasen in Hochpunkten von Rohrleitungen 1997*, ISBN 3-921694-94-9

95   Cirpka, Olaf: *Numerische Methoden zur Simulation des reaktiven Mehrkomponententransports im Grundwasser*, 1997, ISBN 3-921694-95-7, vergriffen

96   Färber, Arne: *Wärmetransport in der ungesättigten Bodenzone: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1997, ISBN 3-921694-96-5

97   Betz, Christoph: *Wasserdampfdestillation von Schadstoffen im porösen Medium: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1998, ISBN 3-921694-97-3

98   Xu, Yichun: *Numerical Modeling of Suspended Sediment Transport in Rivers*, 1998, ISBN 3-921694-98-1, vergriffen

99   Wüst, Wolfgang: *Geochemische Untersuchungen zur Sanierung CKW-kontaminierter Aquifere mit Fe(0)-Reaktionswänden,* 2000, ISBN 3-933761-02-2

100   Sheta, Hussam: *Simulation von Mehrphasenvorgängen in porösen Medien unter Einbeziehung von Hysterese-Effekten*, 2000, ISBN 3-933761-03-4

101   Ayros, Edwin: *Regionalisierung extremer Abflüsse auf der Grundlage statistischer Verfahren*, 2000, ISBN 3-933761-04-2, vergriffen

102   Huber, Ralf: *Compositional Multiphase Flow and Transport in Heterogeneous Porous Media*, 2000, ISBN 3-933761-05-0

103   Braun, Christopherus: *Ein Upscaling-Verfahren für Mehrphasenströmungen in porösen Medien*, 2000, ISBN 3-933761-06-9

104   Hofmann, Bernd: *Entwicklung eines rechnergestützten Managementsystems zur Beurteilung von Grundwasserschadensfällen*, 2000, ISBN 3-933761-07-7

105   Class, Holger: *Theorie und numerische Modellierung nichtisothermer Mehrphasenprozesse in NAPL-kontaminierten porösen Medien,* 2001, ISBN 3-933761-08-5

106   Schmidt, Reinhard: *Wasserdampf- und Heißluftinjektion zur thermischen Sanierung kontaminierter Standorte*, 2001, ISBN 3-933761-09-3

107   Josef, Reinhold:, *Schadstoffextraktion mit hydraulischen Sanierungsverfahren unter Anwendung von grenzflächenaktiven Stoffen*, 2001, ISBN 3-933761-10-7

124   Hrsg.: Schrenk, Volker; Batereau, Katrin; Barczewski, Baldur; Weber, Karolin und Koschitzky, Hans-Peter: *Symposium Ressource Fläche und VEGAS - Statuskolloquium 2003, 30. September und 1. Oktober 2003,* 2003, ISBN 3-933761-27-1

125   Omar Khalil Ouda: *Optimisation of Agricultural Water Use: A Decision Support System for the Gaza Strip,* 2003, ISBN 3-933761-28-0

126   Batereau, Katrin: *Sensorbasierte Bodenluftmessung zur Vor-Ort-Erkundung von Schadensherden im Untergrund*, 2004, ISBN 3-933761-29-8

127   Witt, Oliver: *Erosionsstabilität von Gewässersedimenten mit Auswirkung auf den Stofftransport bei Hochwasser am Beispiel ausgewählter Stauhaltungen des Oberrheins*, 2004, ISBN 3-933761-30-1

128   Jakobs, Hartmut: *Simulation nicht-isothermer Gas-Wasser-Prozesse in komplexen Kluft-Matrix-Systemen*, 2004, ISBN 3-933761-31-X

129   Li, Chen-Chien: *Deterministisch-stochastisches Berechnungskonzept zur Beurteilung der Auswirkungen erosiver Hochwasserereignisse in Flussstauhaltungen,* 2004, ISBN 3-933761-32-8

