



HAL
open science

Integration of uncertainty and definition of critical thresholds for CO₂ storage risk assessment

Tatiana Okhulkova

► **To cite this version:**

Tatiana Okhulkova. Integration of uncertainty and definition of critical thresholds for CO₂ storage risk assessment. Risques. Université Paris Saclay (COMUE), 2015. English. NNT : 2015SACLCO21 . tel-02122811

HAL Id: tel-02122811

<https://theses.hal.science/tel-02122811>

Submitted on 7 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

2015SACLCO21



THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY,
préparée à CentraleSupélec, Laboratoire MSSMat :
Mécanique des Sols, Structures et Matériaux (UMR CNRS 8579)

ÉCOLE DOCTORALE N° 579
Sciences mécaniques et énergétiques, matériaux et géosciences
Spécialité de doctorat : Energétique

Par

Mme Tatiana OKHULKOVA

Integration of uncertainty and definition of critical thresholds
in CO₂ storage risk assessment

Thèse présentée et soutenue à Châtenay-Malabry, le 15/12/2015

Composition du Jury :

M. SOIZE Christian	Professeur, Université Paris-Est	Président
M. NOETINGER Benoît	Professeur, IFPEN	Rapporteur
M. SUDRET Bruno	Professeur, ETH Zürich	Rapporteur
M. PANFILOV Michel	Professeur, Université de Lorraine	Examineur
M. CLOUTEAU Didier	Professeur, CentraleSupélec	Directeur de thèse
M. DE ROCQUIGNY Etienne	Professeur, Aremus&Associés	Encadrant
M. FARRET Régis	Ingénieur, INERIS	Encadrant



Titre : Insertion de l'incertitude et définition de seuils critiques pour l'analyse de risques du stockage du CO₂

Mots clés : Stockage de CO₂, Propagation d'incertitudes, Analyse de risques, Hétérogénéité spatiale, Champs aléatoire anisotrope, Simulation d'écoulements en milieu poreux

Title : Integration of uncertainty and definition of critical thresholds in CO₂ storage risk assessment

Keywords : CO₂ storage, Uncertainty propagation, Risk assessment, Spatial heterogeneity, Anisotropic random field, Flow simulation in porous medium

Résumé

L'objectif principal de la thèse est de définir comment l'incertitude peut être prise en compte dans le processus d'évaluation des risques pour le stockage de CO₂ et de quantifier, à l'aide de modèles numériques, les scénarios de fuite par migration latérale et à travers la couverture. Les scénarios choisis sont quantifiés par l'approche de modélisation de système pour laquelle des modèles numériques prédictifs *ad-hoc* sont développés. Une étude probabiliste de propagation d'incertitude paramétrique par un méta-modèle de polynômes de chaos est réalisée. La problématique de la prise en compte de la variabilité spatiale comme une source d'incertitude est éclairée et une étude comparative entre représentations homogène et hétérogène de la perméabilité est fournie.

Abstract

The main goal of the thesis is to define how the uncertainty can be accounted for in the process of risk assessment for CO₂ storage and to quantify by means of numerical models the scenarios of leakage by lateral migration and through the caprock. The chosen scenarios are quantified using the system modeling approach for which *ad-hoc* predictive numerical models are developed. A probabilistic parametric uncertainty propagation study using polynomial chaos expansion is performed. Matters of spatial variability are also discussed and a comparison between homogeneous and heterogeneous representations of permeability is provided.



Remerciements

J'aimerais remercier les membres du jury d'avoir accepté d'évaluer mon travail de thèse. Ce fut un grand honneur de pouvoir vous présenter mes travaux et connaître vos avis sur les sujets qui sont devenus tant importants à mes yeux durant la préparation de la thèse. En particulier je remercie M.Christian Soize d'avoir présidé le jury. Je suis reconnaissante à M.Benoît Noetinger d'avoir rapporté cette thèse, vous m'avez rappelé mes origines dans le monde de géosciences. Je remercie M.Bruno Sudret pour les précisions importantes apportées au manuscrit, ainsi que pour l'aperçu de perspectives. Je suis reconnaissante à M.Michel Panfilov de m'avoir préparé à cette thèse et d'avoir assisté à la soutenance.

Je remercie du fond du coeur mes encadrants de m'avoir guidé chacun dans son domaine d'expertise et d'avoir été patients lors de la rédaction et la correction du manuscrit. Je remercie mon directeur de thèse M.Didier Clouteau pour la culture scientifique et personnelle que j'ai gagnée à ses cotés. Je dis merci à M.Étienne de Rocquigny pour ces idées précieuses et le travail important de vérification de mes travaux; votre livre 'Modelling under risk and uncertainty' m'a servi de guide. Je remercie M.Régis Cottureau d'avoir appuyé mes recherches au quotidien et de m'avoir apporté un soutien irremplaçable. Et enfin un grand merci à M.Régis FARRET, l'initiateur du sujet de thèse et le moteur principale de son avancement. Je n'oublierai pas l'effort et le temps que tu as consacré à cette thèse et à ma formation. Merci!

Je remercie toute l'équipe du laboratoire MSSMat, en particulier Anne-Sophie Mouronval pour la formation et les secours informatiques, Fernando Lopez Cabalero, Arézou Modaressi et Denis Aubry pour la chaleur humaine et les conversations intéressantes. Je n'oublie pas les doctorants et les post-doctorants qui ont partagé mon sort, je vous aime!

Merci à toute l'équipe de la direction de risques sol et sous-sol d'INERIS.

Je suis reconnaissante envers les professeurs de l'Université d'État de Moscou et de l'ENSG de Nancy de m'avoir donné les bases nécessaires pour parvenir à effectuer mes recherches scientifiques.

Avec beaucoup de gratitude j'ai du accepter l'aide de ma famille pour achever le travail. Je pense à mes parents et beaux-parents, à ma grand-mère, à mon fils et à mon cher mari qui a rempli ma vie de joie. Vini, merci pour tout.

Abstract

In the recent years, the geological storage of carbon dioxide (CO₂) has been proposed as a promising technology of reducing its concentration in the atmosphere. The evaluation of risks related to the storage phase (starting with the injection in a deep aquifer and lasting up to thousands of years) is based on the accurate modeling of the simultaneous flow of the injected gas and the saline water initially occupying the pores of the medium. Such forecasts are highly dependent on our knowledge about the properties of the medium, both fluids and their interaction with the medium. Generally quite limited information is available on these characteristics, especially when the properties of the media such as layers geometry, permeability, porosity, fault locations and transmissibility are concerned. These and many other sources of uncertainty affect the estimates of risk.

The main goal of the thesis is to define how the uncertainty can be accounted for in the process of risk assessment for CO₂ storage and to quantify by means of numerical models the scenarii of leakage by lateral migration and through the caprock. It is found that for qualitative and semi-quantitative risk assessment methodologies, the uncertainty can only be accounted for *a posteriori* giving an approximate relative margin of the evaluated risk. For most quantitative methodologies it is possible to define the risk measures out of evaluated uncertainty on consequence intensity and probability.

The chosen scenarii have been quantified using the system modeling approach. This approach is most suitable for slow processes and provides a wide choice for uncertainty representation and propagation. To make a link with the classical risk model in the form of an Event Tree, the events are expressed as exceedence of critical thresholds (such as fluid overpressure, maximal lateral spread of CO₂ cloud, leakage rates,...). Three different levels of risk measures and critical thresholds are discussed depending on uncertainty representation.

Two predictive numerical models are developed: a 2D rotation-invariant model and coupled 3D/2D rotation-invariant one. The main phenomena accounted for in the numerical models are two-phase immiscible pressure-driven flow, capillarity, gravity and viscosity. These phenomena dominate the repartition of CO₂ in space for several hundreds of years.

The first model is used for probabilistic parametric uncertainty propagation with a stochastic metamodel (using a polynomial chaos expansion). Results are given in terms of threshold exceedence probability curves for two variables of interest: the lateral extent of the CO₂ cloud and fluid overpressure on top of the reservoir after 10 years of injection. It is found that the most influential input variables for the lateral extent are intrinsic permeability, maximal gas saturation and thickness of the reservoir whereas the total porosity shows no relevance. The parameter controlling gas overpressure is the intrinsic permeabil-

ity. It is shown that using a stochastic metamodel together with a Taylor approximation, the level-2 quantities of interest become accessible even if the numerical model is rather time-consuming.

The second model represents a 3D model enhanced by coupling with a 2D rotation invariant formulation far from the injection well. Such coupling makes the model lighter and diminishes the boundary effects. It has been created and used to quantify the effects of 3D permeability heterogeneities on the flow. The spatial variability is represented as an anisotropic tensor-valued random field. Special statistical estimators based on spatial averaging are then defined to reach the quantities of interest such as threshold exceedence probability. The results are compared to homogeneous permeability random model having the same probability density function. In the heterogeneous case a tremendous reduction of variance is observed. To find an equivalent homogeneous variability closely reproducing the heterogeneous case results, the input permeability variance should be divided by 5. It follows that the subjective decision to represent the input uncertainty with or without spatial variability changes significantly the results. Therefore, a multi-variable uncertainty propagation techniques treating spatial variability should be developed.

Résumé

Le présent travail fait le lien entre les domaines de modélisation des écoulements biphasiques (dans les milieux homogènes et hétérogènes), le traitement d'incertitudes et l'analyse de risques dans le contexte de stockage de CO₂ dans les aquifères profonds et salins. La technologie de stockage de CO₂ dans des couches géologiques consiste en un captage du dioxyde de carbone de l'air, sa condensation jusqu'à l'état supercritique et l'injection dans un milieu profond, naturellement poreux (à plus de 1 km de profondeur). L'évaluation de risques liés à la phase de stockage (à partir du moment où le CO₂ commence à être injecté dans l'aquifère) est basée sur une prédiction des flux simultanés du gaz injecté et de l'eau saline initialement présente dans les pores du milieu. Telles prédictions dépendent des propriétés du milieu, des caractéristiques des deux fluides et de leur interaction avec le milieu. Ces caractéristiques sont toutefois mal connues, particulièrement quand il s'agit des propriétés du milieu, telles que la géométrie des couches, la perméabilité, la porosité, la position et la transmissibilité des failles. Les sources d'incertitude mentionnées altèrent l'évaluation de risques.

L'objectif principal de la thèse est de définir comment l'incertitude peut être intégrée dans la procédure d'évaluation de risques dans le contexte de stockage de CO₂. Comme il existe une grande variété des méthodologies d'évaluation de risques, leur classification en fonction de la manière dont est prise en compte l'incertitude est étudiée. L'approche par modélisation de système permet le choix le plus flexible de représentation et de propagation des incertitudes dans l'hypothèse de processus lents. Le lien entre cette approche et un modèle classique de risques sous forme d'un arbre de causes et de conséquences est présenté dans ce travail. Les événements sont formulés en termes de dépassement de seuils critiques (pour la surpression des fluides, l'extension maximale du panache de CO₂, les taux de fuite,...). Trois niveaux des seuils sont définis en accord avec les trois niveaux de traitement d'incertitudes.

La partie appliquée de la thèse comprend le développement de modèles numériques prédictifs et une propagation d'incertitude probabiliste pour la quantification de l'extension maximale du panache de CO₂ et de la surpression au toit de l'aquifère. Les phénomènes inclus dans les modèles numériques sont un écoulement biphasique immiscible provoqué par une différence de pressions, la capillarité, la gravité et la viscosité. Ces facteurs définissent la forme du panache de CO₂ pendant la période d'avancement le plus rapide dans le milieu.

Une partie importante de la thèse est consacrée à la problématique de prise en compte de la variabilité spatiale comme une source d'incertitude. Une étude numérique probabiliste est menée sur un champ aléatoire anisotropique de la perméabilité intrinsèque. Cette étude est réalisée grâce à la création d'un modèle numérique *ad hoc* et la défini-

tion d'estimateurs statistiques pour les quantités d'intérêt. Les résultats sont comparés au cas d'une perméabilité homogène et aléatoire avec une même fonction de densité de probabilité. Une variabilité homogène équivalente est déduite permettant de reproduire les résultats du cas hétérogène.

Contents

1	Introduction	1
1.1	Context	1
1.2	Uncertainty and Risk assessment	2
1.3	Objectives and outline of the Thesis	6
2	Physical and numerical model of CO₂ injection into an aquifer	7
2.1	The overall model and the main phenomena	8
2.2	Two-phase flow in porous media	11
2.3	Closed system of PDEs for a numerical model in 3D formulation	15
2.3.1	The constitutive system of equations	15
2.4	Developed numerical models	17
2.4.1	2D-axisymmetric model	18
2.4.2	Coupling of 2D-axisymmetric and full 3D formulations	20
2.5	Existing simplified analytical models	24
3	Risk assessment for CO₂ storage and uncertainty propagation	27
3.1	Definitions	28
3.1.1	Components of risk management	28
3.1.2	Dimensions of risk	29
3.1.3	Risk scenarios	30
3.2	Review of risk assessment studies applied to CO ₂ storage	32
3.2.1	Types of risk analysis	32
3.2.2	Particularities of CO ₂ Storage risk assessment	34
3.2.3	Methodologies	35
3.3	Methodology used by INERIS for CCS risk assessment	37
3.3.1	The global workflow, the risk model in the form of an event tree	37
3.3.2	Central events and impacting phenomena for CO ₂ storage	38
3.4	From scenario to uncertainty propagation	41
3.4.1	The relation between scenario evaluation and uncertainty propagation	42
3.4.2	Deterministic, probabilistic and level-2 probabilistic quantities of interest and critical thresholds	43
3.5	Classical uncertainty propagation methods illustrated on a simplified physical model	48
3.5.1	Probabilistic analysis: global approach	49
3.5.2	‘Level-2’ probabilistic analysis	57

3.5.3	Comparison of the results	59
3.6	Concluding remarks	61
4	Quantification of scenarios using numerical models	63
4.1	Meta-models for maximal lateral extent and fluid overpressure	64
4.1.1	Uncertainty sources description	65
4.1.2	Uncertainty propagation	66
4.1.3	Sensitivity analysis	70
4.2	Level-2 uncertainty propagation using meta-model and Taylor approximation	72
4.3	Results for homogeneous permeability	74
4.4	Concluding remarks	76
5	Spatial variability	77
5.1	Geological heterogeneity	78
5.1.1	A matter of scale	78
5.1.2	Spatial variability and uncertainty related to its representation as a random field	79
5.2	Random tensor field of permeability: modelling and simulation results . . .	80
5.2.1	Theoretical basis of the generation of anisotropic random field of permeability	80
5.2.2	Simulated anisotropic random field of permeability	84
5.3	Probabilistic study for random tensor field of permeability	86
5.3.1	Results for media with heterogeneous orthotropic and anisotropic permeability	86
5.3.2	Statistics on the maximum radius with variability	88
5.4	Permeability uncertainty: homogeneous versus heterogeneous variation . .	97
5.4.1	Comparison of the results	98
5.4.2	Inference of the homogeneous equivalent permeability distribution .	101
5.5	Concluding remarks	104
	Conclusions and perspectives	105
A	Review of risk assessment methodologies and tools applied to CCS	113
B	Deterministic and local probabilistic results for the simplified physical model	117
B.1	Deterministic response surface	117
B.2	Local probabilistic sensitivity analysis	118
B.3	Analytical computation of the response PDF characteristics.	119
C	Polynomial Chaos Expansion theory	123
D	Meta-model results for seven uncertain inputs	127
E	The coupled model results	131

F Details of the inference of the homogeneous equivalent permeability distribution	133
Bibliographie	135

Notations

a, A - scalars

$\mathbf{a} = (a_1, \dots, a_n)$ - vector in the Euclidean space \mathbb{R}^n

\mathbf{A} - matrix or tensor

Uncertainty modelling

$\mathbf{x} = (x_1, \dots, x_p)$ - vector (of dimension p) of uncertain model inputs representing sources of uncertainties

\mathbf{d} - vector of fixed model inputs representing design or decision variables

$\mathbf{z} = (z_1, \dots, z_q)$ - vector (of dimension q) of the model output variables (variables of interest)

\mathbf{X}, \mathbf{Z} - vector of uncertain random variables corresponding to \mathbf{x} and \mathbf{z}

$G(\cdot)$ - deterministic function representing the system model relying the input vectors to the output variables of interest

$f_X(\mathbf{x}|\boldsymbol{\theta}_X)$ - joint density of random vector \mathbf{X} , parametrized by $\boldsymbol{\theta}_X$

$\boldsymbol{\theta}_X$ - vector of parameters of the uncertainty model of \mathbf{X} (such as parameters of the joint PDF in probabilistic approach)

Θ_X - vector of uncertain random variables corresponding to $\boldsymbol{\theta}_X$

$\pi(\boldsymbol{\theta}_X|\zeta)$ - joint density of random vector Θ_X , modeling epistemic uncertainty in $\boldsymbol{\theta}_X$, such a level-2 distribution is parametrized by the hyper-parameters ζ

Matrix sets

$\mathbb{M}_{n,m}(\mathbb{R})$ - set of all $(n \times m)$ real matrices

$\mathbb{M}_n(\mathbb{R})$ - set of all square $(n \times n)$ real matrices

$\mathbb{M}_n^+(\mathbb{R})$ - set of all square $(n \times n)$ real symmetric positive-definite matrices

Matrix invariants and norms

$\det\{\mathbf{A}\}$ - determinant of a square matrix $\mathbf{A} \in \mathbb{M}_n(\mathbb{R})$

$\text{tr}\{\mathbf{A}\}$ - trace of a matrix

$\|\mathbf{A}\|_F^2 = \text{tr}\{\mathbf{A}^T \mathbf{A}\} = \sum_{i=1}^n \sum_{j=1}^m A_{ij}^2$ - the Frobenius norm

Operators

∇ - gradient

div - divergence

Physical model

The subscript 'w' stands for the wetting fluid, here the salt water, 'nw' - for non-wetting fluid, here the injected supercritical CO₂.

The phase $\alpha =$ 'w' or 'nw'.

K - hydraulic conductivity [m/s]

K_α - effective permeability of the fluid α

k_i - intrinsic permeability of the medium [m²]

$k_{r\alpha}$ - relative permeability for the fluid α [-]

B_α - bulk modulus of the fluid α [Pa]

s_α - saturation of the fluid α [-]

S_α - effective saturation of the fluid α [-]

$s_{r\alpha}$ - residual saturation of the fluid α [-]

ρ_α - density of the fluid α [kg/m³]

ϕ - medium porosity [-]

P_{atm} - atmospheric pressure [Pa]

P_0 - reference pressure [Pa]

$p_{\alpha 0}$ - initial pressure of the fluid α [Pa]

Q_{nw} - gas injection rate [m/s]

Statistical notions

μ - mean value

σ - standart deviation

CV - coefficient of variation

Var - variance

Cov - covariance

$\gamma(h)$ - variogram

$R(h)$ - autocorrelation function

Acronyms

CCDF - complementary cumulative distribution function (exceedance probability curve)
CCS - carbon capture and storage
CDF - cumulative distribution function
CFA - certification framework assessment
CQUESTRA - risk and performance assessment code for geological sequestration of carbon dioxide
DRA - deterministic risk assessment
ICARAS - integrated carbon risk assessment
INERIS - national (French) institute of environmental and industrial risks
IPCC - international panel on climate change
FEP - features, events and processes
HAZOP - hazard and operability study
HSE - health, safety & environment
LOPA - layer of protection analysis
MAUT - multi-attribute utility theory
MIMAH - methodology of identification of major accident hazard
MCA - multi-criteria assessment
MOSAR - organized and systemic method of risk analysis
PCE - polynomial chaos expansion
PDF - probability density function
PRA - abbreviation for both probabilistic risk assessment and preliminary risk analysis
P&RTM - performance and risk assessment approach for well integrity
REV - representative elementary volume
RISQUE - risk identification and strategy using quantitative evaluation
QRA - quantitative risk assessment
SMA - system modeling approach
SRF - screening and ranking framework
SVM - support vector machine
SWIFT - structured 'What-if?' technique

Chapter 1

Introduction

1.1 Context

Carbon Capture and Storage (CCS) is a technique to reduce the air carbon dioxide concentration by means of its injection into geological medium. The CO₂ is captured chemically from the air in the vicinity of industrial emitters, compressed to supercritical state, transported to the injection well and injected into a natural underground porous formation.

This procedure allows to continue the use of fossil energies, giving necessary time for renewable energies to be widely implemented. The main potential receptors of the CO₂ are depleted oil and gas fields, unminable coal seams and deep saline aquifers. According to the Intergovernmental Panel on Climate Change (IPCC, 2005), among the listed receptors the deep saline aquifers have the biggest predicted capacity (see Table 1.1).

Reservoir type	Lower estimate of storage capacity [Gt] ^a	Upper estimate of storage capacity [Gt] ^a
Oil and gas fields	675 ^b	900 ^b
Unminable coal seams	3-15	200
Deep saline formations	1000	Uncertain, but possibly 10 ⁴

Table 1.1: Storage capacity of different reservoirs (IPCC, 2005)

Deep saline aquifers are deep porous sedimentary rocks saturated with brines (saline water). The high concentrations of salt contained in such brines prevents them from being used as water resources for human and agricultural needs, therefore the deep saline aquifers are good candidates to receive the CO₂ captured from the air.

CO₂ geological storage is a recent technology which has seen a rapid development over the last decade. It challenges the researchers to provide appropriate tools to compare and assess potential injection sites in terms of risk (Grobe et al., 2009) as well as to establish a conceptual framework of risk assessment methodology (Lahaie et al., 2009) and a regulatory framework (Solomon et al., 2007). The underground medium is hardly accessible and

^aStorage capacity includes storage options that are not economical.

^bThese numbers would increase by 25% if 'undiscovered' oil and gas fields were included in this assessment.

therefore poorly known. That is why evaluation of risks related to underground processes needs to integrate uncertainty.

1.2 Uncertainty and Risk assessment

A closer look at the relations between risk and uncertainty allows to assert that risk does not exist without uncertainty. Indeed, if the behavior of any system was known precisely for any time moment (i.e. could be described in deterministic way), the risk would be entirely controllable. Risk arises from the lack of knowledge that we have to describe or predict the reality. The International Organization for Standardization defines the risk as the “effect of uncertainty on objectives” (ISO Guide 73:2009, 2009).

In the light of the definition above, risk assessment consists in uncertainty treatment. Every methodology of risk assessment intrinsically relies on a certain uncertainty perception. Conscious or unconscious perception affects the risk evaluation by contributing to the subjectivity of any given risk study. Subjectivity is considered as one of the natures of uncertainty sources itself.

So, strictly speaking, to “integrate” uncertainty in risk assessment procedure means to adapt it to be able to include more advanced risk measures, uncertainty description and propagation. In other words, the practice of looking for an error interval related to a risk value could be replaced by a practice of evaluating uncertainty first and then defining the risk depending on how pessimistic the evaluation should be.

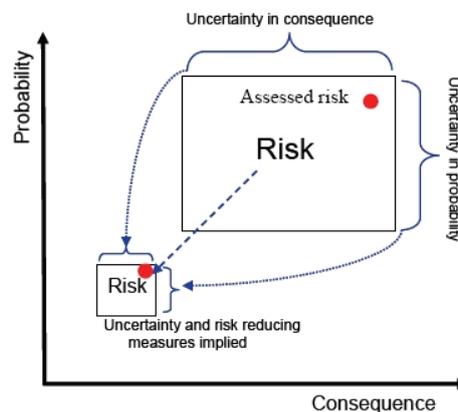


Figure 1.1: Illustration of risk and uncertainty management throughout the life of a CO₂ storage project. Reducing uncertainty shrinks the ‘risk box’, whereas reducing risk moves the ‘risk box’ toward the origin. The red dot illustrates how risk is assessed based on a conservative approach, i.e. a pessimistic evaluation of probability and consequence. If risks are ranked conservatively, any measures that reduce uncertainty will generally also reduce the assessed risk. (DNV, 2010, page 65)

What is uncertainty? The perception of uncertainty has been evolving with time and the domain of application determining the way of its treatment in risk assessment and giving a variety of approaches. At the same time the distinction of different types of uncertainty has appeared. It has become common to make difference of aleatory (or in other

terminologies “random”, “stochastic”, “natural variability”) uncertainty, which describes unpredictable character of natural processes and media, and epistemic (or “systematic”, “subjective”) uncertainty, which takes origin from incomplete knowledge. In practical studies the temporal and spatial scale of the problem make the distinction rather clear.

The typology of uncertainties adapted to the context of risk assessment is widely discussed in the PhD thesis of [Cauvin \(2007\)](#). The principle categories (without considering the relations between them) proposed by [Baecher and Christian \(2003\)](#) are as follows:

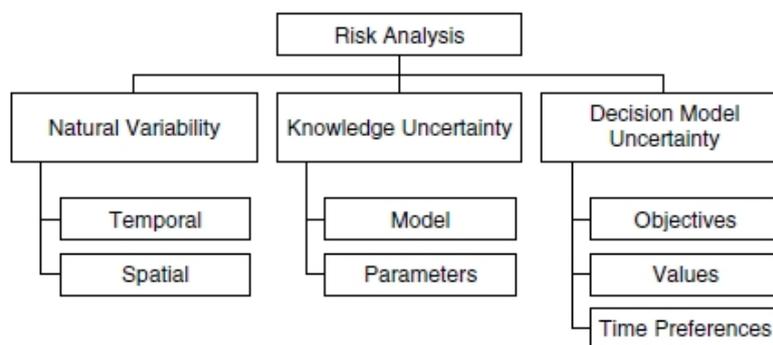


Figure 1.2: Categories of uncertainty entering risk analysis ([Baecher and Christian, 2003](#)).

These categories are not independent of course. The presence of natural spacial heterogeneities is partly in the origin of the imprecision of our knowledge about media properties (it would be much easier to interpret the measurements if they were independent of location). As well as it is impossible to consider exclusively natural variability, because its identification necessitates constraining to data affected by imprecision.

The parameter knowledge uncertainty and less natural variability constitute the main body of uncertainty sources commonly considered for uncertainty propagation in the predictive modeling context.

Nowadays, it is a common practice to engage predictive models in order to get quantitative estimates of the intensity of physical processes and consequently of risks. But “numerical models are a form of highly complex scientific hypothesis” ([Oreskes et al., 1994](#)). Modeling itself introduces uncertainty that should be quantified.

Briefly the stages of modeling process and associated uncertainties are the followings:

[Walker et al. \(2003\)](#) and [Refsgaard et al. \(2007\)](#) discuss a conceptual basis for uncertainty management in model-based decision support. In their terminology the delimitation of the system in time and space would be referred as ‘framing’, further transition from simplifications through physical model to mathematical model as ‘model structure uncertainty’, and transition from mathematical to numerical model results as ‘model technical uncertainty’. These sources of uncertainties put in the process of creation of consequent models is shown on Table 1.2.

For a reservoir engineering prediction (as can be considered CO₂ storage as well) ‘framing’ would correspond to the choice of the lateral extent of the simulated area, selection of the geological layers to be included in the simulation and selection of time period for the prediction. The input uncertainty consists in the uncertainty of the geometry

Context	
Reality	
	delimitation of the system in time and space, choosing geometry, simplifications of the phenomena at stake, hypothesis, neglects
Physical model	
	approximations of empirical laws, mathematical expression
Mathematical model	
	discretization, choice of a numerical resolution scheme, parametrization of the solver, numerical approximations, errors
Numerical model	
	+ Parameter uncertainty (possibly including representation of natural variability)
= Probabilistic model	

of the static geological model: facies thicknesses, fault locations, etc. This information is provided by geologists. Non-uniqueness of data interpretation by geologists always leads to a set of equivalent models of the geological system which could give substantially different results for flow dynamic simulation. Therefore, reservoir engineers are tempted to treat this type of uncertainty, working simultaneously with several interpretations of the geological model.

Context	
Reality	
	Context and framing
	Input uncertainty
	Model structure uncertainty (approximation)
Physical model	
	Model structure uncertainty (simplification)
Mathematical model	
	Model technical uncertainty
Numerical model	
	+Parameter uncertainty
= Probabilistic model	

Table 1.2: Uncertainties associated with the modeling process: sources of uncertainties according to [Walker et al. \(2003\)](#) and [Refsgaard et al. \(2007\)](#)

Model structure uncertainty (incomplete understanding or simplified description of modeled processes) can be demonstrated by comparison of the results of solutions provided by different institutions for the same problem (although the effect would be coupled with the model technical uncertainty, which could be estimated by comparing the results

for a numerical benchmark: the same mathematical model solved by different numerical simulators). The work on benchmarks for two-phase flow in porous media has been conducted by the MoMas group (MoMas, 2012). A framework for dealing with uncertainty due to model structure error is proposed by (Refsgaard et al., 2006). The tenability of CO₂ fate prediction models can be evaluated through a pedigree matrix. The average score of pedigree matrix on the scale from 0 to 4 is 2 (it could rise to 3). As an example, the matrix for the model engaged in the present work would be following: Model exactly addresses the desired predictive variable (Proxy, score 4), is based on indirectly measured data or poor historical/field data (Quality and quantity, score 2), built on the accepted theory with partial nature in view of the phenomenon it describes (Theoretical understanding, score 3), described by equations reflecting acceptable mechanistic process detail (Representation of understood underlying mechanisms, score 3, can go down to aggregated parametrized meta-model for computational needs, score 2), somewhat plausible or reasonably plausible (Plausibility, score 2 or 3), accepted by all colleagues except rebels: competing theories exist, such as streamline simulation, but are not controversial (Colleague consensus, score 3).

The model technical uncertainty constitutes a great body of work of numerical studies (Díez et al., 1998). Some types of technical uncertainties can be expressed exactly, some can be estimated or bounded. In the present work the numerical uncertainty was only addressed by comparing the results of 2D axisymmetric and 3D formulations and simulations on different meshes (see section 2.4.2).

That is why nowadays to perform an uncertainty study one needs to make appropriate choices. A guide to quantitative uncertainty management by De Rocquigny et al. (2008) provides a generic workflow for uncertainty treatment as well as an assistance for making best choices at each step of the workflow. In the present work the ideas of uncertainty management are adopted from the guide cited above. The proposed generic workflow divides an uncertainty study into four steps:

- A. The specification step: What are the quantities of interest to be estimated? Which model is going to be used? What are the inputs and the outputs of this model?...
- B. The uncertainty modeling (or uncertainty source quantification) step: How the uncertain nature of the inputs can be represented?
- C. The uncertainty propagation step: supposing that the uncertainty of the output arises from the uncertainty in inputs, how to compute the uncertainty of the output, the quantity of interest and even the uncertainty in the quantity of interest?
- C'. The sensitivity analysis (or importance ranking) step: supposing again that the uncertainty of the output arises only from the uncertainty in inputs, which is the contribution of each input to the output uncertainty?

The importance of addressing uncertainties in the risk assessment models concerning the geologic storage of CO₂ is stressed in the IPCC special report. A study for the IEA Greenhouse gas programme (IEA GHG, 2007) also concluded that a better understanding is required for the process of conducting a site performance site, for the management of liability and the quantification of site-specific impacts of a CO₂ release and the estimation

of its probability. The listed detected gaps are closely related to the understanding of the uncertainties. The need to handle various kinds of uncertainty is mentioned by numerous authors (e.g. [Stenhouse et al., 2006](#)).

1.3 Objectives and outline of the Thesis

Taking into account the features formulated above, the objectives of the thesis are:

1. To determine how uncertainties impart the risk assessment methodologies currently used in the context of CO₂ storage.
2. In the frame of the methodology followed at INERIS, to contribute to the scenario quantification (based on uncertainty treatment).
3. To develop numerical models allowing of a probabilistic uncertainty propagation and account of spatial variability.
4. To choose the most appropriate uncertainty propagation method and sensitivity analysis technique and apply them using a numerical model.

The document is organized in five chapters. Chapter 2 is devoted to the description of the physical, numerical and semi-analytical models used for prediction of CO₂ migration through an aquifer. Coupling of a 3D model with a rotation-invariant one is presented.

In Chapter 3 the role of uncertainty in risk assessment is further discussed. The scenarios quantified in this work are presented. A preliminary analysis on a semi-analytical model illustrates the choice of an uncertainty propagation technique.

Chapter 4 is entirely devoted to the results of uncertainty propagation and sensitivity analysis performed on 2D rotation-invariant model of CO₂ injection into an aquifer through a PCE (Polynomial Chaos Expansion) metamodel.

In the last Chapter the effect of spatial variability is analyzed. Two probabilistic studies with homogeneous parametric uncertainty and spatial variability are compared.

Chapter 2

Physical and numerical model of CO₂ injection into an aquifer

Contents

2.1	The overall model and the main phenomena	8
2.2	Two-phase flow in porous media	11
2.3	Closed system of PDEs for a numerical model in 3D formulation	15
2.3.1	The constitutive system of equations	15
2.4	Developed numerical models	17
2.4.1	2D-axisymmetric model	18
2.4.2	Coupling of 2D-axisymmetric and full 3D formulations	20
2.5	Existing simplified analytical models	24

In this chapter the main phenomena taking place after injection of CO₂ into an aquifer are discussed. The formulation of the problem in the form of a closed system of partial differential equations is explained. Further, the particularities of numerical models developed during the preparation of the thesis are presented including geometry, values of parameters and expressions for variables, mesh characteristics, resolution scheme. In order to be able to treat the heterogeneities in 3D an *ad-hoc* model has been elaborated. It couples a solution in a 3D cylindrical zone with an axisymmetric solution outside. The justification of the model can be found in Subsection 2.4.2. Analytical models can be useful for preliminary risk screening and ranking. In the last part of the present chapter a literature review of existing analytical models for CO₂ storage is proposed. The model for maximal lateral extent by Nordbotten and Celia (2006) is detailed as it is used in the next chapter for illustration purpose.

2.1 The overall model and the main phenomena

The term of ‘physical model’ is commonly referred to as a smaller (or larger) physical copy of an object designed for experimental needs. In the present work, we are speaking of ‘physical model’ as a conceptual model of interacting physical processes taking place in a spatially and temporally determined system. In practical sense, it is a certain choice of geometry and modeled phenomena, of set of simplifications and hypothesis that give access to mathematical expression.

In the context of CO₂ geological storage, the system includes the storage aquifer and surrounding layers. For some models, all upper layers until the surface are considered. Depending on the available information, the geometry is either fully represented or simplified to horizontal or slightly inclined parallel layers. According to Gasda et al. (2008), for many aquifers of low permeability which are likely to become CO₂ sequestration the effect of slope can be neglected. In the present work, the model is inspired by an aquifer in Paris basin without any precise geometry available and with a single injector, that is why the choice of horizontal parallel cylindrical layers is made.

The lateral boundary conditions for analytical and semi-analytical solutions are often supposed infinite as for Celia and Nordbotten (2011); Okwen et al. (2011); Mathias et al. (2009); Fučík et al. (2007). For numerical models the domain is laterally bounded. The choice of lateral boundary condition influences significantly the solution and should correspond to the aquifer type (Smith et al., 2011; Zhou et al., 2008; Vilarrasa et al., 2010). For open aquifers, a condition of constant pressure is imposed laterally at a distant boundary assuming that there is a limit of pressure perturbation. This is discussed in Section 2.4.2 and a technique allowing to move further the lateral boundary is proposed.

What happens when the supercritical CO₂ is injected in an aquifer? Due to high injection pressure the supercritical fluid goes beyond the entry pressure and penetrates the largest pores of the medium driving out the saline water initially in place. Gradually CO₂ penetrates the pores of smaller radius and occupies more and more volume of the connected porosity. Due to the difference in densities (the injected fluid is lighter than the fluid in place), the CO₂ cloud has a tendency not only to spread laterally but also to rise slowly. As time goes more and more of CO₂ dissolves in water and becomes available for

chemical reactions with the matrix. The commonly accepted opinion is that the brine is lighter than the brine with dissolved carbon dioxide inside. Therefore, at the late stages of the storage the gravity flows for liquid phase should be considered. Nevertheless, certain experimental work do not confirm this point of view.

The most complete model would take into account the following physical and physico-chemical phenomena:

- flow of two fluid phases mainly driven by the injection pressure,
- capillary effect: wetting (w) and non-wetting (nw) phases - brine and gas respectively,
- imbibition/drainage hysteresis
- compressibility of fluids,
- compressibility of the porous medium (pore volume is not constant),
- buoyancy (gravity effect),
- temperature change (thermodynamical effects),
- partial miscibility between phases (naturally gas dissolves in water but also vice versa the water left after the gas drainage is able to evaporate forming so called drying front with inevitable salt precipitation),
- diffusion of dissolved species,
- disappearance of dissolved gas due to the chemical reaction with rock or with other components dissolved in water,
- alteration of the porous medium with time: the properties of the porous medium such as porosity, permeability and compressibility are not invariant due to mobile particles transport and precipitation, mechanical stresses provoking micro-cracking and chemical reactions.

All together these phenomena form a coupled dynamic chemo-thermo-hydro-mechanical (-migratory) problem which to the knowledge of the author has never been solved in its entirety because of the high number of unknowns, difficulty of coupling formalization (in particular for numerical resolution scheme) and difference of scales.

Nevertheless, this modeling problem is not completely new as similar underground processes and phenomena appear in applied science fields such as underground water resources treatment, gas storage, nuclear waste sequestration, oil and gas reservoir engineering and some geomechanical problems. Therefore, the choice of acting phenomena for CO₂ storage modeling is historically strongly related to the field of knowledge which this modeling is based upon: underground water modeling together with geochemistry emphasize the chemical aspects (reactive transport models as for (Zheng et al., 2009)), whereas reservoir engineering provides more experience in thermo-hydrodynamics for multiphase flow, geomechanical studies introduce the coupling with the mechanical response of the

system, nuclear waste sequestration has advances in thermo-hydro-mechanical coupling (Hudson et al., 2005; Fall et al., 2014).

Numerous works have recently appeared covering the chemical aspects of the CO₂ storage. Chemical modeling focuses on the last three phenomena of the above list. That is to say, its goal is predicting of the reaction rates, transport of the dissolved components and generally spatial distribution of concentrations. Some works include the mineralogical description of the reservoir rock. To quote some studies dealing with chemical aspects: Andre et al. (2007); Durst and Kervvan (2011); De Lucia et al. (2011); Dalkhaa et al. (2013); Shen et al. (2013); Xu and Li (2013); Kirste (2013); Mohd Amin et al. (2014). The chemical modeling is quite important for predicting of CO₂ evolution on the short term as well as on the long term because the chemical reaction is recognized as the ultimate and therefore most secure type of trapping (IPCC, 2005). For fully coupled flow-reactive-transport formulation, the reader is referred to the works of Fan et al. (2012); Liu et al. (2011); Durst and Kervvan (2011).

One of the most important factors for secure CO₂ storage is the integrity of the caprock. As the injection modifies the pressure field, there can be a critical pressure which would lead to the caprock fracturing or reactivation of existent fractures. That is why the geomechanical aspects have received a considerable attention as well (Le Gallo, 2009; Shi et al., 2013; Yamamoto et al., 2013; Rohmer and Seyedi, 2010; Rutqvist et al., 2008; Rutqvist, 2012; Vidal-Gilbert et al., 2009; Vilarrasa et al., 2010). At INERIS the numerical approaches for hydro-mechanical coupling were studied by Millard et al. (1995), Souley and Thoraval (2011), Thoraval et al. (2012) and in application to CO₂ storage by Thoraval (2008) and Vidal-Gilbert and Thoraval (2007).

Significant experimental efforts have also been focused on characterization of each one of the phenomena mentioned above both *in-situ* and in laboratory. The relative permeability is studied with the use of new technology: X-ray CT scanner (Perrin et al., 2009). The mobilization of the metals due to CO₂ injection has been studied by Rillard et al. (2013); Kharaka et al. (2009). The changes in porous structure in shales due to CO₂ injection can be found in (Rhenals Garrido et al., 2013). Cornet and Burlet (1992) have worked on the stress field determination.

To summarize, CO₂ has various means of flowing or being transported once injected. First of all, it flows in supercritical state. On its way, a part of the fluid dissolves in water and it is transported by the moving water. Inside the liquid phase CO₂ diffuses equilibrating the concentration gradient. While in liquid phase some quantity of CO₂ disappears due to chemical reactions.

These processes correspond to four trapping mechanisms determining the long-term evolution of the CO₂:

- structural and stratigraphic trapping
- residual gas trapping
- solubility trapping
- mineral trapping

The time scales of these processes are not the same. Some studies show that the dissolution, diffusion and chemical reactions are likely to be preponderant after thousand years (IPCC, 2005). The primary CO₂ spatial distribution in the aquifer is conditioned by the transport in supercritical state (Bachu et al., 1994). Therefore, as the development of a most complete and fully coupled model is not the topic of the present work, only the hydrodynamical model of two-phase compressible flow through incompressible porous medium has been considered. The reference time of 100 years is taken to quantify the influence of uncertainties.

2.2 Two-phase flow in porous media

We first remind how a fluid flow through porous media can be expressed for a single fluid.

Darcy's law for a single fluid

There are several approaches to describe quantitatively a fluid flow through porous medium depending on the scale of interest.

On the smallest scale the fluid flow can be quantified out of particle interactions. The Navier-Stokes equations describe the balance of forces acting at any given region of the fluid: changes in momentum of fluid particles depend only on the external pressure and internal viscous forces acting on the fluid. Such an approach seems fundamental and rigorous but can only be used for pore-scale models for short time periods. As it was already mentioned, the CO₂ storage involves large spatial and temporal dimensions - several tens of kilometers and thousands of years. Therefore, macro-scale laws of fluid flow are required. The classical law which is used to model fluid flow on macro-scale was initially formulated for saturated flow in porous media in an experimental way by Darcy (1856). The historical insight of this discovery is described by Marle (2006).

Darcy's law (Equation 2.1) states that the fluid velocity field (q) is determined by the pressure gradient (∇p), the fluid viscosity (μ), and the structure of the porous medium (expressed through the intrinsic permeability of the medium k_i):

$$q = -\frac{k_i}{\mu}(\nabla p - \rho g \nabla z). \quad (2.1)$$

Nowadays it is proved that the Darcy law can be obtained in a mathematically rigorous way via homogenization of Navier-Stokes equations (Allaire, 1992; Das and Hassanizadeh, 2005; Panfilov, 2000). Darcy's law applies when the gradient in hydraulic potential drives fluid movement in the porous medium. But this generalized linear flux law is only valid for slow, non-turbulent viscous flow. The state of the art in modeling long-term underground gas behavior acknowledges the use of Darcy's law.

Darcy's law for two phases

The formula above provides the momentum of a single fluid through porous medium. Nevertheless, Darcy's law is also largely used to describe a two-phase flow (e.g. in reservoir engineering, the main principles can be found in (Dake, 1978).

If we assume that the pores are entirely filled with two fluids, the volume fractions of each of the phases are introduced in order to find their velocities.

Wetting fluid saturation (s_w) and the non-wetting fluid saturation (s_{nw}) sum to one. Some part of the porosity is not available for the flow, which means that there is a residual saturation: when injecting gas, a part of the initial water remains bounded to the matrix. This leads to define:

- s_{rw} , the wetting fluid (water) residual saturation,
- S_w , the wetting fluid effective saturation,
- s_{rnw} , the non-wetting fluid residual saturation (which is supposed zero) and
- S_{nw} , the non-wetting fluid effective saturation.

The link between the saturation, the effective saturation and the residual saturation for the wetting phase (analogously for the non-wetting phase) is the following:

$$S_w = -\frac{s_w - s_{rw}}{1 - s_{rw} - s_{rnw}} \quad (2.2)$$

Hereafter only the effective saturation is going to be used.

The simultaneous flow of two fluids can be described as a combination of two separate flows (each of them is caused by the difference of fluid pressure and elevation according to 2.1). The contribution of each phase flow to the total flow is introduced through relative wetting and non-wetting phase permeabilities (kr_w and kr_{nw} respectively).

The velocity field of each phase according to extended Darcy's law is expressed as the wetting phase flow rate (q_w) and the non-wetting phase flow rate (q_{nw}):

$$q_w = -\frac{k_i k_{rw}}{\mu_w} (\nabla p_w - \rho_w g \nabla z), \quad (2.3)$$

$$q_{nw} = -\frac{k_i k_{rnw}}{\mu_{nw}} (\nabla p_{nw} - \rho_w g \nabla z). \quad (2.4)$$

For notations and values of the parameters chosen for the numerical model the reader can refer to Tables 2.2, 2.3 and 2.4.

The intrinsic permeability is by definition the permeability of water-saturated medium (for single-phase flow). It is measured experimentally by reversing the Darcy's law (Lake et al., 2007).

The gas relative permeability at maximum gas saturation is lower than the water relative permeability at maximal water saturation. According to Bennion and Bachu (2008), the ratio of the relative permeability endpoints for the studied samples is between 0.15 and 0.6. An example of experimentally measured relative permeability curves for a sample of most similar properties to Paris basin aquifer is shown in Figure 2.1.

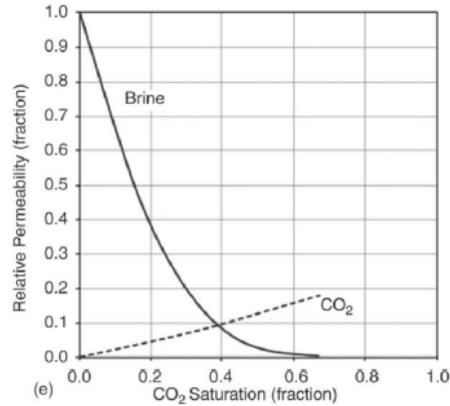


Figure 2.1: Relative permeability data (drainage cycle only) for CO₂/brine at *in-situ* conditions for the sample “Nisku #1”: carbonate at the depth of 2050 m, porosity of 9.7%, pressure of 17.4 MPa and temperature of 56 °C (from Bennion and Bachu, 2008).

Fluid mass conservation laws. The main equation allowing to model fluid flow is the mass conservation law. Its detailed derivation is demonstrated by Bear (1972, 1979).

For saturations defined above the mass conservation law for both phases reads:

$$\phi \frac{\partial(S_w \rho_w)}{\partial t} + \text{div}(\rho_w q_w) = 0, \quad (2.5)$$

$$\phi \frac{\partial(S_{nw} \rho_{nw})}{\partial t} + \text{div}(\rho_{nw} q_{nw}) = 0. \quad (2.6)$$

Even if the densities can be expressed through equation of state, there are still four unknowns in these equations: saturations and pressures of both phases. But there exists a relation between the saturation and the pressures difference. Therefore, in order to be able to solve the problem, two main unknowns should be chosen. This choice is important for numerical stability and the treatment of gas appearance. Some authors claim that choosing the gas pressure and water saturation as principal variables improves the convergence and numerical stability (Tong et al., 2013). In the present work, the pressures of the gas and liquid phases are chosen as two main unknowns. To obtain the equations with the pressure as the principal variable it is needed to express the first summand of Equations 2.5 and 2.6 as function of the capillary pressure change.