130   Reichenberger, Volker; Helmig, Rainer; Jakobs, Hartmut; Bastian, Peter; Niessner, Jennifer: *Complex Gas-Water Processes in Discrete Fracture-Matrix Systems: Upscaling, Mass-Conservative Discretization and Efficient Multilevel Solution,* 2004, ISBN 3-933761-33-6

131   Hrsg.: Barczewski, Baldur; Koschitzky, Hans-Peter; Weber, Karolin; Wege, Ralf: *VEGAS - Statuskolloquium 2004,* Tagungsband zur Veranstaltung am 05. Oktober 2004 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2004, ISBN 3-933761-34-4

132   Asie, Kemal Jabir: *Finite Volume Models for Multiphase Multicomponent Flow through Porous Media*. 2005, ISBN 3-933761-35-2

133   Jacoub, George: *Development of a 2-D Numerical Module for Particulate Contaminant Transport in Flood Retention Reservoirs and Impounded Rivers,* 2004, ISBN 3-933761-36-0

134   Nowak, Wolfgang: *Geostatistical Methods for the Identification of Flow and Transport Parameters in the Subsurface,* 2005, ISBN 3-933761-37-9

135   Süß, Mia: *Analysis of the influence of structures and boundaries on flow and transport processes in fractured porous media*, 2005, ISBN 3-933761-38-7

136   Jose, Surabhin Chackiath: *Experimental Investigations on Longitudinal Dispersive Mixing in Heterogeneous Aquifers,* 2005, ISBN: 3-933761-39-5

137   Filiz, Fulya: *Linking Large-Scale Meteorological Conditions to Floods in Mesoscale Catchments,* 2005, ISBN 3-933761-40-9

138 Qin, Minghao: *Wirklichkeitsnahe und recheneffiziente Ermittlung von Temperatur und Spannungen bei großen RCC-Staumauern*, 2005, ISBN 3-933761-41-7

139 Kobayashi, Kenichiro: *Optimization Methods for Multiphase Systems in the Subsurface - Application to Methane Migration in Coal Mining Areas*, 2005, ISBN 3-933761-42-5

140 Rahman, Md. Arifur: *Experimental Investigations on Transverse Dispersive Mixing in Heterogeneous Porous Media,* 2005, ISBN 3-933761-43-3

141 Schrenk, Volker: *Ökobilanzen zur Bewertung von Altlastensanierungsmaßnahmen,* 2005, ISBN 3-933761-44-1

142 Hundecha, Hirpa Yeshewatesfa: *Regionalization of Parameters of a Conceptual Rainfall-Runoff Model*, 2005, ISBN: 3-933761-45-X

143 Wege, Ralf: *Untersuchungs- und Überwachungsmethoden für die Beurteilung natürlicher Selbstreinigungsprozesse im Grundwasser,* 2005, ISBN 3-933761-46-8

144 Breiting, Thomas: *Techniken und Methoden der Hydroinformatik - Modellierung von komplexen Hydrosystemen im Untergrund*, 2006, 3-933761-47-6

145 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Müller, Martin: *Ressource Untergrund: 10 Jahre VEGAS: Forschung und Technologieentwicklung zum Schutz von Grundwasser und Boden,* Tagungsband zur Veranstaltung am 28. und 29. September 2005 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2005, ISBN 3-933761-48-4

146 Rojanschi, Vlad: *Abflusskonzentration in mesoskaligen Einzugsgebieten unter Berücksichtigung des Sickerraumes,* 2006, ISBN 3-933761-49-2

147 Winkler, Nina Simone: *Optimierung der Steuerung von Hochwasserrückhaltebecken-systemen,* 2006, ISBN 3-933761-50-6

148 Wolf, Jens: *Räumlich differenzierte Modellierung der Grundwasserströmung alluvialer Aquifere für mesoskalige Einzugsgebiete*, 2006, ISBN: 3-933761-51-4