The capillary pressure is by definition the difference between the pressures of the non-wetting and wetting phases:

$$p_c = p_{nw} - p_w \quad (2.7)$$

Let us introduce a specific capacity $C_{p,w}$, which would describe how the effective saturation changes with capillary pressure. If we define the specific capacity as a partial derivative of the effective saturation with respect to the capillary pressure weighted by porosity, then such capacities for wetting and non-wetting phases are complementary:

$$C_{p,w} = -C_{p,nw} = \phi \frac{\partial(S_w)}{\partial p_c} \quad (2.8)$$

The water compressibility is the inverse of the water bulk modulus and can be expressed by definition as:

$$\frac{1}{B_w} = -\frac{1}{V} \frac{\partial V}{\partial p_w} \quad (2.9)$$

Using Equations 2.8 and 2.9, the first summand of Equation 2.5 can be developed as follows:

$$\begin{aligned} \phi \frac{\partial(S_w \rho_w)}{\partial t} &= \phi \rho_w \frac{\partial S_w}{\partial t} + \phi S_w \frac{\partial \rho_w}{\partial t} = \phi \rho_w \frac{\partial S_w}{\partial p_c} \frac{\partial p_c}{\partial t} + \phi S_w \frac{\partial \rho_w}{\partial p_w} \frac{\partial p_w}{\partial t} = \\ &= C_p \rho_w \frac{\partial p_c}{\partial t} + \phi S_w \frac{\partial \rho_w}{\partial V} \frac{\partial V}{\partial p_w} \frac{\partial p_w}{\partial t} = C_p \rho_w \frac{\partial p_{nw}}{\partial t} - C_p \rho_w \frac{\partial p_w}{\partial t} + \phi S_w \left(-\frac{\rho_w}{V}\right) \frac{\partial V}{\partial p_w} \frac{\partial p_w}{\partial t} = \\ &= C_p \rho_w \frac{\partial p_{nw}}{\partial t} + \left(-C_p \rho_w + \frac{\phi S_w \rho_w}{B_w}\right) \frac{\partial p_w}{\partial t} \end{aligned} \quad (2.10)$$

Similarly the first summand of Equation 2.6 can be expressed as follows:

$$\phi \frac{\partial(S_{nw} \rho_w)}{\partial t} = \rho_{nw} \left(C_p \frac{\partial p_w}{\partial t} + \left(-C_p + \frac{\phi S_{nw}}{B_{nw}}\right) \frac{\partial p_{nw}}{\partial t} \right) \quad (2.11)$$

On the right-hand side of Equations 2.10 and 2.11 we still have the water and gas saturations. There are a number of empirical laws describing the relation between capillary pressure and phase saturation as well as phase permeability and phase saturation (Brooks and Corey, 1964; Mualem, 1976; Van Genuchten, 1980). In the simulations presented in this work, the effects of hysteresis are neglected and the well-known Van Genuchten (1980) model for capillary pressure $p_c(S_w)$ is used. Inverting this law leads to the expression of saturation as a function of capillary pressure:

$$S_w(p_w, p_{nw}) = \begin{cases} (1 + (\alpha H_c)^N)^{-M}, & H_c > 0 \\ 1, & H_c \leq 0 \end{cases} \quad (2.12)$$

where H_c is the pressure head (Equation 2.13), α is the inverse of the ‘pseudo air entry pressure’ p_0 (Croisé et al., 2006) weighted by the gravity acceleration and the density (see Equation 2.14), N and M are non-dimensional characteristic parameters of the law.

$$H_c(p_w, p_{nw}) = \frac{p_{nw} - p_w}{\rho_w g} \quad (2.13)$$

$$\alpha = \frac{\rho_w g}{p_0} \quad (2.14)$$

The parameter p_0 is larger for finer material, and its dependency on material properties may be assessed using the Leverett (1941) scaling law (the procedure is described by Nikolaevskij (1990)). This law has the following form:

$$p_0 = \sigma \sqrt{\frac{k_i}{\phi}} \quad (2.15)$$

where σ is the interfacial tension between the two phases.

Finally, we have all the parts of the system of equations entirely expressed in terms of pressure variation. The final problem statement is given in the next section.

2.3 Closed system of PDEs for a numerical model in 3D formulation

2.3.1 The constitutive system of equations

Geometry

For the sake of simplicity we are considering a domain which represents only one geological layer of cylindrical geometry, having the vertical injector as the symmetry axis. Adding other layers would not change the formulations presented hereafter but would only modify the values of physical properties inside each layer. So let us consider a three-dimensional cylindrical domain Ω with a boundary $\partial\Omega$. A schematic view on the domain before revolution around symmetry axis $\mathbf{1z}$ is given in Figure 2.2. The part of the boundary where the injection takes place is noted as $\partial\Omega_{in}$, the opposite side of the border is $\partial\Omega_{out}$.

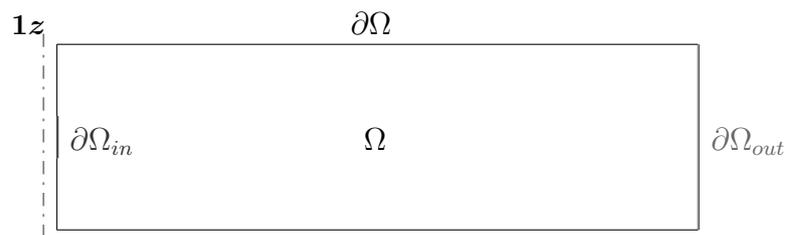


Figure 2.2: Considered domain: horizontal cylindrical layer with the injection well as the revolution axis.

Final mathematical problem statement

The final formulation of the problem is the following: Find fluid pressures p_w and p_{nw} defined on $\Omega \times [0, T]$ such that the systems of equations 2.16 and 2.17 are satisfied.

For the wetting phase:

$$\begin{aligned}
\rho_w(C_p \frac{\partial p_{nw}}{\partial t} + (-C_p + \frac{\phi S_w}{B_w}) \frac{\partial p_w}{\partial t}) + \operatorname{div}(-\frac{k_i k_{rw}}{\mu_w} (\nabla p_w - \rho_w g \nabla z)) &= 0, & \Omega \times [0, T] \\
p_w &= p_{w0}, & \Omega \times (t = 0) \\
n \cdot (-\frac{k_i k_{rw} \rho_w}{\mu_w} \nabla (p_w - \rho_w g z)) &= 0, & \partial\Omega_{in} \times [0, T] \\
n \cdot \nabla p_w &= 0, & \partial\Omega \times [0, T] \\
p_w &= p_{w0}, & \partial\Omega_{out} \times [0, T].
\end{aligned} \tag{2.16}$$

For the non-wetting phase:

$$\begin{aligned}
\rho_{nw}(C_p \frac{\partial p_w}{\partial t} + (-C_p + \frac{\phi S_{nw}}{B_{nw}}) \frac{\partial p_{nw}}{\partial t}) + \operatorname{div}(-\frac{k_i k_{rnw}}{\mu_{nw}} (\nabla p_{nw} - \rho_w g \nabla z)) &= 0, & \Omega \times [0, T] \\
p_{nw} &= p_{nw0}, & \Omega \times (t = 0) \\
n \cdot (-\frac{k_i k_{rnw} \rho_{nw}}{\mu_{nw}} \nabla (p_{nw} - \rho_{nw} g z)) &= Q_{nw} \rho_{nw}, & \partial\Omega_{in} \times [0, T] \\
n \cdot \nabla p_{nw} &= 0, & \partial\Omega \times [0, T] \\
p_{nw} &= p_{nw0}, & \partial\Omega_{out} \times [0, T].
\end{aligned} \tag{2.17}$$

To close the system of equations, the hydraulic properties are expressed as functions of the hydraulic head (H_c) according to the van Genuchten-Mualem laws (Van Genuchten, 1980):

$$H_c(p_w, p_{nw}) = \frac{p_{nw} - p_w}{\rho_w g} \tag{2.18}$$

$$S_w(p_w, p_{nw}) = \begin{cases} (1 + (\alpha H_c)^N)^{-M}, & H_c > 0 \\ 1, & H_c \leq 0 \end{cases} \tag{2.19}$$

$$C_p(p_w, p_{nw}) = \begin{cases} \frac{\alpha M}{1-M} (\phi - \theta_r) S_w^{1/M} \frac{(1 - S_w^{1/M})^M}{\rho_w g}, & H_c > 0 \\ 0, & H_c \leq 0 \end{cases} \tag{2.20}$$

$$\theta_w(p_w, p_{nw}) = \begin{cases} \theta_r + S_w(\phi - \theta_r), & H_c > 0 \\ \phi, & H_c \leq 0 \end{cases} \tag{2.21}$$

$$k_{rw}(p_w, p_{nw}) = \begin{cases} S_w^L (1 - (1 - S_w^{1/M})^M)^2, & H_c > 0 \\ 1, & H_c \leq 0 \end{cases} \tag{2.22}$$

Hydraulic properties for the non-wetting phase:

$$S_{nw}(p_w, p_{nw}) = 1 - S_w, \tag{2.23}$$

$$\theta_{nw}(p_w, p_{nw}) = \phi - \theta_w \quad (2.24)$$

$$k_{rnw}(p_w, p_{nw}) = \begin{cases} (1 - S_w)^{LL} \cdot (1 - S_w^{1/M})^{2M}, & H_c > H_c^o \\ 0, & H_c \leq H_c^o \end{cases} \quad (2.25)$$

It should be mentioned that the expressions for relative permeabilities (Equations 2.22 and 2.25) are normalized. In order to obtain a non-normalized curve, the gas relative permeability is multiplied by its endpoint 0.2 (value taken from the closest sample analyzed by Bennion and Bachu (2008), see Figure 2.1)

Defining the system in phase pressures as main unknowns was initially implemented in the ‘simultaneous solution’ scheme in petroleum reservoirs (Douglas et al., 1959).

The main criticisms of this approach are the strong coupling and non-linearity of the first equations in systems 2.16 and 2.17 and the assumption that the capillary pressure has a unique inverse function for expressing saturation (Equation 2.19). Alternative ways of defining the system are the implicit pressure-explicit saturation (IMPES) scheme and the formulation in a global pressure, described by Chen et al. (2006).

2.4 Developed numerical models

All the numerical models developed in the present project for flow prediction during and after CO₂ injection into an aquifer are based on the same assumptions. The geological context is inspired by Paris sedimentary basin. Dogger reservoir is considered as a potential storage aquifer.

The main two models used for comparative numerical study in the present work are a 2D axisymmetric model of two-phase immiscible flow in the aquifer with a vertical injection well as a rotation axis and a similar 3D model completed with 2D axisymmetric solution for distant zones.

Other developed models have not found their application in the thesis but can be further adapted for evaluation of different scenarios. Among them a 2D rotation-invariant model including three geological layers: aquifer, caprock and overburden; a 2D rotation-invariant model for the aquifer with anticline geometry; a coupled 3D/2D axisymmetric model for three layers and for horizontal injection well (need further numerical optimization). The author also contributed to the verification of a vertical 2D model predicting the leakage rates through the entire column of overlaying formations.

Table 2.1: Model input parameters

Parameters	Values
Reservoir depth	[-2470 m, -2350 m]
Injection zone	[-2420 m, -2400 m]
Initial pressure (hydrostatic)	23.5 MPa (at the top of the reservoir)
Temperature	75°C (for thermal gradient 0.03°C/m)
Boundary conditions	no flux vertically, constant pressure laterally

Variables and parameters The two principal variables to be found by the numerical solution are the fluids pressure fields for any time instant from the beginning of the injection and till 100 years. The constant model parameters describing the aquifer conditions are summarized in Table 2.2.

Table 2.2: Parameters of the model

symbol	value	unit	name
Q_{nw}	$3.68 \cdot 10^{-3}$	m/s	gas injection rate
ρ_w	1000	kg/m ³	water density (surface)
ϕ	0.15	[-]	total porosity
θ_r	0.05	[-]	irreducible water volume fraction
r_1	0.108	m	well radius
r_2	1	m	damaged zone thickness
P_{atm}	10^5	Pa	atmospheric pressure
P_0	10^5	Pa	reference pressure
K_0	10^{-13}	m ²	intrinsic permeability far from the well
M	0.66	[-]	van Genuchten parameter
N	0.34	[-]	van Genuchten parameter
L	0.5	[-]	van Genuchten parameter
LL	0.9	[-]	van Genuchten parameter
α	$\frac{\rho_w \cdot g}{P_0} = 0.0981$	1/m	van Genuchten parameter
kr_{max}	0.2	[-]	gas relative permeability endpoint
g	9.81	m/s ²	gravity acceleration
K_w	$2 \cdot 10^9$	Pa	water bulk modulus

2.4.1 2D-axisymmetric model

In a homogeneous case for simplified geometry of horizontal superposed layers an injection of fluid through a single vertical well can be modeled in axisymmetric setting.

Table 2.3: Variables

name	unit	description	formula	order of magnitude
k_i	m ²	intrinsic permeability with a damaged zone around the well	$\begin{cases} K_0, & r > r_1 + r_2 \\ K_0(10 - 9\frac{r-r_1}{r_2}), & r \leq r_1 + r_2 \end{cases}$	10^{-13}
p_{nw0}	Pa	initial gas pressure	$-\rho_w g z + P_{atm}$	235×10^5
p_{w0}	Pa	initial water pressure	$-\rho_w g z(1 - 10^{-4}) + P_{atm}$	235×10^5

Gas injection rate Q_{nw} corresponds to injection debit of 1Mt/year for the gas density at the middle reservoir depth. For the gravity acceleration, the average on the Earth's surface is taken. At the depth of 1-2 km the gravity acceleration does not change significantly according to the Preliminary Reference Earth Model. Brine density is approximated by water density at surface level. The major simplification made for the construction of the

numerical model is the approximation of fluid properties as functions of depth. It allows taking into account the temperature and pressure gradient. The functions (Table 2.4) are entered in the FEM software by interpolating between point-wise values taken from on-line calculator of CO₂ properties www.carbon-dioxide-properties.com based on the work of Fenghour et al. (1998) and Scalabrin et al. (2006). The resulting functions are presented in Figure 2.3.

Table 2.4: Interpolated functions of the model

name	unit	description
$\rho_{nw}(-z)$	kg/m ³	non-wetting fluid density
$\mu_{nw}(-z)$	Pa·s	non-wetting fluid viscosity
$\mu_w(-z)$	Pa·s	wetting fluid viscosity
$K_{nw}(-z)$	Pa	non-wetting fluid bulk modulus

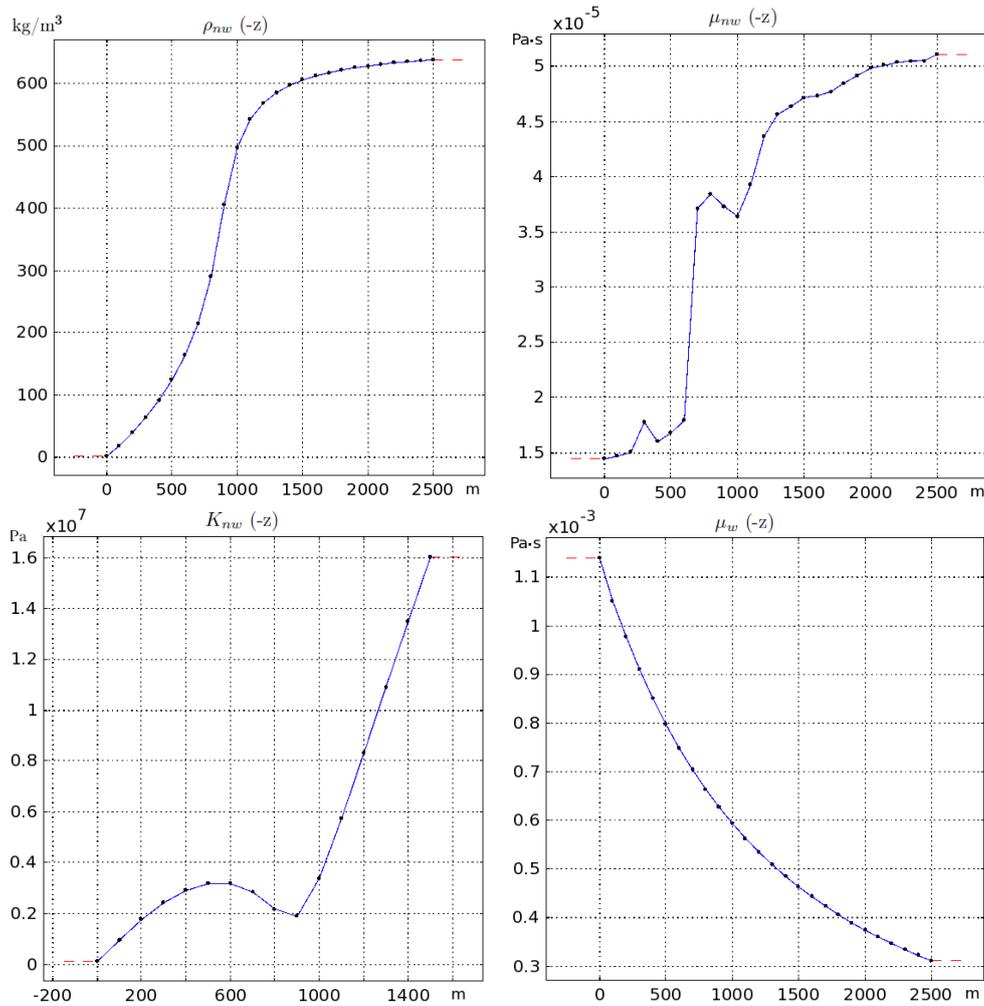


Figure 2.3: Interpolated functions of the model

2.4.2 Coupling of 2D-axisymmetric and full 3D formulations

As a 3D model is computationally intensive, we want to restrict the 3D model to some neighborhood of the injection well that would contain the CO₂ cloud for at least 10 years, typically of the order of 1 km. However, when restricting the model to this 3D zone, the proximity of the boundary $\partial\Omega_{out}$ (Figure 2.2) would strongly modify the solution. In the context of CO₂ storage, Le Gallo (2009) has shown for a given basin the need to extent the simulation domain up to at least 100 km to minimize the influence of the boundary conditions. Hence, we propose in this thesis to couple the 3D solution inside the 3D domain with an axisymmetric solution computed on an *ad-hoc* reduced domain. As this coupled 2D-3D model cannot easily be handled in a standard 3D-FE code, an equivalent 3D model is worked out.

This will be done in three steps: i) stating the equations on the 3D model so that the approximate solution satisfies the rotation-invariant conditions in the outer domain Ω_+ (complement of Ω_- in Figure 2.4); ii) finding the equations which apply on a slice of small thickness e (domain Ω_e) having the same solution as the rotation-invariant one in the outer domain Ω_+ ; iii) coupling together the 3D interior model and the outer slice model in the context of a 3D Finite Element software.

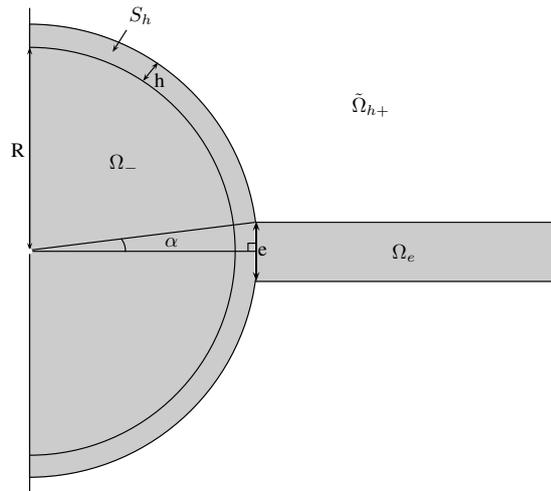


Figure 2.4: View from above of the considered 3D domain consisting of three parts: Ω_- , the domain where the original 3D formulation of the problem applies, S_h - zone with significant tangent conductivity, Ω_e - 3D slice of a little thickness e conceived to represent the equivalent axisymmetric solution of the problem in $\tilde{\Omega}_{h+}$

The development will be done for a generic diffusion equation of the following form:

$$d_a \frac{\partial u}{\partial t} + \text{div}(-c \nabla u + \gamma) = 0 \quad (2.26)$$

applying in a given domain Ω where the variable $u(x, y, z, t)$ can be vector valued, the coefficients d_a and γ - second-order-tensor valued and c - symmetric fourth-order tensor valued. Appropriate boundary conditions on $\partial\Omega$ are also assumed, either Dirichelet or

Neumann type on the boundary.

$$u = u_o, \quad (c\nabla u) \cdot n = q_o$$

where n is the outer normal vector.

Axisymmetric solution at finite distance Let us call $S \subset \Omega$ a cylindrical surface bounded by the two planes $z = z^\pm$ with a circular cross-section of radius R and let $\Omega_- \subset \Omega$ be the inner domain and $\tilde{\Omega}_+$ be the outer domain. In the outer domain both physical properties and boundary conditions are assumed to have a cylindrical symmetry^a. Note however that the solution u is rotation-invariant only when properties and boundary conditions share the same property.

Let us now consider an approximate 3D solution $\tilde{u}_-(x, y, z)$ satisfying the field Equation 2.26 in Ω_- together with the boundary conditions on $\partial\Omega \cap \partial\Omega_-$, and $\tilde{u}_+(r, z)$ an axisymmetrical solution of Equation 2.26 in $\tilde{\Omega}_+$ satisfying the boundary conditions on $\partial\Omega \cap \partial\tilde{\Omega}_+$ with the following compatibility conditions on S :

$$\frac{1}{2\pi} \int_0^{2\pi} \tilde{u}_- d\theta = \tilde{u}_+(z, R) \quad , \quad \forall z \in]z^-, z^+[\quad (2.27)$$

$$\int_0^{2\pi} (c_- \nabla \tilde{u}_-) \cdot e_r R d\theta = q_R(z) \quad , \quad \forall z \in]z^-, z^+[\quad (2.28)$$

$$\frac{1}{R} \partial_\theta \tilde{u}_- = 0 \quad \text{on } S \quad (2.29)$$

$$c_{rr} \partial_r \tilde{u}_+ = \frac{q_R(z)}{2\pi R} \quad , \quad \forall z \in]z^-, z^+[\quad (2.30)$$

where e_r is the outer normal vector of S . Due to orthotropy c_{rr} is the radial hydraulic conductivity and $q_R(z)$ is the vector of radial fluxes crossing S at a given depth z , the fluxes tangent to S vanish. Equation 2.29 stating that the solution is rotation-invariant at R is difficult to implement in a Finite Element framework. However, noticing that its left hand side is one of the components of the gradient, a penalization technique is used.

Penalization of rotation-invariance Let S_h be a tube of thickness h and internal radius R , $\tilde{\Omega}_{h+} = \tilde{\Omega}_+ \setminus S_h$ and let c_{eh} be the orthotropic hydraulic conductivity tensor defined on S_h as:

$$c_{eh} = c_{zz} e_z \otimes e_z + \epsilon^{-1} (c_{\theta\theta} e_\theta \otimes e_\theta + c_{rr} e_r \otimes e_r)$$

for any small ϵ . Equation 2.29 is replaced by the following field equations and boundary conditions on S_h

$$\begin{cases} \operatorname{div}(-c_{eh} \nabla u_h) = 0 & \text{in } S_h \\ u_h = \tilde{u}_- & \text{for } r = R \\ u_h = \tilde{u}_+ & \text{for } r = R + h \\ (c_- \nabla \tilde{u}_- - c_{eh} \nabla u_h) \cdot e_r = 0 & \text{for } r = R \\ (c_+ \nabla \tilde{u}_+ - c_{eh} \nabla u_h) \cdot e_r = 0 & \text{for } r = R + h \end{cases} \quad (2.31)$$

^aTensor c has to be orthotropic with respect to the z axis.

Taking the variational form for any virtual field $v = v_o(z) + v_1(\theta, z) \frac{x-R}{h}$ leads to:

$$\int_S \left(\frac{R+h}{R} c_+ \nabla \tilde{u}_+ - c_- \nabla \tilde{u}_- \right) \cdot e_r v_o dS + \int_S \left(\frac{R+h}{R} c_+ \nabla \tilde{u}_+ \right) \cdot e_r v_1 dS = \int_{S_h} \left(c_{zz} \partial_z u_h \cdot \partial_z v + \epsilon^{-1} c_{rr} \left(\frac{1}{h} \partial_r u_h v_1 + \frac{r-R}{hr^2} \partial_\theta u_h \partial_\theta v_1 \right) \right) dV \quad (2.32)$$

Taking $v_1 = 0$ and $h \rightarrow 0$ leads to the balance of fluxes assuming that $\partial_z u_h \cdot \partial_z v$ is bounded:

$$(c_+ \nabla \tilde{u}_+ - c_- \nabla \tilde{u}_-) \cdot e_r = 0 \quad \text{on } S \quad (2.33)$$

leading to Equation 2.28.

Hence taking now $v_o = 0$ yields:

$$\left(1 + \frac{h}{R}\right) \int_S (c_- \nabla \tilde{u}_-) \cdot e_r v_1 dS = \int_S \left(c_{zz} \partial_z \left(\int_R^{R+h} u_h(r-R) \frac{r}{R} dr \right) \partial_z v_1 dS + \int_{S_h} (\epsilon h)^{-1} c_{rr} \left(\partial_r u_h v_1 + \frac{r-R}{r^2} \partial_\theta u_h \partial_\theta v_1 \right) dV \right) \quad (2.34)$$

After some calculations the leading terms are:

$$\int_S (c_- \nabla \tilde{u}_-) \cdot e_r v_1 dS = \int_S \left(c_{zz} \partial_z u_h \partial_z v_1 dS + (\epsilon)^{-1} \int_S c_{rr} \left(\partial_r u_h v_1 + \frac{1}{R} \partial_\theta u_h \partial_\theta v_1 \right) dS \right) \quad (2.35)$$

Taking the limit for $\epsilon \rightarrow 0$ leads to:

$$\partial_r u_h = 0 \quad (2.36)$$

$$\frac{1}{R} \partial_\theta u_h = 0 \quad (2.37)$$

Hence $\tilde{u}_+ = u_h = \tilde{u}_-$ and $\frac{1}{R} \partial_\theta \tilde{u}_- = 0$ yielding Equations 2.27 and 2.29.

Equivalent equations on an outer slice Since for any $g(r, z)$ one knows that:

$$\operatorname{div} g = \frac{1}{r} (\partial_r + \partial_z) (rg) \stackrel{\text{def}}{=} \frac{1}{r} \operatorname{div}_{2D} (rg) \quad (2.38)$$

Hence Equation 2.26 for \tilde{u}_+ on Ω_+ can be written as:

$$r d_a \frac{\partial \tilde{u}_+}{\partial t} + \operatorname{div}_{2D} (-cr \nabla \tilde{u}_+ + r\gamma) = 0 \quad (2.39)$$

Therefore the field u_e is defined as the solution of

$$d_{ae} \frac{\partial \tilde{u}_e}{\partial t} + \operatorname{div}_{2D} (-c_e \nabla \tilde{u}_e + \gamma_e) = 0 \quad (2.40)$$

on the 3D thin vertical slice $\Omega_e = \Omega \cap S_e$ with $S_e = \{(x, y, z) : R \cos \alpha < x, -e/2 < y < e/2, z^- < z < z^+\}$, where $e = 2(R+h) \sin \alpha$ is the thickness of the slice. The new

coefficients are $d_{ae}(x, y, z) = \frac{2\pi x}{e} d_a(r = x, z)$, $c_e(x, y, z) = \frac{2\pi x}{e} c(r = x, z)$, $\gamma_e(x, y, z) = \frac{2\pi x}{e} \gamma(r = x, z)$. The following boundary conditions apply:

$$\begin{cases} \partial_y \tilde{u}_e = 0 & \text{for } y = \pm e/2 \\ \tilde{u}_e = \tilde{u}_+(R+h, z) & \text{for } x = (R+h) \cos \alpha \end{cases} \quad (2.41)$$

Since the slice and the boundary conditions are invariant with respect to y so is the field \tilde{u}_e . In addition, it obviously satisfy

$$\tilde{u}_e(x, y, z) = \tilde{u}_+(r = x, z)$$

Moreover the two solutions being equal we can also show that the related fluxes coincide:

$$q_R = \int_0^{2\pi} c \nabla \tilde{u}_+(R+h) d\theta = 2\pi(R+h) c \partial_r \tilde{u}_+ = e c_e(x = (R+h), y, z) \partial_x u_e = \int_{-e/2}^{e/2} c_e \partial_x u_e dy$$

Approximate 3D solution on a cylinder-slice domain Combining the two above methods, we can define the approximation solution \tilde{u}_{he} satisfying Equation 2.26 on a composite domain $\Omega_- \cup S_h \cup \Omega_e$ shown in Figure 2.4

$$\tilde{u}_{he} = \tilde{u}_- \quad \text{in } \Omega_- \quad (2.42)$$

$$\tilde{u}_{he} = \tilde{u}_h \quad \text{in } S_h \quad (2.43)$$

$$\tilde{u}_{he} = \tilde{u}_e \quad \text{in } \Omega_e \quad (2.44)$$

with the following vanishing flux boundary conditions:

$$\partial_r \tilde{u}_{he} = 0 \quad \text{for } r = R+h, \theta = [\alpha, 2\pi - \alpha] \quad (2.45)$$

$$\partial_y \tilde{u}_{he} = 0 \quad \text{for } y = \pm e/2, x > (R+h) \cos \alpha \quad (2.46)$$

$$\partial_z \tilde{u}_{he} = 0 \quad \text{for } z = z^\pm \quad (2.47)$$

The geometry and the mesh of the coupled model are presented in Figure 2.5. The simplified geometry significantly eases the task of the meshing and the computation of the solution. For more complex geometries the proposed methodology will still be working until the boundary between the 3D and rotation-invariant part is axisymmetric. For meshing of more complex geometries advanced techniques exist such as described by [Juntunen and Wheeler \(2012\)](#).

Several meshes have been compared for the coupled model having the characteristics given hereafter. Figure 2.6 illustrates the front form at a fixed time instant for the three meshes. Due to the particularities of the problem, the cylindrical mesh even with smaller number of elements presents better results than the tetraedric meshes.

Cylindric mesh	1.5·10 ⁴ elements	(1.1·10 ⁴ in 3D domain)
Tetraedric mesh "Finer"	15·10 ⁴ elements	(14.8·10 ⁴ in 3D domain)
Tetraedric mesh	1.9·10 ⁴ elements	(1.7·10 ⁴ in 3D domain)

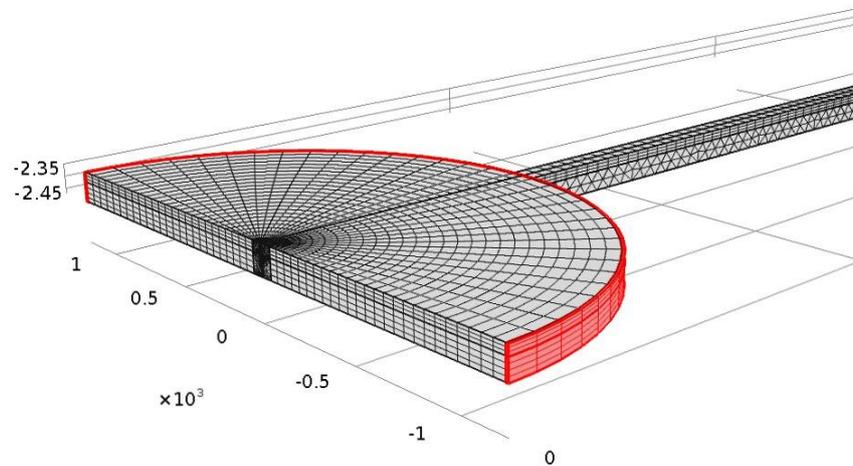


Figure 2.5: The mesh of the coupled model: 3D zone and rotation-invariant continuation, the transition zone (S_h) is highlighted in red.

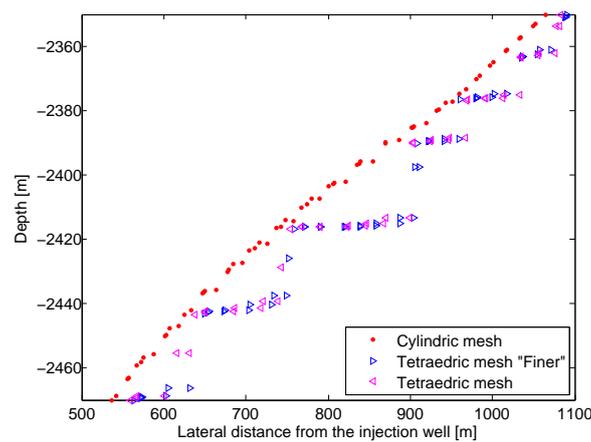


Figure 2.6: Vertical cut of the front for different meshes

2.5 Existing simplified analytical models

Analytical solution of Nordbotten for maximal lateral spread

A review of models for CO₂ storage has been done by [Stenhouse et al. \(2006\)](#).

An analytical model proposed by ([Nordbotten and Celia, 2006](#)) allows a simple and efficient prediction of the furthest extent of the injected supercritical CO₂ into a brine-saturated deep aquifer. First the derivation of this model was based on the energy minimization principle and was valid only for systems with high injection rates compared to aquifer permeability, height and fluid density difference ([Nordbotten et al., 2005](#)). But in further works ([Nordbotten and Celia, 2006](#); [Celia and Nordbotten, 2009](#)) the authors

have overcome these limitations switching to the use of similarity solutions for fluids with different mobilities.

Hypothesis and equations of the model The equations leading to the analytical solution are:

- Pressure variation in the vertical direction (due to vertical equilibrium assumption)
- Vertically integrated Darcy law
- Volume balance for each phase integrated vertically

The major assumptions of the model are essentially horizontal flow, radial symmetry and sharp front

Analytical solution The analytical formula expresses the maximal lateral extent of the injected gas cloud r_{max} as a function of injection time t and volumetric debit Q_v , endpoint of the gas relative permeability curve kr_{max} , reservoir thickness H , porosity ϕ , maximal gas saturation s_{nw}^{max} and fluid viscosities μ_w and μ_{nw} :

$$r_{max} = \sqrt{\frac{Q \cdot t}{\pi} \frac{kr_{max} \cdot \mu_w}{H \cdot \phi \cdot s_{nw}^{max} \cdot \mu_{nw}}} \quad (2.48)$$

Validity limits Among the assumptions made in order to obtain the analytical solution there is a condition of vertical equilibrium. Such a condition requires to verify that the gravity effect is negligible. In the work of Nordbotten and Celia (2006) a dimensionless gravity factor Γ is introduced to express the ratio between the buoyancy and viscous and pressure forces effects.

$$\Gamma = \frac{2\pi(\rho_w - \rho_{nw})gkH^2}{Q\mu_w} \quad (2.49)$$

In the same work a cutoff of $\Gamma = 0.5$ is proposed to identify when buoyancy cannot be neglected. The set of parameters chosen for numerical model (see section 2.4) gives the value of the gravity factor which exceeds the threshold of 0.5. Therefore, to make the use of this analytical model possible the reservoir thickness is reduced to 50 m. In this case the value of the gravity factor is 0.32. Besides, the model is only valid for injection along all the reservoir height. That is why the results of the analytical model are compared to an adapted numerical model with the injection interval extended to the entire reservoir thickness of 50 m.

Comparison of numerical and analytical results The maximal lateral extent of the CO₂ cloud after 10 years of injection estimated by analytical solution is 1168 m. The results obtained through the numerical solution is of 1430 m. This difference is not unexpected and can be explained by the phenomena neglected in the analytical model. The numerical model takes into account the gravity effect: the gas rises faster increasing the radius of the cloud on the top of the reservoir. In the analytical model the gas saturation after the front

is constant and equal to the maximal possible saturation whereas in numerical model the CO_2 saturation is variable due to the relative permeability curves. With the same injected quantity the area swept by the gas in the numerical model would be greater than those in the analytical model.

Chapter 3

Risk assessment for CO₂ storage and uncertainty propagation

Contents

3.1	Definitions	28
3.1.1	Components of risk management	28
3.1.2	Dimensions of risk	29
3.1.3	Risk scenarios	30
3.2	Review of risk assessment studies applied to CO₂ storage . .	32
3.2.1	Types of risk analysis	32
3.2.2	Particularities of CO ₂ Storage risk assessment	34
3.2.3	Methodologies	35
3.3	Methodology used by INERIS for CCS risk assessment . . .	37
3.3.1	The global workflow, the risk model in the form of an event tree	37
3.3.2	Central events and impacting phenomena for CO ₂ storage . . .	38
3.4	From scenario to uncertainty propagation	41
3.4.1	The relation between scenario evaluation and uncertainty prop- agation	42
3.4.2	Deterministic, probabilistic and level-2 probabilistic quantities of interest and critical thresholds	43
3.5	Classical uncertainty propagation methods illustrated on a simplified physical model	48
3.5.1	Probabilistic analysis: global approach	49
3.5.2	‘Level-2’ probabilistic analysis	57
3.5.3	Comparison of the results	59
3.6	Concluding remarks	61

As it was mentioned in the introduction, risk quantification is closely related to uncertainty treatment. Each methodology is intrinsically based on certain perception of uncertainty (including disregard of its existence). To ‘integrate’ uncertainty it could be necessary to choose a different methodology and/or make different choices for the complexity of the risk measure, uncertainty sources description and their propagation.

In this chapter, first, main notions of risk, risk assessment, scenario parts are introduced. Then the principal types of risk assessment methodologies are discussed. The three levels of the complexity of the risk measure are associated with the corresponding representation of uncertainty sources and the types of risk assessment methodologies allowing to estimate this risk measure. Taking the case of a risk measure in the form of threshold exceedance, it is demonstrated how critical thresholds can be determined from vulnerability back to causes. Depending on the type of the risk measure three levels of thresholds are considered.

The methodologies applied so far to the CO₂ storage are reviewed in the light of uncertainty treatment. The methodology employed in the project to which the present work contributes is presented including the relevant scenarios. The central event of leakage by lateral migration is studied through an analytical model of CO₂ injection into a deep aquifer, allowing to compare the performance of main uncertainty propagation techniques.

3.1 Definitions

The perception of risk is quite personal as it depends on the values that humans have. We are speaking of risk only when something of value is in danger. The definition of risk as ‘effect of uncertainty on the objectives’ by ISO Standardization ([ISO 31000:2009](#)) seems to reflect the subtlety of perception. It is possible and important to converge to the unique and rigorous terminology for risk-related domain. A ‘dictionary’ specific to CO₂ geological storage was proposed by [Korre and Durucan \(2009\)](#). Unfortunately, this terminology is not in accordance with the ISO Guide as it is based on another standard ([WHO, 2004](#)) dealing with health impacts. The two terminologies were published the same year. In the present work the choice of ISO terminology was made as it is meant to be universal for risk management in any domain. Hereafter, some notions related to risk management are recalled.

3.1.1 Components of risk management

According to the standard [ISO 31000:2009](#) Risk management is a continuous and iterative cycle which contains the following steps:

Risk management	{	<ul style="list-style-type: none"> Establishment of the context Risk assessment Risk treatment Communication Monitoring
-----------------	---	--

It has to be noticed that the notions of risk management, risk assessment and risk

analysis are different. From the definition above, risk management is the most general term and it includes the risk assessment step. Risk assessment on its turn includes the risk analysis phase:

$$\text{Risk assessment} \left\{ \begin{array}{l} \text{Risk identification} \\ \text{Risk analysis (quantifying the likelihood and the consequences)} \\ \text{Risk evaluation} \\ \text{Risk hierarchisation} \end{array} \right.$$

The main difference between risk assessment and risk analysis is that the first includes a decision of acceptability of the risk made during the risk evaluation step. So, the level of risks estimated during the risk analysis step is generally compared with a critical threshold.

The present work focuses on the risk assessment (mostly the risk analysis step, with modelling tasks and integration of uncertainties in models). The goal of this step of risk management according to the ISO standard consists in information supply in order to allow an informed decision-making regarding the level of risk. It follows that the quantitative measures of risk should be defined.

3.1.2 Dimensions of risk

The measures of risk are rather non-uniform depending on the context. But the quantitative dimensions of risk can be deduced from the goals of risk analysis. It is generally accepted that risk analysis for a given system is supposed to answer the following three questions, formulated by [Kaplan and Garrick \(1981\)](#) :

1. What can go wrong?
2. How likely is it that this will happen?
3. If it does what are the consequences?

There is the fourth question which can be added addressing the subjectivity of any performed risk analysis:

4. What is the confidence on the obtained risk estimations?

Answering the first question includes purely qualitative analysis of the system, including the identification and ranking of possible failure events and the establishment of the relations between them. The common name for this stage of risk analysis in the chemical industry and in the nuclear industry is ‘hazard analysis’. It gives access to possible risk scenarios.

The second question is aimed to give an appreciation (estimate) of the probability of the occurrence of the identified failure events.

The response to the third question requires an estimation of the severity (or gravity) of the consequences. Both second and third questions can be answered either in qualitative or quantitative ways. Different approaches to answering these first three questions give birth to the variety of risk assessment methodologies ([Kirchsteiger, 1999](#)).

Therefore, principal quantitative dimensions of risk of an impacting phenomenon are its probability and its severity (ISO 31000:2009, 2009).

$$Risk = Probability \times Severity \quad (3.1)$$

The severity can be expressed as the intensity of the physico-chemical phenomenon coupled with the vulnerability of the asset at stake (sensible element) towards this phenomenon. Such distinction allows to separate the potential vulnerable element from the hazard which is quite common for the domain of natural risk. Engaging this term an equivalent definition of risk is:

$$Risk = Hazard \times Exposure\ of\ vulnerable\ elements \quad (3.2)$$

According to Field et al. (2012), the exposure is defined as the presence of the assets at stake, namely: people, livelihoods, environmental services and resources, infrastructure, or economic, social, or cultural assets in places that could be adversely affected. The intrinsic propensity or predisposition to be adversely affected is defined as the vulnerability.

Evidently this definition should not be understood as a formula. Hazard and vulnerability can be related especially in the case of CO₂ storage risks where the most hazards can be described like slow coupled continuous processes more than punctual events. The definition 3.1.2 will be used hereafter in a subdivision of a scenario: usually the major part of a scenario involves the hazard characterization whereas only its last element concerns the asset at stake and its vulnerability. In the present work the hazard characterization part is concerned (Section 3.5.3 and Chapter 4), focusing on the probability of events and the intensity of physico-chemical mechanisms.

3.1.3 Risk scenarios

As mentioned in the previous section, risk analysis deals with a number of risk scenarios. A scenario is a single pathway from a given cause to a precise consequence. The goal of using scenarios is, first, to be able to evaluate the probability of the impacting phenomena given the probabilities of causes and, secondly, to try to intervene in the scenario starting from the causes in order to diminish the probability or the severity of the final impacting phenomena. The most exhaustive ensemble of risk scenarios composes a ‘risk model’ of the system. Regardless of the domain of application or regulatory constraints, a scenario always has the same components which are summarized in Figure 3.1.

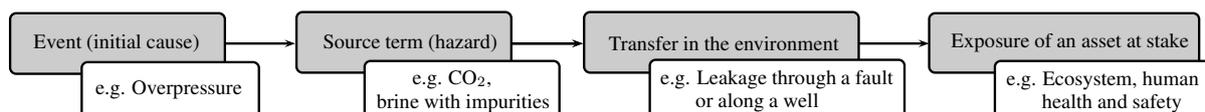


Figure 3.1: Generic representation of a risk scenario with an example for CO₂ storage, source INERIS

A more global subdivision of a scenario is: Cause - Central Event - Impacting Phenomenon - Asset at stake. The notion of central event can be defined as an event that

deviates the system behavior from normal and certainly causes an effect on the assets at stake. The central events, contrary to the preceding events, are likely to be detected. They initiate the transfer in the environment. The typical example is the loss of containment, i.e. leakage as shown in Figure 3.1 (see also Section 3.3.2 for the list of central events for the CO₂ storage). An impacting phenomenon is a process launched by a central event and having a direct impact on an asset at stake (ecosystem, human health, performance of storage, other human activities).

The subdivision of a risk scenario with central event and impacting phenomena is convenient to cluster scenarios according to their trigger mechanism or their final impact. In the case of CO₂ storage, all the central events (and consequently the impacting phenomena) happen during or after the ‘Transfer in the environment’ term. This is the reason why the definition ‘Event-Source-Transfer-Exposure’ is convenient for the description of hazards, focusing on the left-hand side of the scenario.

Generally the scenarios are not independent and have some of the elements above in common. So a schematic representation of a ‘risk model’ would look like Figure 3.2. The phases of risk treatment (measures to diminish the risk) and monitoring can be reflected on the risk model as well.

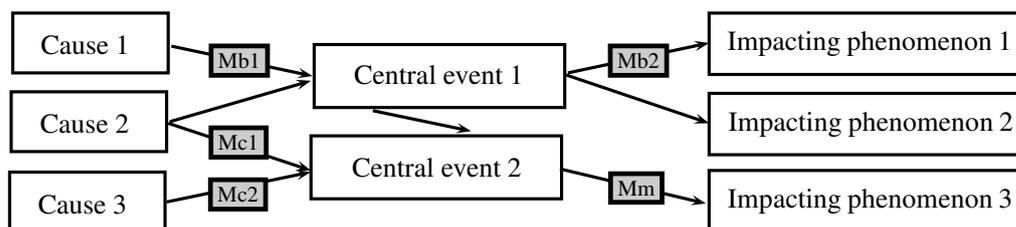


Figure 3.2: Schematic representation of a risk model including risk treatment (safety barriers Mb and conception barriers Mc) and monitoring (Mm) (adapted from Gombert and Farret, 2009)

In order to identify and characterize the risk scenarios there should be a guidance for expertise : a ‘**methodology**’. Some of the methodologies are standardized. Each methodology is often associated with its ‘**tools**’. A tool serves a methodology and helps in calculating, structuring and presenting the results in an operational and user-friendly way. A methodology constitutes the core exercise of risk assessment but it makes part of a more global ‘**methodological framework**’. Often represented as a workflow, it sets the actions and rules to be followed in order to reach the desired result. The methodological framework is often imposed by the regulation (then, those who follow it should prove that the regulatory conditions are fulfilled and the acceptability criteria are met). Also it can be imposed by the state of the art in a given technical domain. Suggested workflow for a global methodological framework for CO₂ storage will be discussed further (Table 3.4).

3.2 Review of risk assessment studies applied to CO₂ storage

This part is devoted to methodologies already applied to CO₂ storage risk assessment and accessible from literature review. In order to better understand the diversity of the methodologies, the main classification modes adopted from other application fields are first discussed. Peculiarities of CO₂ storage differentiating it from other sectors are stated.

3.2.1 Types of risk analysis

A wide spectrum of qualitative and quantitative methods as well as various assumptions and data can be used in risk analysis. The level of detail and quality of a risk analysis depends on the detail and quality of the available information, the validity of assumptions and first of all the objectives of the analysis.

[Tixier et al. \(2002\)](#) proposed a review and classification of risk analysis methodologies for industrial plants. The authors identified 62 methodologies and ranked them according to the following criteria:

- the phases of risk assessment included (identification phase; identification and evaluation phases; identification, evaluation and hierarchisation phases),
- types of method (qualitative, quantitative, deterministic, probabilistic),
- types of input data (text, plan, probabilities...),
- types of output data (recommendations, lists, scenarios probabilities, hierarchy).

These criteria can be used for classification of risk assessment methodologies in any domain of application. Dwelling on the types of method, there exist three main classes : qualitative, semi-quantitative and quantitative. The quantitative class can be subdivided to deterministic, probabilistic, mixed deterministic and probabilistic, and fuzzy logic (which alternatively can be attributed to semi-quantitative category).

There could be another important criterion of classification: 'frequentist' or 'degree-of-belief' logic. Nevertheless, a study can combine both approaches.

Qualitative approach

Qualitative risk analysis use narrative scales to describe probabilities and consequences of particular hazards. It is of a particular interest for preliminary screening or assessment of risks. It can also be the only applicable approach when very little information is available. An example of qualitative scale for likelihood and consequences in the context of geological storage of CO₂ is proposed by [DNV \(2010, page 67\)](#)(see [Table 3.1](#)).

Semi-quantitative approach

Semi-quantitative methods can be divided in two types. The first one is based on defining of risk indexes leading to numerical scales for comparing risks. Although these risk indexes

Description	Very low	Low	Medium	High
Impact on storage integrity	None	Unexpected migration of CO ₂ inside the storage volume	Unexpected migration of CO ₂ outside the storage volume	Leakage to vadose zone
Impact on local environment	Minor or no damage	Local damage of short duration, time for restitution < 1 year	Time for restitution of ecological resource < 2 years	Time for restitution of ecological resource > 2 years

Table 3.1: Examples of definitions that could be used to qualitatively rank consequences of risk scenarios (DNV, 2010, page 67)

are numerical, their values do not reflect exactly the actual magnitude of probabilities and risks. If the risk index is carefully defined, justified and applied, such an approach is quite effective for ranking and prioritizing risks.

The second one is based on defining categories with corresponding quantitative intervals and distributing the scenarios among these categories according to an expert judgment. The limitation of semi-quantitative categorization is that its reliability depends on the justification and support of the expert judgment.

In order to make a qualitative or semi-quantitative study more objective and reliable, it is important to incorporate the feedback on passed incidents/accidents. An example of a semi-quantitative scale is presented in Table 3.2.

Quantitative probability	Level of likelihood	Description	Objective elements
$P \approx 1$	Certain	Certain or almost certain	Normal evolution, according to present knowledge. Event likely to occur once or even several times, either in the longer term, or on surface facilities during operation phase.
$P > 0.1$	C	Event likely to occur once (on surface facilities during the operation phase, or later in the underground system)	Event that may occur in the longer term and that was already observed on similar underground activities OR event/process that seems inescapable in the long-term. For surface facilities, Frequency of occurrence estimated about 100 years
$P > 0.01$	B	Unlikely (on surface facilities during the operation phase, or later in the underground system)	Not very likely on surface facilities during the injection phase, not very likely on underground system Event already observed on similar industrial facilities. For surface facilities, Frequency of occurrence estimated about 1000 years
$P > 0.001$	A	Very unlikely but plausible	Never observed (or observed in a very different configuration), but not impossible given of current knowledge (or lack of knowledge), either on surface equipment or in the longer term
$P \approx 0$	0	Not plausible	Never observed worldwide and not considered plausible on this site given our present knowledge

Table 3.2: A semi-qualitative quotation scale for likelihood that integrates all life stages of CTSC and any equipment (source INERIS).