149 Kohler, Beate: *Externe Effekte der Laufwasserkraftnutzung,* 2006, ISBN 3-933761-52-2

150 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS-Statuskolloquium 2006,* Tagungsband zur Veranstaltung am 28. September 2006 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2006, ISBN 3-933761-53-0

151 Niessner, Jennifer: *Multi-Scale Modeling of Multi-Phase - Multi-Component Processes in Heterogeneous Porous Media*, 2006, ISBN 3-933761-54-9

152 Fischer, Markus: *Beanspruchung eingeerdeter Rohrleitungen infolge Austrocknung bindiger Böden,* 2006, ISBN 3-933761-55-7

153   Schneck, Alexander: *Optimierung der Grundwasserbewirtschaftung unter Berück-sichtigung der Belange der Wasserversorgung, der Landwirtschaft und des Natur-schutzes*, 2006, ISBN 3-933761-56-5

154   Das, Tapash: *The Impact of Spatial Variability of Precipitation on the Predictive Uncertainty of Hydrological Models,* 2006, ISBN 3-933761-57-3

155   Bielinski, Andreas: *Numerical Simulation of CO_2 sequestration in geological for-mations*, 2007, ISBN 3-933761-58-1

156   Mödinger, Jens: *Entwicklung eines Bewertungs- und Entscheidungsunterstüt-zungssystems für eine nachhaltige regionale Grundwasserbewirtschaftung*, 2006, ISBN 3-933761-60-3

157   Manthey, Sabine: *Two-phase flow processes with dynamic effects in porous media - parameter estimation and simulation,* 2007, ISBN 3-933761-61-1

158   Pozos Estrada, Oscar: *Investigation on the Effects of Entrained Air in Pipelines,* 2007, ISBN 3-933761-62-X

159   Ochs, Steffen Oliver: *Steam injection into saturated porous media – process anal-ysis including experimental and numerical investigations*, 2007, ISBN 3-933761-63-8

160   Marx, Andreas: *Einsatz gekoppelter Modelle und Wetterradar zur Abschätzung von Niederschlagsintensitäten und zur Abflussvorhersage, 2007,* ISBN 3-933761-64-6

161   Hartmann, Gabriele Maria: *Investigation of Evapotranspiration Concepts in Hydro-logical Modelling for Climate Change Impact Assessment*, 2007, ISBN 3-933761-65-4

162   Kebede Gurmessa, Tesfaye: *Numerical Investigation on Flow and Transport Characteristics to Improve Long-Term Simulation of Reservoir Sedimentation*, 2007, ISBN 3-933761-66-2

163   Trifković, Aleksandar: *Multi-objective and Risk-based Modelling Methodology for Planning, Design and Operation of Water Supply Systems*, 2007, ISBN 3-933761-67-0

164   Götzinger, Jens: *Distributed Conceptual Hydrological Modelling - Simulation of Climate, Land Use Change Impact and Uncertainty Analysis*, 2007, ISBN 3-933761-68-9

165   Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS – Kolloquium 2007,* Tagungsband zur Veranstaltung am 26. September 2007 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2007, ISBN 3-933761-69-7

166   Freeman, Beau: *Modernization Criteria Assessment for Water Resources Plan-ning; Klamath Irrigation Project, U.S.*, 2008, ISBN 3-933761-70-0

167 Dreher, Thomas: *Selektive Sedimentation von Feinstschwebstoffen in Wechsel-wirkung mit wandnahen turbulenten Strömungsbedingungen*, 2008,
ISBN 3-933761-71-9

168 Yang, Wei: *Discrete-Continuous Downscaling Model for Generating Daily Precipitation Time Series*, 2008, ISBN 3-933761-72-7

169 Kopecki, Ianina: *Calculational Approach to FST-Hemispheres for Multiparametrical Benthos Habitat Modelling,* 2008*,* ISBN 3-933761-73-5

170 Brommundt, Jürgen: *Stochastische Generierung räumlich zusammenhängender Niederschlagszeitreihen*, 2008, ISBN 3-933761-74-3