Quantitative approach

Quantitative risk analysis uses estimates of numerical values of probability and severity. For deterministic methods there is one value for probability and one for severity. For probabilistic methods both probability and severity represent curves, because the uncertainty

introduces a range of possible values for the intensity. Therefore, probability and severity can be expressed as functions of the variable of interest describing the intensity.

Probability can be estimated in several ways: historical statistics/frequencies, professional judgment, simulation assuming a random distribution of failures/events, experimentally established failure data. The probability of a specific element of an event tree is often made up from probabilities of contributing elements. For systems with known conditional probabilities between elements and oriented tree structure the direct and inverse computation of probabilities is rather friendly. Unfortunately, for geological systems, even if the event tree is linearized (not presenting any cyclic structures), obtaining conditional probabilities turns into an impossible quest due to rarity/slowness of the events and limitations of learning from experience.

Severity is usually defined in precise terms, such as number of injuries (or deaths), magnitude of adverse effects on the receiving ecosystem, extent of damage on structures. Therefore, supplementary models may need to be engaged to assess the specific consequences. For example, an air dispersion model in the underground structures (caves, basements, tunnels...) in order to assess the critical endpoint concentrations in the event of release. Or a model of gas accumulation in the overlying layers close to the surface level (in the surface waters most of all).

Risk as follows would be measured by extent of damage within a given area per time period combining severity and probability of risk scenarios. Presence of cumulative risks should be taken into account. The way of risk expression is often defined by norms.

Risk curve is a continuous representation of risk as a function of severity versus probability. It can be found from the curves of severity of intensity and probability of intensity.

3.2.2 Particularities of CO₂ Storage risk assessment

The risk assessment for the whole CO₂ Capture-Transport-Injection-Storage chain is quite complex as it involves at the same time technological risk (mostly for the Capture-Transport-Injection part), natural risk (which can occur during the life-time of the storage or which can be induced by CO₂ injection) and finally environmental and sanitary impacts (concerning most of all the storage stage).

The knowledge and experience in dealing with technological risk exist in the domain of industrial security. Management of the risks related to the drilling and injection can be learned from the petroleum industry. The carbon dioxide is an aggressive corrosive agent, so its injection differs from water, oil or natural gas, but its particularities are still being studied for the application of enhanced oil recovery. The environmental and sanitary impact estimations are less common than technological risks. However a know-how exists in the sectors of industrial releases and nuclear waste sequestration.

Nevertheless, the particularities of the CO₂ injection and storage risk assessment which make it difficult to apply the existing general methodologies from other sectors are many:

- Few full scale tests: the first project ('Sleipner' in Norway) has been operating since 1996. Currently only a few projects are operating (Wildenborg et al., 2013). As a consequence, estimates of likelihood from past events are very coarse.
- Long time scale: the behavior of the injected gas should be predicted for several thousands of years. But the interaction of the slow coupled processes is not known in the long-term (see section 2.1) and a quantitative prediction by numerical models becomes rather heavy and time-consuming.
- Strong dependency of the risk on the natural properties of the geological media and their evolution over time; poorly known interactions between parameters. The statistical data needed for the treatment of uncertainty (both lack of knowledge and natural spatial variability) is difficult to obtain. Even the scenario of normal evolution is strongly subjected to uncertainties.
- System design is almost uncontrolled: there are very few operating variables which do not cover the majority of processes and events. The list of operating variables can possibly be reduced to: well location and characteristics, injection debit/pressure, injection time, injection temperature, fractions if mixed with other substances. After the injection phase the spread and the interaction of the gas with the medium are out of control: one can only observe the evolution through monitoring. So there will be few conception and security barriers at the stage of storage, whereas the monitoring data should be progressively taken into account to adjust predictions.
- The methodology of risk assessment should be systemic (not to drop any hazards from consideration, even those of low probability) and at the same time site-specific (in order to take into account the quality and quantity of data available on the site).
- A numerical model of chemico-physical processes (even simplified), which would describe the long-term evolution of the injected fluid in all the compartments from the reservoir up to the surface, is very expensive in terms of computational resources. The computational power available today is not sufficient to account for uncertainties in a probabilistic way with the use of such models. Therefore, in order to estimate the consequences, the entire system would be divided in compartments, though obviously the compartments are dependent. The interaction between successive models is oversimplified (this contributes to model uncertainty and the subjectivity of analysis). Yet, such models are useful for the first estimation of sites and their risks.
- Some events are sub-states of the system, therefore, knowing the complexity of the processes, the event trees are not straightforward (they can even contain cycles).

3.2.3 Methodologies

The particularities of the CO₂ storage systems highlighted above (Section 3.2.2) together with the diversity of risk analysis types and purposes described in Section 3.2.1 lead to a variety of choices in how to perform a risk assessment. As the present work aims at

studying the possibility of including uncertainties in risk assessment, it is important to review what are the actual methodologies already applied to CO₂ storage.

Qualitative		FEP (tool) (Savage et al., 2004) VEF (U.S.EPA, 2008) ‘What-If’ and SWIFT (Vendrig et al., 2003 ; DNV, 2010) Delphi (Wassermann et al., 2011)
	Semi-quantitative	RISQUE QRA (Bowden and Rigg, 2004) SRF (HSE) (Oldenburg, 2005, 2007) MCA (Gough and Shackley, 2006 ; Jakobsen et al., 2013) similar to MAUT ARAMIS,LOPA (Wilday and Farret, 2010) Risk matrix tool (Gombert and Farret, 2009) OSQAR (tool) used in projects MANAUS (2011) and iNTeg-Risk
Quantitative	Probabilistic	Event Tree Analysis (tool) CO2PENS (SMA) (Stauffer et al., 2009 ; Viswanathan et al., 2008) CQUESTRA (Le Neveu, 2008) QRTT (BP product)(Dodds et al., 2011) FEP + Scenario approach (Wildenborg et al., 2004, 2005 ; Stenhouse et al., 2005) FEP + Markov Chains (Nepveu et al., 2009) MANAUS (MANAUS, 2011) CCDF (tool) (present work) ESL (Paulley et al., 2012 ; Metcalf et al., 2013) P&R TM (Le Guen et al., 2008, 2011 ; Meyer et al., 2009) PRA (Rish, 2005)
	Det&Prob	MOSAR (Cherkaoui and Lopez, 2009) Delphi+RISQUE (Wyatt et al., 2009) CFA (Oldenburg, 2009)

Table 3.3: Risk assessment methodologies and tools applied to Carbon Capture and Storage (CCS). For brief descriptions refer to Appendix A. ‘+’ signifies that a systematic approach (FEP or Delphi) is followed for the quantification step by another method

The literature on the methodologies of risk assessment applied to Carbon Capture and Storage (CCS) has been reviewed in [CSLF \(2009\)](#). Since, several new methodologies have been applied. Another review with brief characteristics of the methodologies can be found in [Lahaie et al. \(2009\)](#). [Condor et al. \(2011\)](#) differentiate the methodologies in terms of the objectives of application for geological storage (which is strongly related to the phases

of risk assessment included in a study in the Tixier's classification). The authors provide examples of types of input and output data for several methodologies. Concerning 'types of methodology' some authors refute the distinction between qualitative and quantitative methods and sort the methodologies in main types: scenario analysis (scenarios of how a CO₂ storage system might evolve are analyzed in terms of CO₂ migration/leakage), fault/event tree analysis (CO₂ release is evaluated as a combination of possible steps, analysis is based on the sub-states of the system as for Markov chains), expert judgment and screening-level analysis.

[Gerstenberger et al. \(2009\)](#) also recognize that the distinction between qualitative and quantitative methods is quite subjective and suggests to focus on the precision of the analysis (level of detail used to describe the system). The less precise methods would be those using descriptive scale combined with fewer system components, the most complicated methods would require the assignment of probability distributions to describe the behavior of multiple components within a system.

We propose an update of the review of risk assessment and risk analysis methodologies and tools applied so far to CCS. The literature references are presented in the form of a table (see [Table 3.3](#)). For brief descriptions of the methodologies refer to [Appendix A](#). The methodologies are classified in three groups: qualitative, semi-quantitative and quantitative. Quantitative branch contains probabilistic and mixed (deterministic and probabilistic) methodologies, the later having at least one of the dimensions of risk evaluated in deterministic way. As mentioned above the differentiation is not rigorous.

3.3 Methodology used by INERIS for CCS risk assessment

The present work contributes to the project of knowledge development for the geological storage of CO₂ and related risks at INERIS. In order to better position the contribution of the present work in the project, the author proposes a quick glance at the global workflow and the main identified central events and impacting phenomena. The procedure for CO₂ risk analysis followed in this project was proposed by [Farret et al. \(2010\)](#). It shares its principles with the methodology [MANAUS](#). It consists in identifying all the possible events which can occur due to the underground CO₂ storage, identifying the possible causes of such events and trying to relate these elements to form event-tree charts. The parts of the charts are quantified through numerical modelling. The expertise of INERIS in risk management for underground systems suggests that the phase of identification should be performed via a systematic approach which gathers all the scientific knowledge about phenomena and events and which ensures that all the possible sub-systems and events are considered, even if the probability of some events is low ([Lahaie et al., 2009](#)).

3.3.1 The global workflow, the risk model in the form of an event tree

Given the objectives of full risk assessment that covers all the chain of CO₂ Capture-Transport-Injection-Storage and its specificities, INERIS exploits the workflow shown in

Table 3.4 for a global methodological framework. The table includes the tools that can be engaged in the different steps.

1. Definition of the system, its boundaries, its life stages, its potential hazards	Conceptual model
2. Identification of the assets that can be exposed. Choice of adequate criteria for severity and likelihood	
3. Collect of knowledge from past events and accidents (learning from experience)	
4. Study of the scenario of normal evolution and its uncertainties	Numerical model
5. Preliminary risk analysis construction of scenarios of altered evolution	Risk Model (event trees)
6. Quantification of scenarios (detailed risk analysis)	Numerical model
7. Definition of measures for risk treatment (mitigation measures, safety barriers)	
8. Overview of risk scenarios to allow a risk evaluation	Risk matrix

Table 3.4: A global workflow for CO₂ storage risk assessment, source INERIS

The construction of a risk model is based on expert panels that follow a systematic approach such as ‘What-if’ or Preliminary Risk Analysis (PRA). Therefore, for the fifth step of the global workflow (Table 3.4) the results of a collective expertise are gathered and represented in event-tree charts. The analysis starts either by main events (see the section below), either by the failure modes of concrete components of the capture-transport-injection-storage chain. The approach for the failure modes analysis is similar to HAZOP (Hazard and Operability study) or FMEA (Failure Mode and Effects Analysis, standard CEI:60812). These methods operate with keywords to identify the failure modes. Then downstream analysis (identifying consequences) and upstream analysis (identifying the initial causes) are performed giving access to scenario consideration. The obtained event-tree charts are reviewed and completed to have the whole collection of risk scenarios defining the ‘risk model’ of the system.

3.3.2 Central events and impacting phenomena for CO₂ storage

The central events (detectable events that deviate the system behavior from normal and certainly cause an effect on the assets at stake) identified for the CO₂ storage are:

1. loss of containment (breach, leakage) on a surface equipment,

2. leakage along an injection or another operating well,
3. leakage through a closed or abandoned well,
4. leakage along a fault,
5. leakage through the caprock,
6. leakage by lateral migration,
7. flow perturbation (including transmission of fluid overpressure),
8. mechanical perturbation and its transmission.

MANAUS is not only a methodology but also a project which proposes a set of basic scenarios on which the present work relies. The impacting phenomena (processes launched by a central event and having a direct impact on an asset at stake) identified for the CCS chain at short and long-term are presented in Table 3.5.

The adduced impacting phenomena cover surface equipment for capture and transport as well as the underground system for injection and storage. It should be mentioned that not only the CO₂ itself but also the annex substances mobilized by CO₂ should be considered (such as in impacting phenomena 5 and 7 of the Table 3.5). These annex substances can be of different origins: impurities injected together with CO₂, impurities mobilized by the reaction between the matrix material and the brine acidified by CO₂.

Finally, focusing on the leakage by lateral migration and leakage through a fault considered in this work, the extract of the event tree constructed in **MANAUS** project is shown in Figure 3.3. One can notice the considerable complexity of the relations between the elements. Unless the Monte Carlo chains are used to estimate the probability of each element, the event tree needs to be linearized.

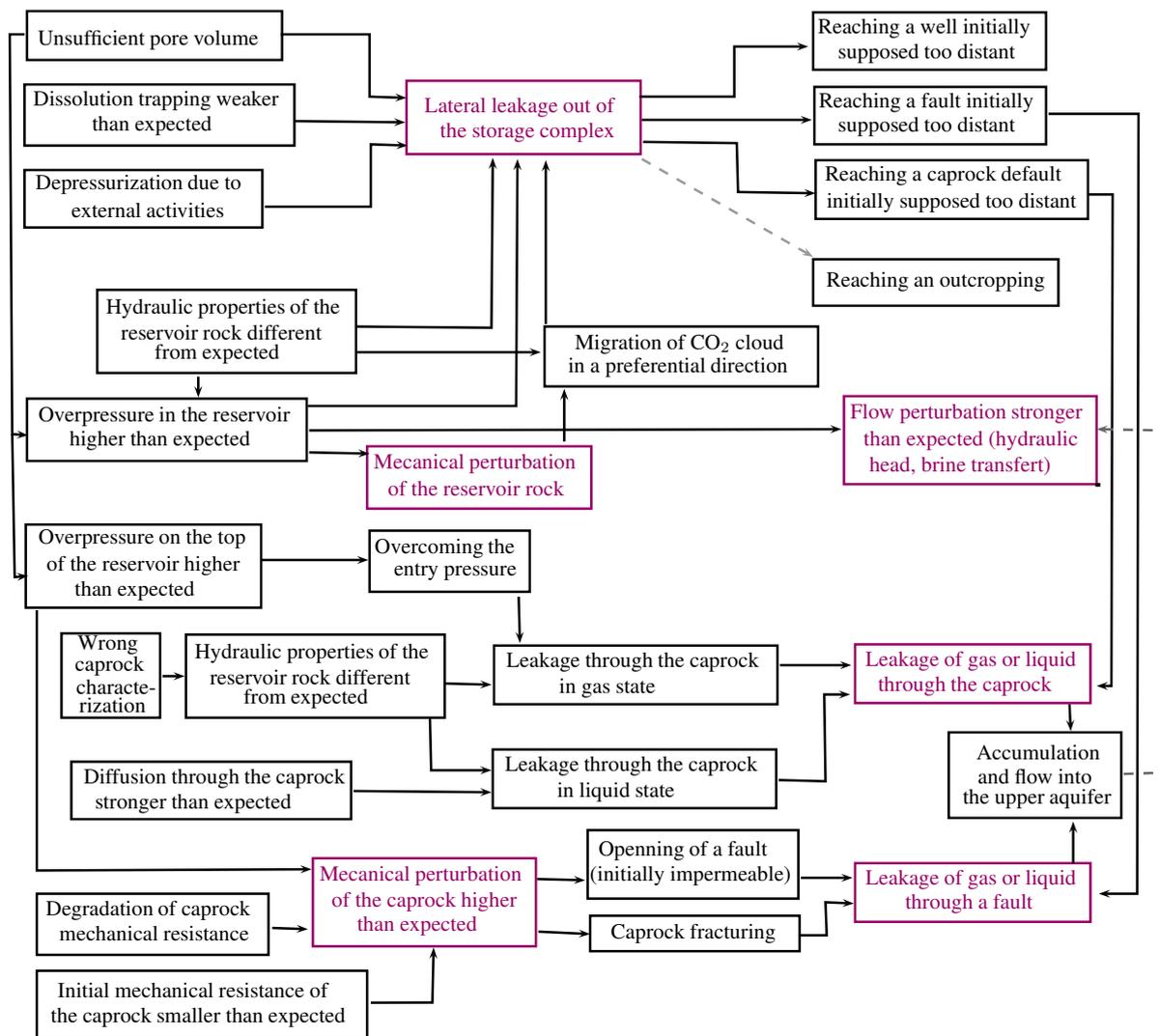


Figure 3.3: Example of an event tree (causes) for CO₂ migration through the caprock and the geological system. Extracted from MANAUS project without geochemical impacts. Violet boxes correspond to the central events

Table 3.5: The eight central events for risk analysis, their relation with eight Impacting phenomena and their potential effects (adapted from [Lahaie et al., 2009](#); [Farret and Thoraval, 2012, 2013](#)). The highlighted elements are those to which the present work contributes.

Central event	Impacting phenomenon	Sensitive compartments	Assets at stake	Potential effects
Loss of containment (breach, leakage) on a surface equipment	Sudden leakage in the air compartment Explosion or Fire	Atmosphere	Humans	Pressure, thermic impact
Leakage along an injection or another operating well	Sudden leakage in the air compartment	Atmosphere	Humans, performance	Toxic
Leakage through a closed or abandoned well	Slow emanation of CO ₂ to the surface	Atmosphere, poorly vented zones	Ecosystem, humans, performance	Ecotoxic, toxic
Leakage along a fault	Slow emanation of annexed substances	Atmosphere, poorly vented zones	Ecosystem, humans	Ecotoxic, toxic
Leakage through the caprock	Pollution by CO ₂ (pure or dissolved in another fluid)	Geological layers, surface water	Ecosystem, economic resources	Ecotoxic perturbation
Leakage by lateral migration	Pollution by annex substances	Geological layers, surface water	Ecosystem, economic resources	Ecotoxic perturbation
Flow perturbation (including transmission of fluid overpressure)	Disturbance of the regional hydraulic regime	Surface and sub-surface waters	Ecosystem, economic resources	hydraulic perturbation
Mechanical perturbation and its transmission	Progressive mechanical effect (slow surface elevation)	Surface	Property, humans (indirectly)	Mechanical
	Sudden mechanical effect provoked by the storage	Surface	Property, humans (indirectly)	Mechanical

3.4 From scenario to uncertainty propagation

The present section is devoted to the following questions:

What is the relation between scenario evaluation and uncertainty propagation? How the risk curve (combining probability and gravity of the impacting phenomenon) can be computed through modelling and uncertainty propagation? What are the levels of complexity of risk measures and critical thresholds? How the uncertainty can be 'taken

into account' in the qualitative and semi-quantitative methods?

3.4.1 The relation between scenario evaluation and uncertainty propagation

Scenario evaluation consists in quantification of probability and severity of events. Scenario evaluation does not necessarily include uncertainty, there are deterministic methods to estimate the probability (in frequentist sense) and severity. Such results can be associated with an approximate range of imprecision. But if the uncertainty sources are described in probabilistic or possibilistic way, then uncertainty propagation can be used as a tool for scenario evaluation. However, it requires some agreement: the processes are supposed to be known but subjected to different kinds of uncertainties.

In an event/fault tree the elements represent keywords corresponding to Boolean operators (of something happening or not happening) such as in the event tree in Figure 3.3. Except the events external to the system (e.g. meteorite impact, drilling of a new well,...), which can be described as Booleans, most of the events being internal to the system are actually related to continuous processes of system evolution in time. For such elements of an event/fault tree the Boolean approach presumes fixing a criterion on process intensity (first level of thresholds, see the next subsection).

The first event tree is based on collective exercise of brainstorming between experts, taking into account their knowledge and the feedback from existing sites, organized through a systematic method and formalized in a rigorous tool (either template or event tree). Modelling and uncertainty propagation can on their turn refine the event tree by pointing out the most essential and less important events.

Risk assessment methodologies based on scenarios evaluation should include a procedure of sub-division of the entire risk model into groups of scenarios whose severity (and possibly probability) can be evaluated within the same predictive model. For uncertainty propagation it is advisable for the model to be global and all the different inputs to be considered simultaneously through their whole potential scope of variation.

For a model which includes several elements of a scenario at a time, the intermediate elements of the scenario become intermediate model outputs themselves (or some of them could be included in the model as conditions creating cases). Even if the elements of a scenario are converted into mathematical expressions, the logical relation of the elements is not preserved.

Thus the transition from verbalized scenarios to uncertainty study demands a certain work of adaptation in terms of specification of the model and the variables of interest, as well as an appropriate expression of uncertainty sources.

Let us consider an example of a scenario piece which has as the cause the fluid overpressure on the top of the reservoir exceeding the entry pressure, as the source term it evidently has the supercritical CO₂, and the central event to evaluate is the leakage through the caprock. After attributing PDFs to the uncertain inputs of the model, which correspond to the uncertainty related to the incorrect characterization of hydraulic properties of the medium, we will see that propagating this uncertainty in order to describe the leakage debit will also give access to the cause characterization (e.g. probability that overpressure exceeds the entry pressure). So, establishing intermediate outputs for the

model can quantify the elements in the scenario.

3.4.2 Deterministic, probabilistic and level-2 probabilistic quantities of interest and critical thresholds

A potential site of CO₂ storage should meet the regulatory requirements. Whether these requirements are satisfied or not can be demonstrated through the quantities of interest. For CO₂ storage the regulatory requirements have not been globally validated yet. Therefore, several types and levels of complexity for the quantities of interest can be imagined.

As the leakage represents long continuous process (except sudden leakage on the surface equipment), the type of the quantities of interest is likely to be governed by thresholds. And again due to the nature of the CO₂ spread in the underground compartment, some of the critical thresholds are related in retrograde way: thresholds on vulnerability determine thresholds on transport characteristics, which on their turn determine thresholds on the source term (see the upper part of Figure 3.4). Hereafter, two examples of how vulnerability/exposure part of scenario can influence the quantities of interest of hazard quantification.

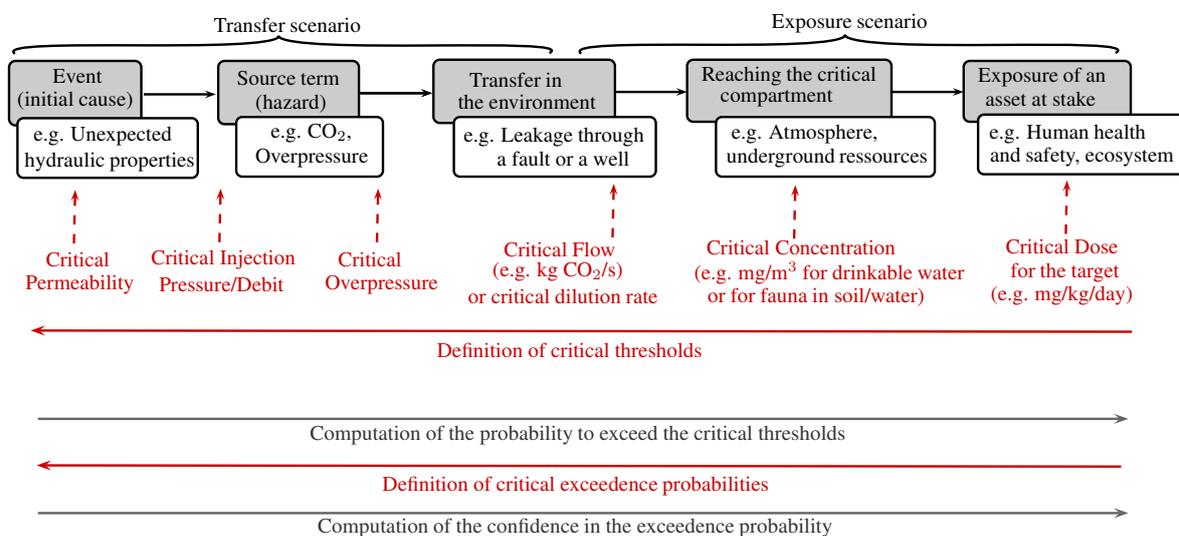


Figure 3.4: Direction of scenario deployment and uncertainty propagation, and inverse direction of critical thresholds definition. The same three levels of complexity as for the quantities of interest are needed to define the three types of critical thresholds: thresholds for physical variables, thresholds for exceedance probability and thresholds for confidence intervals on exceedance probability.