171 Papafotiou, *Alexandros: Numerical Investigations of the Role of Hysteresis in Heterogeneous Two-Phase Flow Systems*, 2008, ISBN 3-933761-75-1

172 He, Yi: *Application of a Non-Parametric Classification Scheme to Catchment Hydrology*, 2008, ISBN 978-3-933761-76-7

173 Wagner, Sven: *Water Balance in a Poorly Gauged Basin in West Africa Using Atmospheric Modelling and Remote Sensing Information*, 2008,
ISBN 978-3-933761-77-4

174 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias; Schrenk, Volker: *VEGAS-Kolloquium 2008 Ressource Fläche III*, Tagungsband zur Veranstaltung am 01. Oktober 2008 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2008, ISBN 978-3-933761-78-1

175 Patil, Sachin: *Regionalization of an Event Based Nash Cascade Model for Flood Predictions in Ungauged Basins*, 2008, ISBN 978-3-933761-79-8

176 Assteerawatt, Anongnart: *Flow and Transport Modelling of Fractured Aquifers based on a Geostatistical Approach,* 2008, ISBN 978-3-933761-80-4

177 Karnahl, Joachim Alexander: *2D numerische Modellierung von multifraktionalem Schwebstoff- und Schadstofftransport in Flüssen*, 2008, ISBN 978-3-933761-81-1

178 Hiester, Uwe: *Technologieentwicklung zur In-situ-Sanierung der ungesättigten Bodenzone mit festen Wärmequellen,* 2009, ISBN 978-3-933761-82-8

179 Laux, Patrick: *Statistical Modeling of Precipitation for Agricultural Planning in the Volta Basin of West Africa,* 2009, ISBN 978-3-933761-83-5

180 Ehsan, Saqib: *Evaluation of Life Safety Risks Related to Severe Flooding,* 2009, ISBN 978-3-933761-84-2

181 Prohaska, Sandra: *Development and Application of a 1D Multi-Strip Fine Sediment Transport Model for Regulated Rivers,* 2009, ISBN 978-3-933761-85-9

182 Kopp, Andreas: *Evaluation of CO$_2$ Injection Processes in Geological Formations for Site Screening,* 2009, ISBN 978-3-933761-86-6

183 Ebigbo, Anozie: *Modelling of biofilm growth and its influence on CO$_2$ and water (two-phase) flow in porous media,* 2009, ISBN 978-3-933761-87-3

184 Freiboth, Sandra: *A phenomenological model for the numerical simulation of multiphase multicomponent processes considering structural alterations of porous media*, 2009, ISBN 978-3-933761-88-0

185 Zöllner, Frank: *Implementierung und Anwendung netzfreier Methoden im Konstruktiven Wasserbau und in der Hydromechanik*, 2009, ISBN 978-3-933761-89-7

186 Vasin, Milos: *Influence of the soil structure and property contrast on flow and transport in the unsaturated zone*, 2010, ISBN 978-3-933761-90-3

187 Li, Jing: *Application of Copulas as a New Geostatistical Tool*, 2010, ISBN 978-3-933761-91-0

188 AghaKouchak, Amir: *Simulation of Remotely Sensed Rainfall Fields Using Copulas*, 2010, ISBN 978-3-933761-92-7

189 Thapa, Pawan Kumar: *Physically-based spatially distributed rainfall runoff modelling for soil erosion estimation*, 2010, ISBN 978-3-933761-93-4

190 Wurms, Sven: *Numerische Modellierung der Sedimentationsprozesse in Retentionsanlagen zur Steuerung von Stoffströmen bei extremen Hochwasserabflussereignissen*, 2011, ISBN 978-3-933761-94-1

191 Merkel, Uwe: *Unsicherheitsanalyse hydraulischer Einwirkungen auf Hochwasserschutzdeiche und Steigerung der Leistungsfähigkeit durch adaptive Strömungsmodellierung*, 2011, ISBN 978-3-933761-95-8

192 Fritz, Jochen: *A Decoupled Model for Compositional Non-Isothermal Multiphase Flow in Porous Media and Multiphysics Approaches for Two-Phase Flow*, 2010, ISBN 978-3-933761-96-5

193 Weber, Karolin (Hrsg.): *12. Treffen junger WissenschaftlerInnen an Wasserbauinstituten*, 2010, ISBN 978-3-933761-97-2

194 Bliefernicht, Jan-Geert: *Probability Forecasts of Daily Areal Precipitation for Small River Basins,* 2011, ISBN 978-3-933761-98-9

195 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2010 In-situ-Sanierung - Stand und Entwicklung Nano und ISCO -,* Tagungsband zur Veranstaltung am 07. Oktober 2010 an der Universität Stuttgart, Campus Stuttgart-Vaihingen*, 2010, ISBN 978-3-933761-99-6

196 Gafurov, Abror: *Water Balance Modeling Using Remote Sensing Information - Focus on Central Asia,* 2010, ISBN 978-3-942036-00-9

197 Mackenberg, Sylvia: *Die Quellstärke in der Sickerwasserprognose: Möglichkeiten und Grenzen von Labor- und Freilanduntersuchungen,* 2010, ISBN 978-3-942036-01-6

198 Singh, Shailesh Kumar: *Robust Parameter Estimation in Gauged and Ungauged Basins,* 2010, ISBN 978-3-942036-02-3

199 Doğan, Mehmet Onur: *Coupling of porous media flow with pipe flow,* 2011, ISBN 978-3-942036-03-0

200 Liu, Min: *Study of Topographic Effects on Hydrological Patterns and the Implication on Hydrological Modeling and Data Interpolation,* 2011, ISBN 978-3-942036-04-7

201 Geleta, Habtamu Itefa: *Watershed Sediment Yield Modeling for Data Scarce Areas,* 2011, ISBN 978-3-942036-05-4

202 Franke, Jörg: *Einfluss der Überwachung auf die Versagenswahrscheinlichkeit von Staustufen,* 2011, ISBN 978-3-942036-06-1

203 Bakimchandra, Oinam: *Integrated Fuzzy-GIS approach for assessing regional soil erosion risks,* 2011, ISBN 978-3-942036-07-8

204 Alam, Muhammad Mahboob: *Statistical Downscaling of Extremes of Precipitation in Mesoscale Catchments from Different RCMs and Their Effects on Local Hydrology,* 2011, ISBN 978-3-942036-08-5

205 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2011 Flache Geothermie - Perspektiven und Risiken,* Tagungsband zur Veranstaltung am 06. Oktober 2011 an der Universität Stuttgart, Campus Stuttgart-Vaihingen*, 2011,* ISBN 978-3-933761-09-2

206 Haslauer, Claus: *Analysis of Real-World Spatial Dependence of Subsurface Hydraulic Properties Using Copulas with a Focus on Solute Transport Behaviour,* 2011, ISBN 978-3-942036-10-8

207 Dung, Nguyen Viet: *Multi-objective automatic calibration of hydrodynamic models – development of the concept and an application in the Mekong Delta,* 2011, ISBN 978-3-942036-11-5

208 Hung, Nguyen Nghia: *Sediment dynamics in the floodplain of the Mekong Delta, Vietnam,* 2011, ISBN 978-3-942036-12-2

209 Kuhlmann, Anna: *Influence of soil structure and root water uptake on flow in the unsaturated zone,* 2012, ISBN 978-3-942036-13-9

210 Tuhtan, Jeffrey Andrew: *Including the Second Law Inequality in Aquatic Ecody-namics: A Modeling Approach for Alpine Rivers Impacted by Hydropeaking,* 2012, ISBN 978-3-942036-14-6