Example 1: if the assets at stake are human beings and we are studying the impacting phenomenon of slow emanation of CO₂ at surface level, the vulnerability curve could be

the probability of death (or the number of deaths if statistics are available) depending on the concentration of CO₂ in a cave. There is a critical concentration corresponding to the lethal effect. Then the hazard of leakage should be evaluated in terms of the probability of giving the concentration higher than the critical one.

Example 2: if the assets at stake are underground water resources and we are studying the impacting phenomenon of contamination, the vulnerability curve could be the risk of developing a disease by ingesting a certain amount of heavy metals (this information can be found in the norms for water quality). Again the probability of hazard would be computed as the probability of exceeding the established thresholds after transfer in the environment.

The quantities of interest determine which type of risk analysis should be performed and in which way the uncertainty sources should be represented. So, in the specification step of uncertainty analysis it would be advisable first to set the quantity of interest (which is conditioned by the vulnerability), then construct the model able to compute this quantity of interest, and determine what are the inputs for this model and the related uncertainty.

The levels of complexity of the regulatory requirements (and therefore the quantities of interest) can be the followings: the first level would be a simple exceedance of a critical threshold by a variable, the second level could be the probability that this variable exceeds the threshold, and finally the third level - confidence interval for an estimated exceedance probability. For nuclear safety the requirements are of the third type demanding reasonable expectation that the exceedance probability curve does not exceed certain fixed points (Helton et al., 1996).

Figure 3.5 illustrates three levels of the quantities of interest related to exceedance of a critical threshold on the example of the leakage through the caprock due to the overpressure on the top of the reservoir exceeding the entry pressure. The levels are embedded one in another. The first internal level contained in every analysis consists in running the physical model (representing for example the transfer of the injected substance in the environment) with fixed input variables. The choice of the values for the input variables can be based on different assumptions such as accumulation of worst-case assumptions to perform a ‘worst-case’ study, the maximum-likelihood estimates to perform a central value analysis. In the work of Paté-Cornell (1996) on the uncertainties in risk analysis, the author differentiates the ‘worst-case’ approach, quasi-worst cases and the best estimates (central values) as the first three levels of treatment of uncertainty (after the level zero corresponding to the hazard identification). What these methods have in common is the need to run the model only once for chosen inputs. In terms of quantities of interest, using such an approach allows of answering the question whether the leaking rate Q exceeds the critical value Q_c for the chosen inputs or not. In terms of cause event, the computation would help to answer whether the fluid overpressure exceeds the entry pressure.

The second box in Figure 3.5 corresponds to the probabilistic setting. The uncertainty in input variables is introduced through probability density functions. The uncertainty propagation consists in sampling the input variables according to the probability density and running the physical model for each set of inputs. Such an approach can provide a support for finding the probability that the leaking rate Q exceeds a given critical value.

The last level (the biggest box in Figure 3.5) displays the uncertainty in risk curve. It

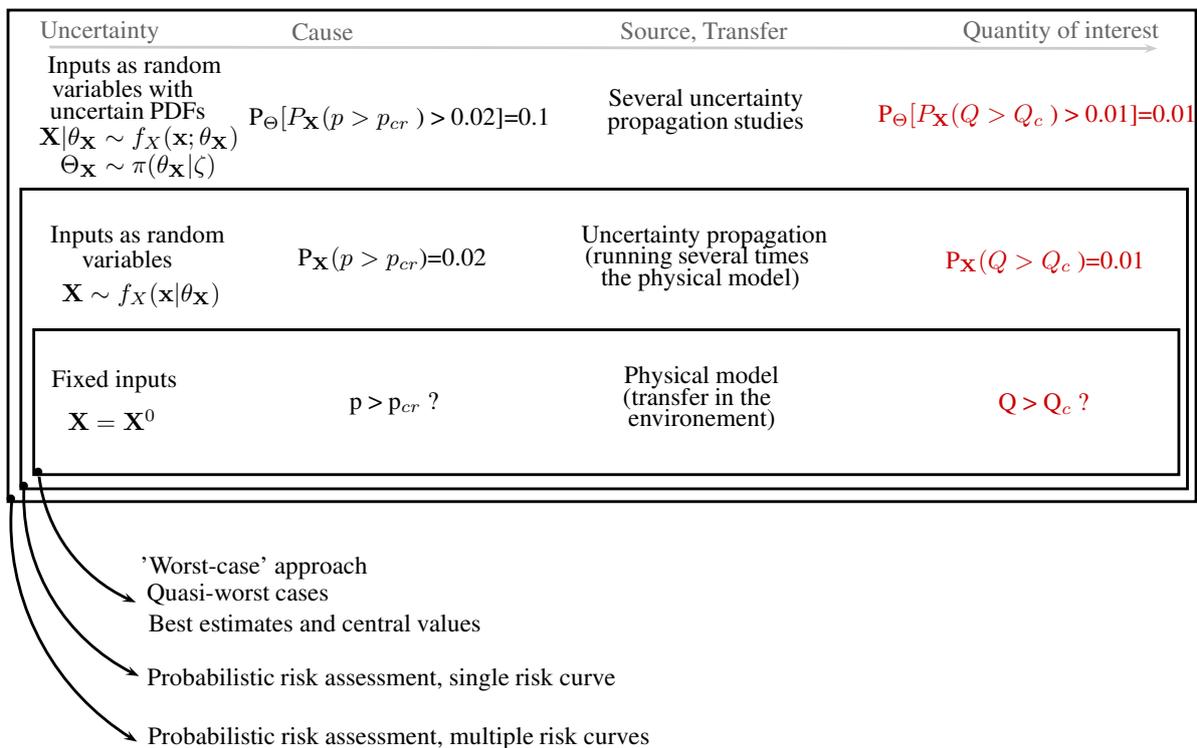


Figure 3.5: Three levels of possible quantities of interest, approaches allowing to reach these quantities of interest and corresponding uncertainty representation for input variables. The cause is an intermediate output of the model, therefore, it takes the same level of description as the quantity of interest. $P_{\mathbf{X}}$ stands for level-1 probability, P_{Θ} stands for level-2 probability, Q is the leakage rate, p is the fluid overpressure.

allows an estimation of the probability that the exceedance probability has been under- or over-evaluated. The uncertainty in input variables is introduced through a set of probability density functions. An ensemble of output risk curves provides a rich information for decision making.

The scenario evaluation with the quantities of interest related to exceedance of critical thresholds implies that these thresholds are fixed a priori (by regulatory requirements for example). But how should be computed the thresholds on the hazard part knowing the vulnerability thresholds? For the first level of thresholds this problem represents the standard inverse problem in reliability analysis. If the variable of interest is a monotonic function, it is mapped in the space of input variables and the failure zone is found. In order to find the next level threshold the conditional cumulative complementary distribution functions (CCDFs) should be computed. The values of the conditional CCDF for each point of the failure curve would give the exceedance probability thresholds.

The inversion example for one uncertain variable is given in the Figure 3.6. The first level corresponds to the following problem: consider the leakage rate as a monotonic function of caprock permeability $Q = M(K)$. Knowing the critical rate Q_{cr} determine the critical permeability K_{cr} . The solution comes from the inversion of the function M .

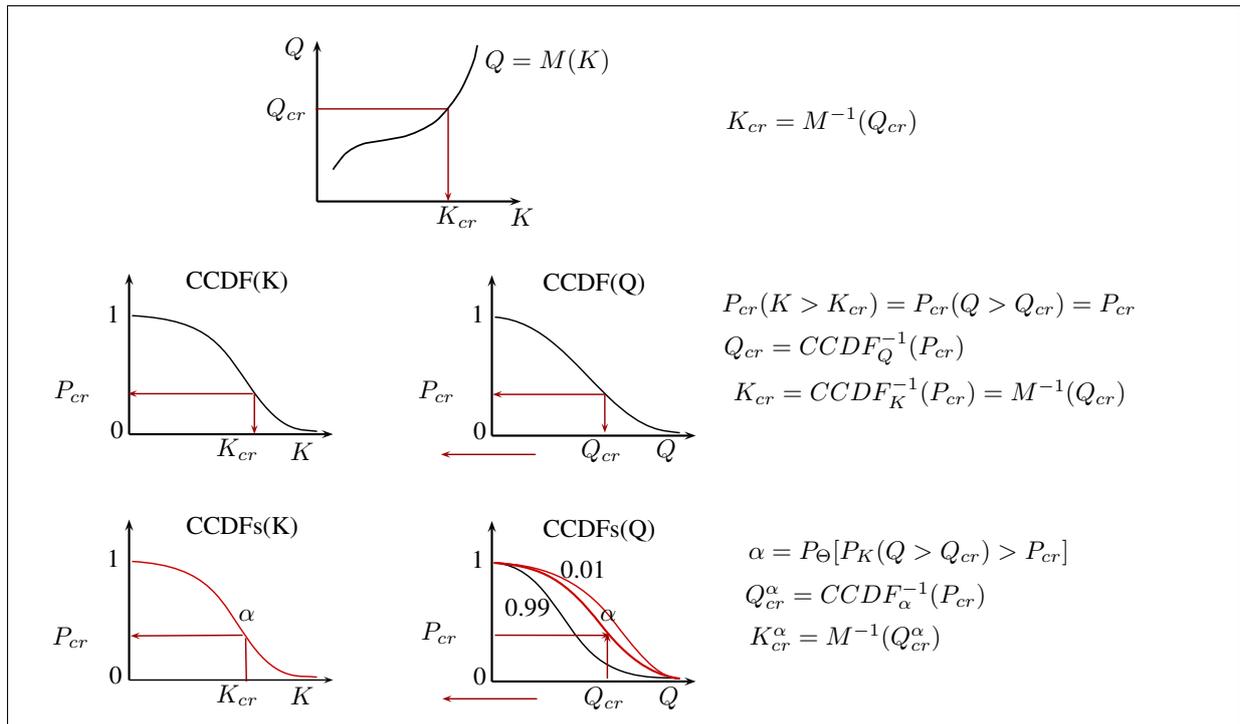


Figure 3.6: Definition of critical thresholds by inversion presuming that the leakage rate Q is a monotonic function of permeability K and $P(K; \theta_K)$ is a monotonic function of parameters θ_K . First level thresholds are obtained by inversion of the function of the model, for the first probabilistic level the exceedance probabilities of the input and the output are equal, for the second probabilistic level the inversion depends on which thresholds are fixed a priori.

The second level problem (still presuming that the leakage rate is a monotonic function

of caprock permeability $Q = M(K)$) contains four thresholds: a critical leakage rate, a critical permeability value, a critical probability that the leakage rate overcomes a certain threshold $P_{cr}(Q > Q_{cr})$, a critical probability that the permeability overcomes a certain value $P_{cr}(K > K_{cr})$. If we require one of these quantities to be fixed, all the others will be fixed as well and can be easily determined. For a monotonic function the input quantile always coincides with the output quantile and vice versa. This fact simplifies the probabilistic inversion problem as the critical probability to overcome a permeability threshold equals to the probability to overcome corresponding leakage rate: $P_{cr}(K > K_{cr}) = P_{cr}(Q > Q_{cr})$.

The last level problem can be stated in different ways depending on the available information and the thresholds that we would like to set. As an example, the leakage rate is a monotonic function of caprock permeability $Q = M(K)$ and the permeability cumulative probability is a monotonic function of parameters θ_K . In this case there exists a CCDF of the leakage rate which corresponds to the α confidence and there exists a zone of values of θ_K which delimits the family of CCDFs of leakage rate which α confidence is not respected.

How the uncertainty can be 'taken into account' in the qualitative and semi-quantitative methods? In the previous section the notion of scenario was explicit. Each scenario includes a central event. What is an event? Is it a change in the behavior of the system, whose likelihood can be estimated from the past history of similar events and whose severity could be given by an expert judgment based also on history or/and on deterministic modelling? Or is it an ensemble of possible states of the system itself for which the evolution differs from normal and which corresponds to a certain zone in the space of input parameters called in structural reliability 'failure domain'?

The first point of view is common to classical risk analysis even when events internal to the system are considered. These are qualitative and semi-quantitative approaches for which the final result of risk attributed to a given scenario is expressed through an verbal appreciation: "the scenario is rare but highly devastating" or placed in the categorical grid - the risk matrix of gravity versus probability (for examples of the methods the reader can refer to Table 3.3). For this group of methodologies Gombert and Farret (2009) propose to introduce the uncertainty related to the risk for final visualization in the form of ellipses. The authors also stress the importance of considering the variation of risk with time. As shown in the Figure 3.7 the risk matrix for the memory period about the storage site (the time before the existence of a storage site could be forgotten, which is commonly estimated up to 200-300 years) would be different from the matrix of long-term evaluation (up to 1000 years).

The meaning of the associated uncertainty is the relative appreciation of uncertainty in the severity estimation versus the uncertainty in probability estimation for a given event (oblongness and orientation of an ellipse) as well as the comparison of the uncertainty between different events (size of ellipses). As an example in Figure 3.7 (third panel), the authors propose that there is a greater uncertainty in the estimation of the impact of a seism on the low leakage (S → LL) than in the estimation of the probability of a seism; the risk estimate for the impact of drilling of a new well on low leakage (D → LL) is more uncertain than the risk caused by the physico-chemical alteration (PC → LL).

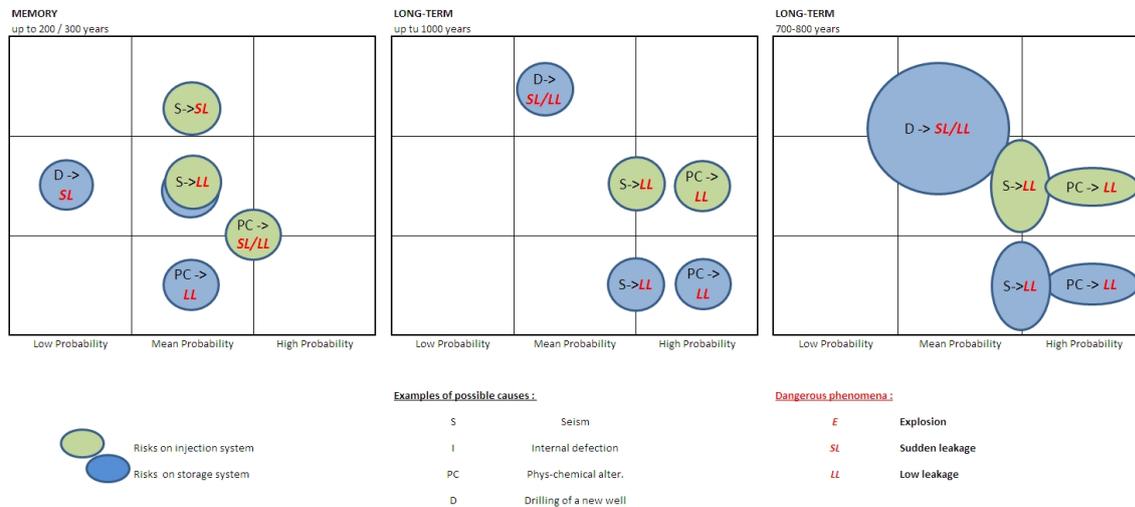


Figure 3.7: Representing the uncertainty factor of semi-quantitative risk analysis results on a risk matrix (Gombert and Farret, 2009). Example of risk matrix: i) memory phase, ii) long-term, iii) long-term with uncertainty

3.5 Classical uncertainty propagation methods illustrated on a simplified physical model

This section is devoted to the comparison of several existing techniques of uncertainty propagation and sensitivity analysis. Each presented technique is illustrated on the analytical model of (Nordbotten and Celia, 2006) for the quantity of interest of an excessive lateral spread of the CO₂ cloud in the receiving aquifer (reservoir). The simplicity of the model allows any kind of study without significant constraints on computation. The particular form of the analytical formula also allows to derive analytically the exact probability density function of the variable of interest in the case of lognormally distributed inputs. Therefore, the convergence rates of the applied techniques can be analyzed and compared to this reference. In order to lighten the text the results of the performed deterministic study and the details of the analytical computation of the response PDF are given in Appendix B. Here we will focus on traditional probabilistic approach including double probabilistic setting for estimating the influence of subjective uncertainty. The emerging theories (Demster-Shafer theory, imprecise probabilities, p-boxes, etc.) would not be discussed in this work. The comparison of results for all the applied techniques are presented in Section 3.5.3.

Uncertain parameters

All the parameters of the model except the control parameters of the duration and the injection rate have been considered uncertain: the maximal relative gas permeability kr_{max} , the reservoir thickness H , the reservoir total porosity ϕ , the irreducible water saturation (or maximal gas saturation Snw_{max}), brine viscosity μ_w and gas viscosity μ_{nw} . These pa-

rameters form the random vector $\mathbf{x} : \mathbf{x} = (kr_{max}, H, \phi, Snw_{max}, \mu_w, \mu_{nw})$. Let us consider that the expected value $E(x_i)$ and the coefficient of variation $CV(x_i)$ for each parameter are fixed by an expert. The values chosen for this particular study are summarized in Table 3.6. As soon as these characteristics are fixed, the variance is calculated as: $Var(x_i) = (E(x_i) \cdot CV(x_i))^2$.

Table 3.6: The choice of the mean value and the coefficient of variation for the input parameters.

Parameter x_i	Description	$E(x_i)$	$CV(x_i)$	$Var(x_i)$
kr_{max}	gas relative permeability endpoint	0.2 [-]	25%	$2.5 \cdot 10^{-3}$
H	reservoir thickness	50 [m]	10%	25
ϕ	porosity	0.15 [-]	10%	$2.5 \cdot 10^{-4}$
Snw_{max}	maximal gas saturation	0.66 [-]	10%	$4.4 \cdot 10^{-3}$
μ_w	water viscosity	$3.5 \cdot 10^{-4}$ [Pa·s]	10%	$1.2 \cdot 10^{-9}$
μ_{nw}	gas viscosity	$5 \cdot 10^{-5}$ [Pa·s]	20%	10^{-10}

For probabilistic analysis lognormal probability density functions are associated to the uncertain input parameters. Assuming that the initial probabilistic information available on the input variables are the mean value and the coefficient of variation (see Table 3.6), the lognormal parameters μ and σ are computed^a (see Table 3.7). The model responses for sets of input parameters modes, medians and mean values are almost coinciding (less than 0.1% difference). These values are computed to be later compared to the mean response in a probabilistic set.

Table 3.7: Calculated characteristics of the lognormal input parameter PDFs and the response r_{max} calculated for the modes of input parameters, the medians and the mean values.

parameter x_i	μ	σ^2	mode	median	mean= $E(x_i)$
kr_{max}	-1.64	$6.06 \cdot 10^{-2}$	$18.3 \cdot 10^{-2}$	$19.4 \cdot 10^{-2}$	$20 \cdot 10^{-2}$
H	3.90	$0.99 \cdot 10^{-2}$	49.3	49.8	50
ϕ	-1.9	$0.99 \cdot 10^{-2}$	$14.8 \cdot 10^{-2}$	$14.9 \cdot 10^{-2}$	$15 \cdot 10^{-2}$
snw_{max}	-0.42	$0.99 \cdot 10^{-2}$	$66 \cdot 10^{-2}$	$67 \cdot 10^{-2}$	$67 \cdot 10^{-2}$
μ_w	-7.96	$0.99 \cdot 10^{-2}$	$3.5 \cdot 10^{-4}$	$3.5 \cdot 10^{-4}$	$3.5 \cdot 10^{-4}$
μ_{nw}	-9.92	$3.92 \cdot 10^{-2}$	$4.7 \cdot 10^{-5}$	$4.9 \cdot 10^{-5}$	$5 \cdot 10^{-5}$
corresponding $r_{max}, [m]$			1167	1168	1168

3.5.1 Probabilistic analysis: global approach

As opposed to the deterministic response surface (the examples and results are reported in Appendix B) the goal of a probabilistic analysis is not only to explore the values that

^aKnowing the mathematical expectation and the coefficient of variation of one of the inputs x_i , the parameters of its lognormal probability density function (μ and σ) are computed as follows: $\mu = \ln(E(x_i)) - \frac{1}{2} (1 + \ln(CV^2(x_i)))$; $\sigma^2 = 1 + \ln(CV^2(x_i))$.

the variable of interest can take, but also to know what is the related probability of these values. The main hypothesis is that the resulting response uncertainty is entirely defined by the input uncertainty and the model.

Probabilistic analysis can be divided in local and global approaches. In *local* sensitivity analysis the behavior of a function is studied only around a given point (for more details and results for the simplified physical model the reader is invited to refer to the continuation of Appendix B, where an analytical computation of the response PDF is demonstrated as well). In *global* approach the entire domain of input parameters is of interest.

The objective of the global sensitivity analysis is to identify and to rank the input variables that drive the uncertainty of the model output. The practical objective is to detect the most influential parameters and if possible to reduce the probabilistic model dimension by neglecting the less influential parameters. Global sensitivity analysis is a stochastic approach based on the joint probability density functions of the output and the inputs. The classical method for uncertainty propagation and global sensitivity analysis is the Monte Carlo sampling discussed hereafter.

Monte Carlo simulation

Monte Carlo simulation is a classical method of approximating integrals ([Metropolis and Ulam, 1949](#)). As probability is an integral of its density, Monte Carlo has become a classical method for uncertainty and sensitivity analysis in computational stochastic studies as well ([Rubinstein, 1981](#)).

There exists a variety of techniques for accelerating the convergence of Monte Carlo simulation. Generally it consists in a particular way of sampling (which is no more completely random). The most known is Latin hypercube sampling ([Helton and Davis, 2003](#)).

Uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling were compared by [Helton et al. \(2005\)](#) for a two-phase fluid flow. The study has shown that uncertainty and sensitivity analysis results with the two sampling procedures are similar and stable across the replicated samples. In the same article the authors come to an important conclusion that the effects of subjective uncertainty can be assessed with much smaller sample sizes than the effects of stochastic uncertainty. This point is discussed further in the context of a ‘two-level’ uncertainty propagation and its simplification.

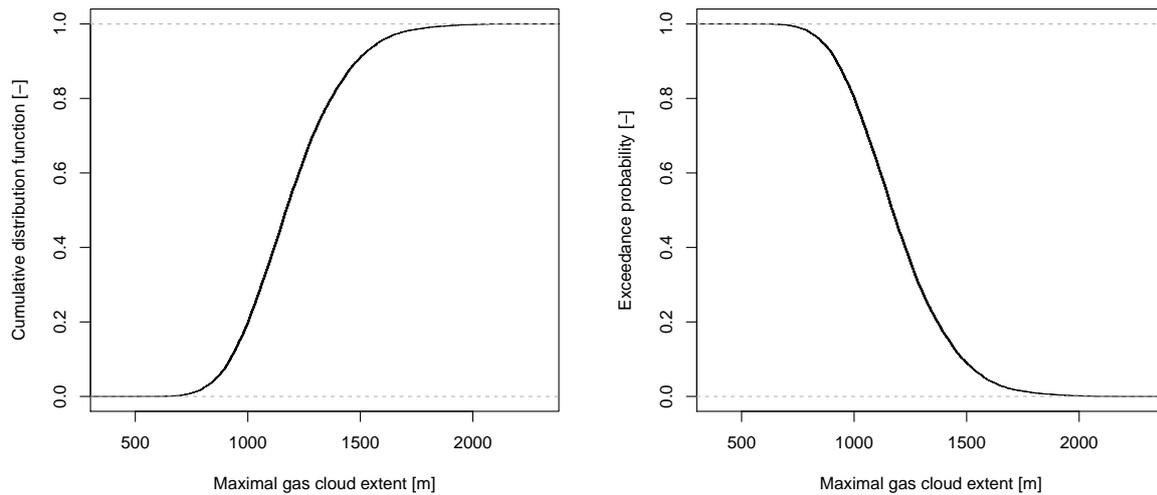
In the present work the standard Monte Carlo method has been applied to the simplified physical model of CO₂ injection into a deep aquifer for the same input parameter PDFs as detailed in Table 3.7. The estimates of the quantities of interest and their standard errors for the performed standard Monte Carlo simulation with N=10⁴ model runs are summarized in Table 3.8. Among the quantities of interest the threshold probability estimation is the least accurate: the relative standard error goes up to 3.4%.

The estimated cumulative distribution function is shown in Figure 3.8. Its complementary curve (CCDF) represents the threshold exceedence probability.