211 Tolossa, Habtamu: *Sediment Transport Computation Using a Data-Driven Adaptive Neuro-Fuzzy Modelling Approach,* 2012, ISBN 978-3-942036-15-3

212 Tatomir, Alexandru-Bodgan: *From Discrete to Continuum Concepts of Flow in Fractured Porous Media,* 2012, ISBN 978-3-942036-16-0

213 Erbertseder, Karin: *A Multi-Scale Model for Describing Cancer-Therapeutic Transport in the Human Lung,* 2012, ISBN 978-3-942036-17-7

214 Noack, Markus: *Modelling Approach for Interstitial Sediment Dynamics and Reproduction of Gravel Spawning Fish,* 2012, ISBN 978-3-942036-18-4

215 De Boer, Cjestmir Volkert: *Transport of Nano Sized Zero Valent Iron Colloids during Injection into the Subsurface*, 2012, ISBN 978-3-942036-19-1

216 Pfaff, Thomas: *Processing and Analysis of Weather Radar Data for Use in Hydrology*, 2013, ISBN 978-3-942036-20-7

217 Lebrenz, Hans-Henning: *Addressing the Input Uncertainty for Hydrological Modeling by a New Geostatistical Method*, 2013, ISBN 978-3-942036-21-4

218 Darcis, Melanie Yvonne: *Coupling Models of Different Complexity for the Simulation of CO_2 Storage in Deep Saline Aquifers*, 2013, ISBN 978-3-942036-22-1

219 Beck, Ferdinand: *Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach*, 2013, ISBN 978-3-942036-23-8

220 Guthke, Philipp: *Non-multi-Gaussian spatial structures: Process-driven natural genesis, manifestation, modeling approaches, and influences on dependent processes*, 2013, ISBN 978-3-942036-24-5

221 Walter, Lena: *Uncertainty studies and risk assessment for CO_2 storage in geological formations*, 2013, ISBN 978-3-942036-25-2

222 Wolff, Markus: *Multi-scale modeling of two-phase flow in porous media including capillary pressure effects,* 2013, ISBN 978-3-942036-26-9

223 Mosthaf, Klaus Roland: *Modeling and analysis of coupled porous-medium and free flow with application to evaporation processes,* 2014, ISBN 978-3-942036-27-6

224 Leube, Philipp Christoph: *Methods for Physically-Based Model Reduction in Time: Analysis, Comparison of Methods and Application*, 2013, ISBN 978-3-942036-28-3

225 Rodríguez Fernández, Jhan Ignacio: *High Order Interactions among environmental variables: Diagnostics and initial steps towards modeling*, 2013, ISBN 978-3-942036-29-0

226 Eder, Maria Magdalena: *Climate Sensitivity of a Large Lake*, 2013,
     ISBN 978-3-942036-30-6

227 Greiner, Philipp: *Alkoholinjektion zur In-situ-Sanierung von CKW Schadensherden
     in Grundwasserleitern: Charakterisierung der relevanten Prozesse auf unter-
     schiedlichen Skalen*, 2014, ISBN 978-3-942036-31-3

228 Lauser, Andreas: *Theory and Numerical Applications of Compositional Multi-
     Phase Flow in Porous Media*, 2014, ISBN 978-3-942036-32-0

229 Enzenhöfer, Rainer: *Risk Quantification and Management in Water Production
     and Supply Systems*, 2014, ISBN 978-3-942036-33-7

230 Faigle, Benjamin: *Adaptive modelling of compositional multi-phase flow with capil-
     lary pressure*, 2014, ISBN 978-3-942036-34-4

231 Oladyshkin, Sergey: *Efficient modeling of environmental systems in the face of
     complexity and uncertainty*, 2014, ISBN 978-3-942036-35-1

*Die Mitteilungshefte ab der Nr. 134 (Jg. 2005) stehen als pdf-Datei über die Homepage
des Instituts: www.iws.uni-stuttgart.de zur Verfügung.*