Figure 3.9 illustrates the variability between the results of ten replications of Monte Carlo simulations with the same number of samples and different random seeds. The coefficient of variation for the mean of r_{max} of these 10 replications is 0.19% compared to the asymptotic coefficient of variation of the estimation 0.22%. For the standard deviation

Table 3.8: Monte Carlo estimates of the quantities of interest for N=10⁴ model runs and the corresponding propagation uncertainty (De Rocquigny, 2012, p.290).

Quantity of interest	Value	Asymptotic standard deviation	
		Expression	Value
Median(r_{\max})	1168 m	$\sqrt{\frac{\pi}{2}} \frac{\sigma(r_{\max})}{\sqrt{N}}$	2.8 m
$\mu(r_{\max})$	1188 m	$\frac{\sigma(r_{\max})}{\sqrt{N}}$	2.2 m
Var(r_{\max})	$4.9 \cdot 10^4 \text{ m}^2$	$\frac{1}{N} \text{Var}(r_{\max})$	4.9 m^2
$\sigma(r_{\max})$	222 m	$\frac{\sigma(r_{\max})}{\sqrt{2N}}$	1.6 m
$P(r_{\max} > 1500 \text{ m})$	$8.9 \cdot 10^{-2}$	$\sqrt{\frac{P(1-P)}{N}}$	$3 \cdot 10^{-3}$

Figure 3.8: Left: the empirical cumulative distribution function of the maximal horizontal spread of the injected CO₂ (r_{max}) for Monte Carlo study with 10000 samples. Right: the empirical complementary cumulative distribution function (CCDF), this complementary function allows of reading directly the values of threshold exceedance probabilities.

the respective values are 0.59% and 0.71%.

The resulting scatterplots are shown in Figure 3.10. The scatterplots and the Pearson correlation coefficients show that the most significant input variables are the gas relative permeability endpoint (kr_{max}) and the gas viscosity (μ_{nw}). Therefore, in Figure 3.11 the output is plotted against these two inputs. The values of r_{max} exceeding 1500 m are highlighted in red. Returning to the question of thresholds definition, we remind that in the case of a multiple uncertain inputs no threshold can be computed for each input separately, but only for the totality of the input space a 'failure domain' can be determined. Such as in Figure 3.11 (left) there is no boundary delimiting a 'failure domain' because

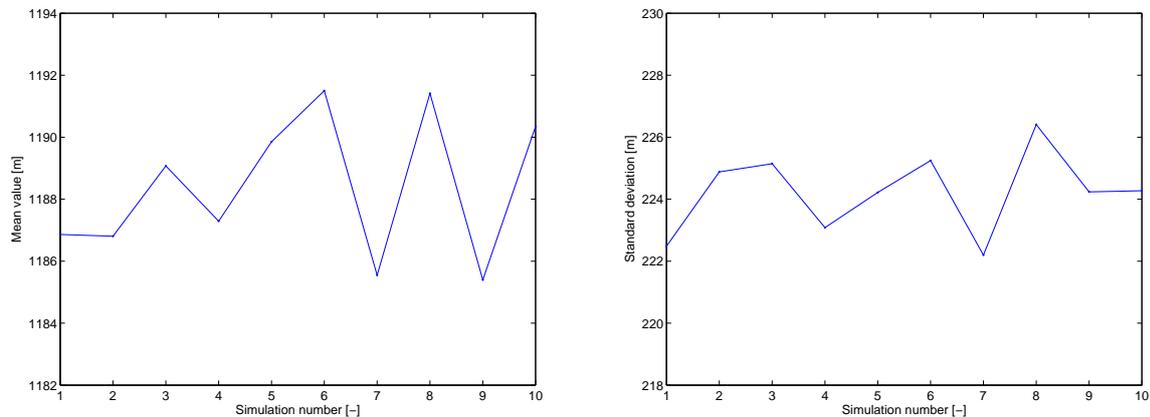


Figure 3.9: Mean value (left) and standard deviation (right) of the variable of interest r_{max} for 10 Monte Carlo replications of 10000 samples each.

of the structure of the function r_{max} and the contribution of the uncertainty of the other four inputs. Whereas if we consider only kr_{max} and μ_{nw} as uncertain input variables (see the left panel of the Figure 3.11), the boundary becomes distinct.

In flow problems the ratio of the relative permeability and viscosity of a phase is often considered characterizing the mobility of the phase. The model under consideration is not an exception as the boundary in Figure 3.11 (right) is linear. Therefore, an attempt to determine a threshold on the ratio the relative permeability and viscosity can be undertaken. Figure 3.12 illustrates the points for which the maximal lateral extent of the CO₂ cloud exceeds 1500 m (red points) as function of the ratio kr_{max}/μ_{nw} . For the right panel of the Figure a threshold on the output defines the threshold on the ratio $kr_{max}/\mu_{nw} < 1.5 \cdot 10^{-4}$, whereas for the left panel defining a threshold on the ratio ($kr_{max}/\mu_{nw} < 2.6 \cdot 10^{-4}$) eliminates also the points which are not in the ‘failure domain’. Nevertheless, such a threshold can be fixed in situations when it is important to insure the avoidance of the ‘failure domain’. Another way to consider the problem would be introduction of *partial safety factors*.

To summarize the results of Monte Carlo simulation:

- The most influential input parameters are the gas relative permeability (45%) and gas viscosity (26%).
- Any level-1 probabilistic quantity of interest can be estimated through Monte Carlo simulation. The estimation of threshold exceedence probability requires a high number of samples.
- The estimations of quantities of interest together with the propagation errors are reported in Table 3.8.

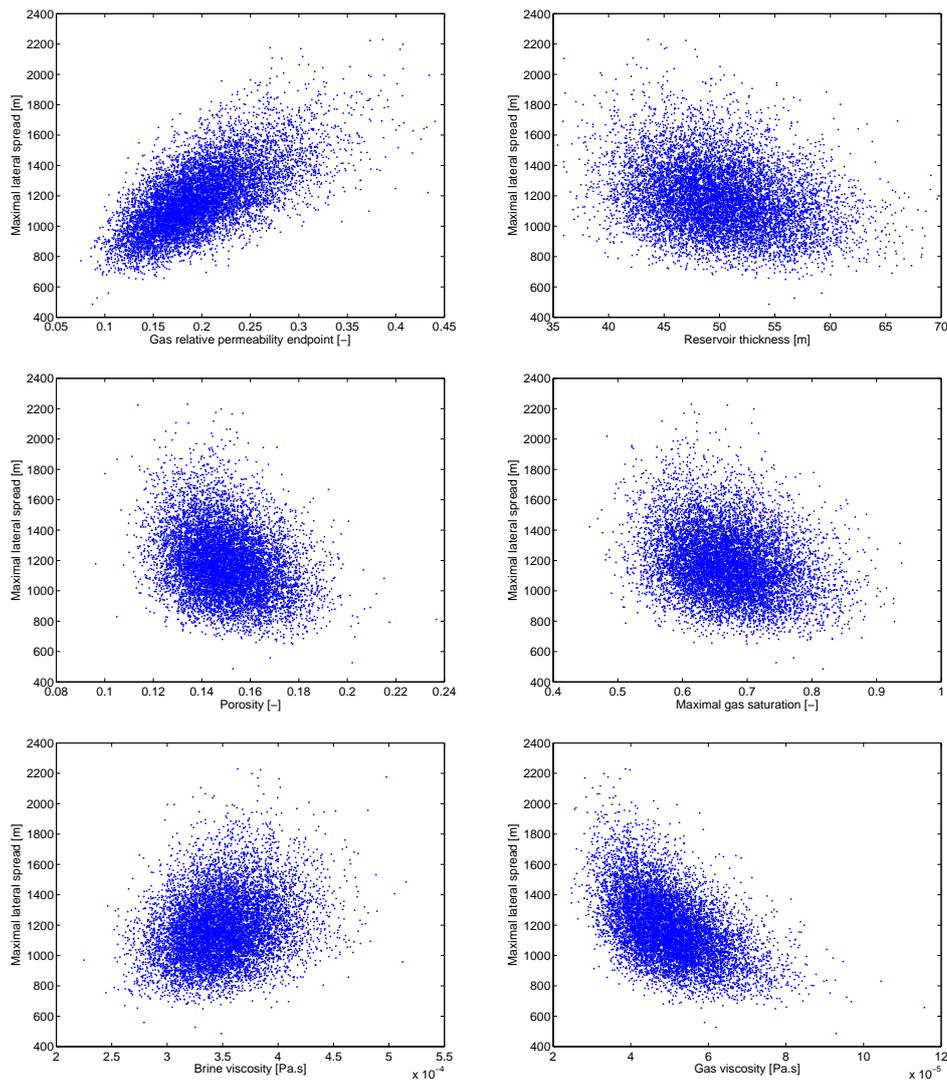


Figure 3.10: Experimental design of the 10000 standard Monte Carlo simulation. Variable of interest is the maximal lateral extent of the CO₂ cloud. Pearson correlation coefficients are respectively [0.66 -0.28 -0.27 -0.26 0.25 -0.50].

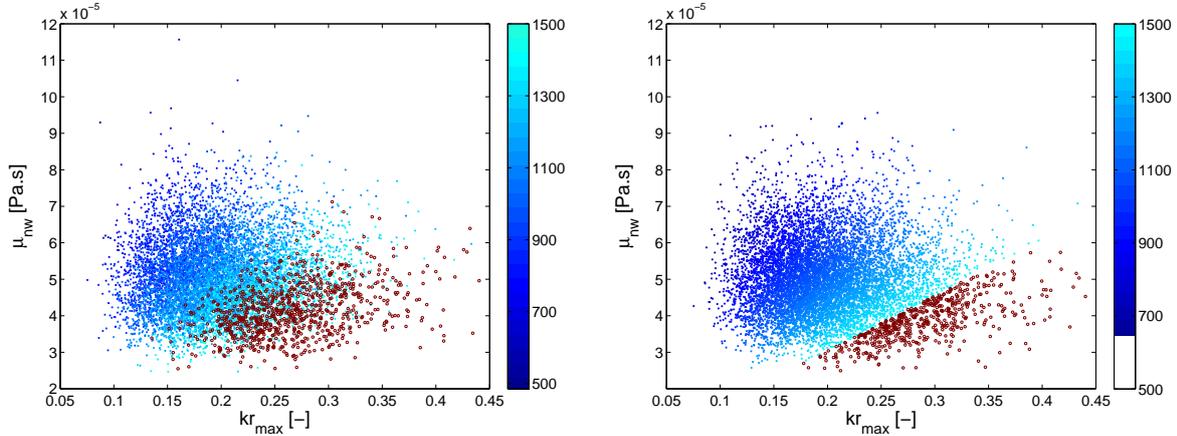


Figure 3.11: Left: the output as a function of the two most important inputs: the gas relative permeability endpoint (kr_{max}) and the gas viscosity (μ_{nw}). The red circles indicate the points for which the maximal lateral extent of the CO₂ cloud exceeds 1500 m. Right: the same figure if kr_{max} and μ_{nw} were the only uncertain inputs.

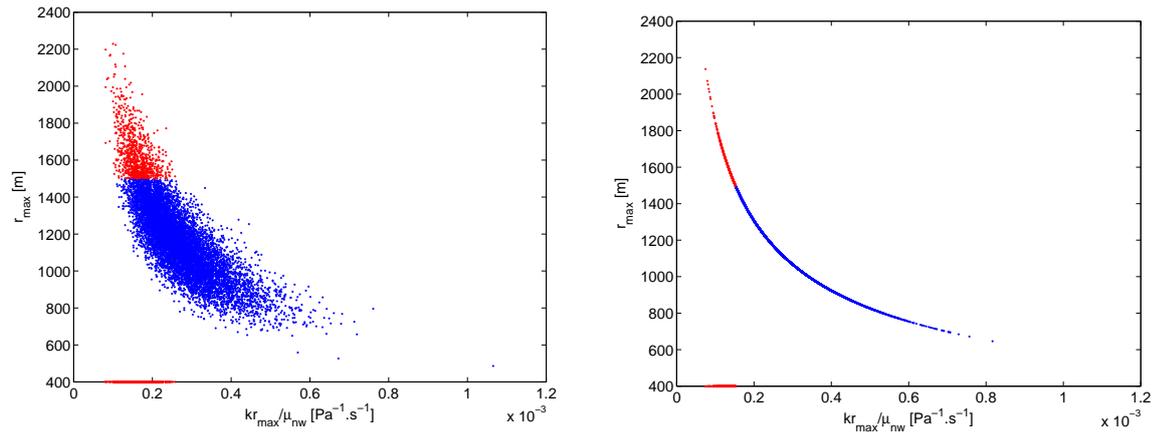


Figure 3.12: Left: the maximal lateral extent of the CO₂ cloud as a function of the ratio of the gas relative permeability endpoint (kr_{max}) and the gas viscosity (μ_{nw}). All the output points exceeding 1500 m have the ratio $< 2.6 \cdot 10^{-4}$. Right: the same figure if kr_{max} and μ_{nw} were the only uncertain inputs. The output points exceeding 1500 m have the ratio $< 1.5 \cdot 10^{-4}$.

Meta-model by Polynomial Chaos Expansion

Meta-models (or response surfaces, surrogate models) are widely used in computational sciences to predict the output of a black-box model at any point, knowing the response of the model at a number of other points (called 'design of numerical experiments' or 'learning sample'). The advantage of meta-models is their computational lightness compared to actual models. The goal of engaging a meta-model is to extract as much information as possible from the simulations in a limited number of points and use it for prediction of the model response. There are several approaches to build a meta-model. Recently, the following methods have received a lot of attention in the domain of structural reliability: Polynomial Chaos Expansion (PCE) reminded by Roger and Ghanem (1998) and further studied and applied in different fields by Xiu and Karniadakis (2003); Sudret (2008a,b); Dossantos-Uzarralde and Guittet (2008); Blatman (2009); Formaggia et al. (2012); Ashraf et al. (2013), Kriging (such as Gaussian process Marrel et al. (2008, 2009); Sergienko et al. (2013)) and Support Vector Machine (SVM).

Hereafter, the reader can find an example of PCE meta-model construction for the simplified physical problem of CO₂ injection in an aquifer (semi-analytical solution described in Section 2.5). The Polynomial Chaos Expansion theory is provided in Appendix C.

Application: The six parameters form the random vector \mathbf{x} appearing in (C.1): $\mathbf{x} = (kr_{max}, H, \phi, S_{nw_{max}}, \mu_w, \mu_{nw})$. For the simplicity of result comparison the same lognormal probability density functions as in previous examples are attributed to each parameter (see Table 3.7). The variable of interest is still the maximal spread of the gas cloud, r_{max} . Once a set of input parameters $\mathbf{x}^0 = (kr_{max}^0, H^0, \phi^0, S_{nw_{max}}^0, \mu_w^0, \mu_{nw}^0)$ is fixed, the corresponding model output r_{max}^0 is calculated through the simplified physical model of Nordbotten (Equation 2.48).

In order to determine the least-square estimate of the expansion coefficients β as shown in (C.6), the experimental designs $\{\ln(r_{max}^j), j = 1, \dots, N\}$ of size $N = 50, 100, 1000$ have been considered.

In Figure 3.13 the empirical CDF of the output samples on which the meta-model is built (solid green line) is compared to the exact CDF calculated analytically (dashed blue line, for the details of computation see Subsection B.3), for $N = 50, 100, 1000$ (subfigures (a), (b) and (c) respectively). Even the meta-model built out of 50 samples reproduces the response variability quite closely; for $N = 100$ the approximation is very accurate without showing significant losses with respect to $N = 1000$. Therefore, the number of $N = 100$ simulations has been retained for meta-model construction for the numerical predictive model, as it seems to balance between the precision in the CDF estimate and the corresponding computational cost.

The coefficients to be included in PCE have been selected following the sparse PCE approach proposed by Blatman (2009); Blatman and Sudret (2010). Having fixed the maximal degree of polynomials to $p = 4$, 45 coefficients proved to be significant for the expansion (with the last index of 206). The most important seven first coefficients have the following values 1187.9, 146.2, -58.9, -59.1, -58.8, 58.9, -117.7. The estimates of the

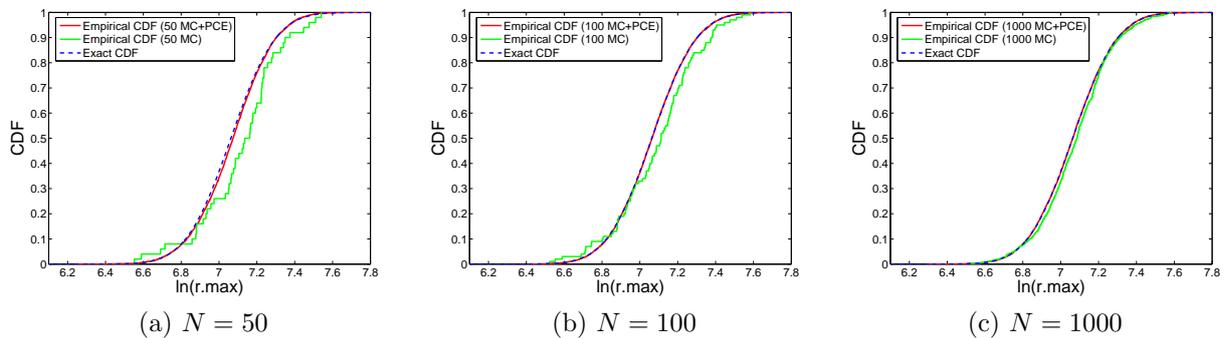


Figure 3.13: Comparison of the meta-model CDF (red line) with the exact CDF (dashed blue line) and with the empirical CDF of the samples used for the meta-model construction (green line). The figures correspond to the analytical solution with varying number of the samples for PCE decomposition: (a) 50 samples, (b) 100 samples, (c) 1000 samples.

Table 3.9: Sobol' indices

Parameter	Sobol' index	Sobol' total index
kr_{max}	43%	44%
H	7%	7%
ϕ	7%	7%
Snw_{max}	7%	7%
μ_w	7%	7%
μ_{nw}	28%	29%

first two moments of the response PDF have been derived by using (C.5), obtaining:

$$\begin{aligned}\mu(r_{max}) &= 1188 \text{ m} \\ \sigma(r_{max}) &= 224 \text{ m}.\end{aligned}$$

These estimates are compared with a standard Monte Carlo simulation and other techniques results in section 3.5.3.

Partial and total Sobol' indices have been computed using (C.11) and (C.10); the results are reported in Table 3.9.

According to the Sobol' total indexes (Table 3.9, second column), the parameters which seem to influence the most the response variability are the maximal relative gas permeability kr_{max} (43%) and gas viscosity μ_{nw} (28%). The same conclusions can be drawn by considering the Sobol' indexes reported in the first column of the Table 3.9, meaning that the influence of the input parameters on the output is mainly due to single effects, while mixed effects play a minor role.

3.5.2 ‘Level-2’ probabilistic analysis

As it was mentioned above, in risk assessment uncertainty is often divided in two types: aleatory (or random, stochastic, variability) due to intrinsic variability in a system behavior and epistemic (or subjective, state of knowledge) due to lack of knowledge.

There exists a point of view that these two types of uncertainties are hierarchically different. The models of aleatory events contain parameters which are poorly known. Therefore, so-called ‘two-level’ setting (or also ‘two-tier sampling’, ‘sampling of double probabilistic measures’ (De Rocquigny, 2012, p.298), ‘double loop’) can be applied, for which the first level refers to the aleatory variables and the second level helps to account for the information scarcity used for the probabilistic description of the input variables. Commonly, if only the first level is considered, the uncertainty model in probabilistic setting, is defined by the probability density function parametrized with $\theta_{\mathbf{X}}$:

$$\mathbf{X} \sim f_X(\mathbf{X} = \mathbf{x}|\theta_{\mathbf{X}}).$$

If both levels are considered, the sample space is extended to $\Omega_x \times \Omega_\theta$ which contains all the possible combination of the values of \mathbf{X} and $\theta_{\mathbf{X}}$. The uncertainty model becomes conditional to the random vector $\theta_{\mathbf{X}}$:

$$\begin{aligned} \mathbf{X}|\theta_{\mathbf{X}} &\sim f_X(\mathbf{x}; \theta_{\mathbf{X}}) \\ \Theta_{\mathbf{X}} &\sim \pi(\theta_{\mathbf{X}}|\zeta) \end{aligned} \quad (3.3)$$

where $\pi(\theta_{\mathbf{X}}|\zeta)$ is a given PDF parametrized by ζ representing the epistemic uncertainty, whereas f_X stands for the aleatory part. The sampling of such a structure can be done by two embedded loops: the outer loop samples $\Theta_{\mathbf{X}}$ and for each set of parameters $\theta_{\mathbf{X}}$ the inner loop performs a standard probabilistic study on the aleatory variables. The interest of a two-level setting is that it gives an estimate of the confidence interval of the output variability characteristics, assuming a certain level-2 uncertainty.

In order to add to the study the level-2 uncertainty the parameters of the lognormal PDFs in Table 3.7 are considered uncertain themselves.

First, let us consider the situation where the mean values for the input variables are given with an error of 10% whereas the coefficient of variation is left unchanged. The PDF variability in this case is shown in Figure 3.14. A uniform law in an interval $[\text{Mean}(x_i) - 0.1 \text{Mean}(x_i); \text{Mean}(x_i) + 0.1 \text{Mean}(x_i)]$ is attributed to the mean value of each input variable. For each sample the characteristics of corresponding lognormal PDF are calculated as in Equations a and a.

In this case introducing the uncertainty on PDFs leads to a mean CCDF shifted to higher risk values than the initial one (see Figure 3.15). It means that not considering the level-2 uncertainty can lead to underestimation of the risk by typically 5%.

The results for the uncertainty on the coefficient of variation are presented in the Figure 3.16

Sankararaman and Mahadevan (2013) propose a variance-based global sensitivity analysis method for evaluating the level-1 and level-2 variability contributions.

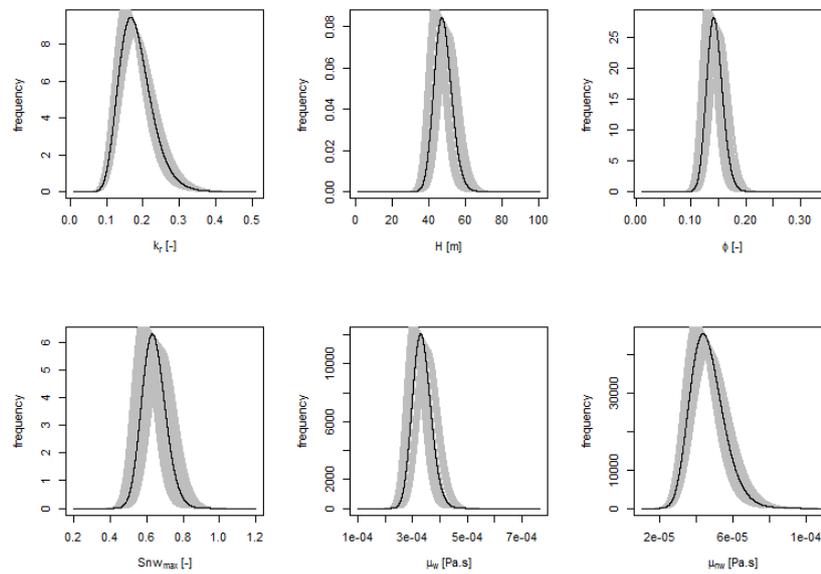


Figure 3.14: The initial lognormal PDFs of the input variables (in black) and 100 PDFs used for level-2 uncertainty evaluation (in gray).

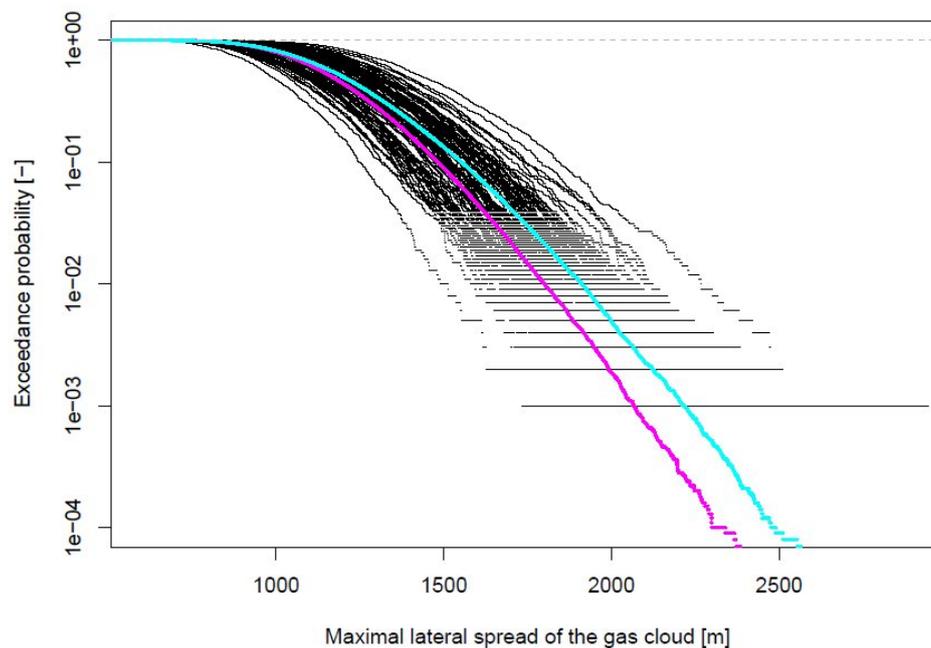


Figure 3.15: Two-level setting results for mean value error of 10% for each of 7 input variables: ensemble of 100 CCDFs of the maximal lateral spread of CO₂ cloud after 10 years of injection. Magenta line - Monte Carlo level-1 10⁵ simulation; black lines - results for each perturbed set of entry PDFs obtained with 10³ simulations each; aqua line - the mean of black lines (the level-2 mean): 10² × 10³ = 10⁵ simulations.

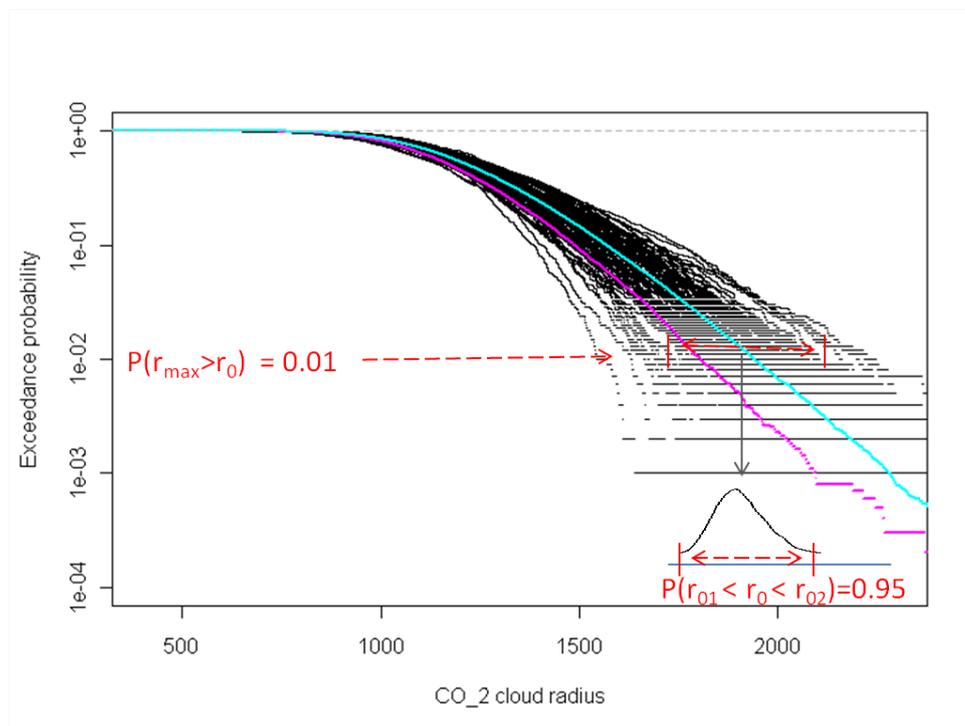


Figure 3.16: Two-level setting results for variation coefficient error of $\pm 7\%$. Illustration of the confidence interval 95% that exceedance probability is of 0.01. $r_0=1900$, $[r_{01}, r_{02}]=[1650, 2150]$

3.5.3 Comparison of the results

The first statistical moments as well as exceedance probability computed through previously described uncertainty propagation techniques are summarized in Table 3.10.

Let us first remind that for lognormal input PDFs, the PDF of the output (maximal lateral extent of CO₂ cloud after ten years of injection) can be computed exactly, which gives the first row in Table 3.10. The second row of the Table, the first order Taylor approximation, requiring only 12 model runs (two times the number of parameters) reaches an estimation of standard deviation which is only 1.5% different from analytically calculated value. Since lognormal PDFs have been considered for the entry parameters, the nominal point (the output for the mean values of inputs, taken as the mean value for the first order Taylor approximation) differs from the output mean. For this method the exceedance probability is not accessible.

Monte Carlo simulation shows a good convergence rate for the mean value whereas the standard deviation and exceedance probability are still underestimated after 10^5 model runs.

Among the applied techniques PCE metamodel gives the best evaluation of the standard deviation and exceedance probability demanding only 10^2 calls to the model. The estimation of the mean value has a relative error of only 0.1%.

Considering the second level uncertainty gives an appreciation of an error related to the incorrect choice of PDF characteristics for the input variables. Allowing the input

Table 3.10: Comparison of uncertainty propagation results from the studies on the simplified model of CO₂ injection into saline aquifer. The same entry PDFs are taken for all methods except double Monte Carlo sampling for which the mean values of input PDFs are considered uncertain following uniform law in an interval $\pm 10\%$. The nominal value equals to 1168m.

Methods	Mean [m]	Median [m]	Std [m]	$P(r_{\max} > 1500 \text{ m})$
Analytically computed output PDF	1189	1168	224	$9.0 \cdot 10^{-2}$
First order Taylor approximation	1168	-	221	-
Monte Carlo simulation 10^5 model runs	1189	1168	222	$8.9 \cdot 10^{-2}$
PCE Metamodel 10^2 model + 10^4 metamodel runs	1188	1169	224	$9.1 \cdot 10^{-2}$
Double Monte Carlo sampling $10^2 \times 10^3$ model runs				
Bilateral confidence interval 95% [quantile 2.5% - quantile 97.5%]	[1085- 1422]	[1073 -1402]	[198 -263]	$[3.2-35.8] \cdot 10^{-2}$
Characteristics of 10^5 outputs	1224	1200	247	$13.3 \cdot 10^{-2}$
Maximal relative error of a value within the 95% interval with respect to the initial exact output PDF	19.7%	20.1%	17.5%	297.1%

mean values to vary in a range of $\pm 10\%$, the confidence intervals can be computed for output mean, standard deviation and exceedance probability. One can notice that these intervals are rather large: the maximal relative error of a value from the 95% confidence interval with respect to the initial PDF (first line of the Table) is of order of 20% for the output mean, median and standard deviation. But the most sensible quantity of interest is the exceedance probability whose relative error can go up to 300%

In order to analyze the influence of the PDF type on the quantities of interest, log-normal, normal, triangular and mixte laws have been compared (see the results in Table 3.12). The relative error between the mean value estimations is of 0.4%.

Table 3.11: Comparison of the sensitivity indexes.

Parameter	Taylor approximation Normalized importance	Monte Carlo Pearson index	Monte Carlo Spearman index	Metamodel Sobol' total index
kr_{max}	44%	45%	45%	44%
H	7%	8%	8%	7%
ϕ	7%	8%	7%	7%
Snw_{max}	7%	7%	7%	7%
μ_w	7%	7%	6%	7%
μ_{nw}	28%	26%	26%	29%

Table 3.12: Comparison of uncertainty propagation results for Monte Carlo simulations on different types of input PDFs. Each simulation consists of 10^4 runs

Entry PDFs type	Mean [m]	Median [m]	Std [m]	P($r_{max} > 1500$ m) and 95% confidence interval	P($r_{max} > 2000$ m) and 95% confidence interval
Lognormal	1188	1168	222	$8.9 \cdot 10^{-2}$ [8.3 9.5] · 10^{-2}	$1.7 \cdot 10^{-3}$ [0.9 2.5] · 10^{-3}
Normal	1190	1174	236	$9.7 \cdot 10^{-2}$ [9.1 10.3] · 10^{-2}	$4.4 \cdot 10^{-3}$ [3.1 5.7] · 10^{-3}
Triangular	1188	1169	236	$9.8 \cdot 10^{-2}$ [9.2 10.4] · 10^{-2}	$2.3 \cdot 10^{-3}$ [1.3 3.2] · 10^{-3}
Mixte	1185	1165	224	$8.7 \cdot 10^{-2}$ [8.2 9.3] · 10^{-2}	$2.2 \cdot 10^{-3}$ [1.2 3.1] · 10^{-3}

3.6 Concluding remarks

General remark about uncertainty and critical thresholds in risk assessment

All the quantitative risk assessment methodologies can be differentiated according to their relation towards uncertainty. The most common practice of deterministic risk assessment does not explicitly include uncertainty. Its output is a single value of risk for each scenario. To give an appreciation of imprecision in this single value of risk, an expert can attribute an error bound for probability and severity on the basis of intuition. It can be done even for qualitative and semi-quantitative methodologies.

For methodologies introducing parameter uncertainty in a probabilistic way the output represents a risk curve (risk as a function of the variable of interest which value is uncertain due to the uncertain character of inputs). The main approaches in this group are: scenario-based, system-model-based, Markov chains. The correspondence between scenario-based and system-model approach can be established. The system-model approach provides the most unrestricted choice of the uncertainty propagation technique.

Finally, to appreciate how different would be the risk curve if the probabilistic laws of input parameters were uncertain, a level-2 probabilistic study can be conducted. The result

represents a set of risk curves. This level can be seen as a way of level-1 probabilistic results justification. The choice of input uncertainty representation is often based on insufficient data and, therefore, is quite subjective. The level-2 analysis gives idea of how sensible is the response to the input PDFs changes.

Each of these three levels of uncertainty treatment give access to a quantity of interest of the corresponding level. The same three levels apply on the critical thresholds in risk assessment: a threshold can be set on a variable itself, on its exceedance probability or on the confidence of its exceedance probability evaluation. As illustrated in this Chapter the definition of most critical thresholds requires inversion. In the case of one input variable, deterministic and level-1 probabilistic thresholds can be computed instantly. Even 'level-2' can be treated easily assuming monotony. Nevertheless, for the problem of underground multi-phase flows, a great number of factors influences the response. Fixing a threshold on one of the inputs disregarding the values of the others influential entries is conservative, it helps to avoid the critical zone but it also eliminates a part of response which is not critical. Furthermore, fixing such thresholds on several inputs can reduce the zone of accepted response to non-existence. Therefore, instead of fixing the thresholds on some of the inputs separately, it is preferable to delimit the critical zone in the space of all or most influential input variables. This fact complicates the definition of the regulatory requirements.

Performances of uncertainty propagation and sensitivity analysis techniques

In this chapter the central event of leakage by lateral migration is used to illustrate and compare the ability of main uncertainty propagation techniques to reach the quantity of interest of exceedance probability. Taking into account the particularities of the problem (such as high non-linearity and large space and time scales) the metamodelling seems to be the only means of probabilistic uncertainty propagation applicable on geological CO₂ storage. It is demonstrated that the polynomial chaos expansion metamodel provides a good approximation of the exceedance probability even being built with a quite limited number of model runs. Therefore, this technique is retained for the uncertainty propagation on the numerical model presented in the next chapter.

The further quantitative analyzes through numerical models focus on level-1 probabilistic setting. We address level-2 variability in Section 4.2.

Chapter 4

Quantification of scenarios using numerical models

Contents

4.1	Meta-models for maximal lateral extent and fluid overpressure	64
4.1.1	Uncertainty sources description	65
4.1.2	Uncertainty propagation	66
4.1.3	Sensitivity analysis	70
4.2	Level-2 uncertainty propagation using meta-model and Taylor approximation	72
4.3	Results for homogeneous permeability	74
4.4	Concluding remarks	76

In Chapter 3 a methodology of risk assessment for CO₂ storage elaborated by INERIS has been presented as well as the main techniques of uncertainty propagation. The methodology is based on scenario analysis, therefore the contribution of the thesis consists in quantifying chosen scenarios defined in the global workflow. The scenarios considered in the present chapter are normal evolution with leakage through slightly permeable caprock and large lateral extent of the injected fluid possibly reaching a fault. Among the uncertainty propagation techniques for global analysis described in Section 3.5.3 only deterministic or stochastic response surface (meta-model) seem applicable to computationally expensive numerical models as it requires a rather limited number of runs compared to other techniques.

A stochastic meta-model is constructed to represent the response of the system through polynomial chaos expansion. The first goal is to evaluate the mean gas overpressure on the top of the reservoir to be able to predict the average flow escaping the caprock. The second goal is to find the probability density function of the maximal lateral extent to be coupled with the probability of reaching a fault to give an estimate of leakage through a fault.

4.1 Meta-models for maximal lateral extent and fluid overpressure

The theory of polynomial chaos was developed by (Wiener, 1938, 1962). The first application of the method was performed by (Meecham and Siegel, 1964; Siegel et al., 1965; Meecham and Jeng, 1968) to study turbulence. At that time it was criticized to have a slow convergence rate (Orszag and Bissonnette, 1967; Crow and Canavan, 1970; Chorin, 1974). Therefore, it was not popular till the work of (Ghanem and Spanos, 1991), who proposed to use the Polynomial Chaos Expansion for spectral representation of uncertainty in the context of finite element methods. Since then it has been applied in several domains allowing uncertainty and sensitivity analysis for complex systems (Blatman and Sudret, 2010; Sudret, 2008b; Formaggia et al., 2012; Oladyskhin et al., 2011). Some overviews can be found in (Sudret and Der Kiureghian, 2000; Xiu and Karniadakis, 2002; Xiu et al., 2002).

In order to quantify the two central events, the lateral leakage out of the storage complex and slow percolation through the caprock, two meta-models have been constructed using the same experimental design of 100 simulations with a numerical model.

To describe the study and report the computation results, the four steps of uncertainty treatment (De Rocquigny, 2012) have been followed:

- A. The specification step: quantities of interest, model, inputs and outputs of the model
- B. The quantification of uncertainty sources
- C. The uncertainty propagation
- C'. The sensitivity analysis

Specification step

The variables of interest chosen for the analysis are i) maximal lateral extent of the CO₂ cloud after 10 years of injection, ii) the overpressure on the top of the reservoir integrated on the surface of 1km of radius and over 10 years iii) the maximal fluid overpressure reached on the top of the aquifer and in the proximity of injection zone.

The quantities of interest are i) the threshold exceedance probability curve (CCDF) for the maximal lateral extent of the CO₂ cloud, ii) mean and standard deviation of the averaged overpressure on the top of the reservoir, probability of exceeding 1MPa, iii) mean value, standard deviation and maximal value of the maximal overpressure (in space and time) in the whole aquifer and on the top of it.

The model used in this study is the numerical model in 2D axisymmetric formulation described in section 2.4.1.

Input variables. Seven model parameters are considered uncertain:

- Maximal gas relative permeability kr_{max} [-]
- Porosity ϕ [-]
- Thickness of the reservoir H [m]
- Maximal gas saturation Snw_{max} [-]
- Intrinsic permeability K_{int} [m²]
- Thickness of the damaged zone around the injection well EDZ [m]
- Injection debit Q [Mt/year]

4.1.1 Uncertainty sources description

The uncertainty on the homogeneous inputs of the numerical model are described through probabilistic marginal laws. The quantitative details of the chosen laws are presented in the Table 4.1.

The three lognormal PDFs (for gas relative permeability endpoint, intrinsic permeability and the thickness of the damaged zone) are constrained to the first two statistical moments by fixed mean values (corresponding to the initial set of inputs) and chosen coefficients of variation in the same way as described previously in section 3.5.1. For the rest of input uncertain parameters (porosity, reservoir thickness, maximal gas saturation and injection rate) the triangular PDFs are chosen as most often the information available on these parameters is presented in the form of ‘minimum - mean - maximum’. To compute the characteristics of these triangular laws the intervals of plausible values (which could be fixed by an expert or found from literature) are associated with the bilateral 99.7% confidence intervals. The most probable values are set to the initial values.

Table 4.1: Chosen characteristics of the input variables. ‘LogN’ and ‘Tr’ stand for lognormal and triangular PDFs respectively

Input variable	Mean	Bilateral interval 99.7%	CV	Attributed PDF
kr_{max}	0.2 [-]	[0.09, 0.40]	25%	LogN(-1.64, 0.25)
ϕ	0.15 [-]	[0.115, 0.185]	10%	Tr(0.113, 0.15, 0.187)
H	120 [m]	[108, 132]	4.3%	Tr(107.31, 120, 132.69)
Snw_{max}	0.67[-]	[0.45, 0.89]	14%	Tr(0.44, 0.67, 0.90)
K_{int}	$1 \cdot 10^{-13}$ [m ²]	$[2.2 \cdot 10^{-14}, 3.53 \cdot 10^{-13}]$	50%	LogN(-30.05, 0.47)
EDZ	10 [m]	[2.2, 36.3]	50%	LogN(2.19, 0.47)
Q	1 [Mt]	[0.8, 1.2]	8.6%	Tr(0.79, 1, 1.21)

For the level-2 uncertainty study the reader can refer to the section 4.2.

The geometry of the domain has been explored separately and has shown little influence on the lateral cloud spread, whereas the influence on the pressure is more significant.

4.1.2 Uncertainty propagation

In order to propagate the described sources of uncertainties to the variables of interest a meta-model is constructed using sparse polynomial chaos expansion implemented in [OpenTURNS \(2012\)](#). To do so, first a standard Monte Carlo simulation is performed with a limited number of samples (N=100). Then the basis multivariate orthogonal functions (Hermite polynomials) are generated. And finally, the coefficients of the expansion are computed according to the least squares strategy. The theory of polynomial chaos expansion and further explanations are given in [Appendix C](#). Once the meta-model is constructed it is explored through Monte Carlo simulations.

The meta-model built for the quantity of interest of the maximal lateral spread contains 9 significant coefficients (the last significant coefficient has the index 196). The meta-model for the averaged overpressure numbers 23 coefficients (with the last coefficient having the index 327).

Results

1. The empirical CCDF of **the maximal lateral extent** of the CO₂ cloud is shown in [Figure 4.1](#).

The mean value of the maximal radius in the present analysis (1123 m) coincides with the response for the mean values of the input variables. It underlines that the relation between r_{max} and input variables for the numerical model differs from the analytical solution of Nordbotten (results in [Table 3.12](#)), for which the use of any entry PDFs (lognormal, normal, triangular, mix of lognormal and triangular) gives a mean value always higher than the nominal value by at least 1.4% in accordance with the Jensen’s inequality ([Jensen, 1906](#)). Therefore, it can be assumed that the numerical model does not represent a strongly convex function of its inputs.

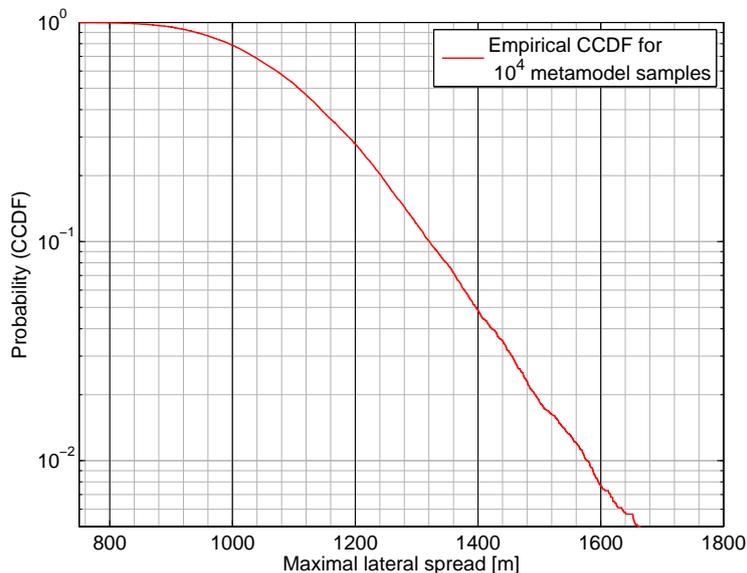


Figure 4.1: Empirical CCDF (or threshold exceedance probability curve) of the maximal lateral spread of the CO_2 cloud after 10 years of injection. The probability that r_{max} exceeds 1500 m estimated with the meta-model equals to 0.0185.

The analytical model is limited by the gravity factor, therefore, to test uncertainty propagation techniques on this model a reduced thickness of the aquifer has been considered (with the mean value of 50 m versus 120 m in the present numerical study). That is why the CCDF of the maximal lateral extent obtained with analytical model is shifted to the right: to occupy the same porous volume with lower layer thickness larger lateral distance needs to be swept by the injected gas. In both cases the uncertainty propagation leads to rather large variance of the maximal lateral extent: the distance doubles from 99% to 1% quantile. The standard deviation obtained through the analytical model (224 m) is higher than for the numerical model (155 m). It can still be explained by the reduced aquifer thickness as well as difference in considered input variables.

2. Averaged overpressure on the top of the reservoir:

The mean value of the fluid overpressure on the top of the reservoir obtained with the metamodel is 0.63 MPa, which is smaller than the nominal value 0.70 MPa. Such result is less surprising when the scatterplots are analysed (see the second panel of Figure 4.5): the averaged fluid overpressure on the top of the reservoir is negatively correlated with the most influential input variable which takes 94% of importance (intrinsic permeability). The estimate of the standard deviation is 0.24 MPa. The entire empirical CCDF is presented in Figure 4.2.

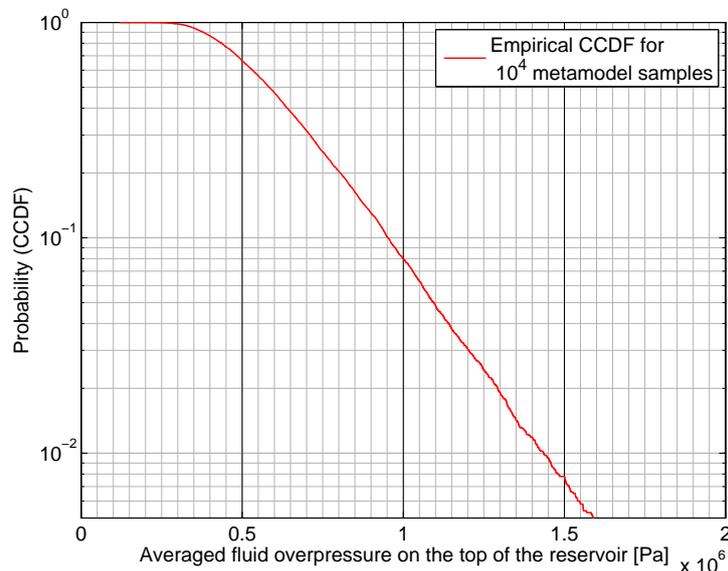


Figure 4.2: Empirical CCDF (the threshold exceedance probability curve) of the averaged overpressure on the top of the reservoir. The probability that the averaged overpressure exceeds 1 MPa estimated with the meta-model equals to 0.08

Some authors consider the value of 1 MPa as the critical threshold for the fluid pressure in the caprock which provokes fracturing. This value depends on the compression regime of the rock in any particular site. The injection-induced changes in the horizontal stress inside the reservoir are much more relevant than in the vertical stress which may be small because the ground surface can move freely. [Rutqvist J. \(2007\)](#) evaluate a critical sustainable injection pressure as 27.2 MPa for a compressional stress regime. For extensional regime the critical pressure is much higher. In our case the injection pressure is not constant as the injection is controlled by the injection rate. But taking the mean overpressure along the injection zone 4.9MPa together with the initial hydrostatic pressure 23.5MPa makes the mean injection pressure of 28.4 MPa. It means that if the reservoir is in compression regime the injection debit of 1 Mt per year is too high.

3. Maximal overpressure characteristics:

The mean value of the maximal overpressure in the reservoir (which corresponds to the center of the injection zone) is 4.9 MPa (with the standard deviation of 2.1 MPa). The maximum of the overpressure in the reservoir is 11 MPa.

On the top of the reservoir the mean value of the maximal overpressure is 1.6 MPa with the standard deviation of 0.6 MPa. This pressure is higher than supposed threshold on the caprock resistance (1 MPa). Which underlines once more that the

chosen injection rate is overstated. The maximum of the overpressure on the top of the reservoir is 3.6 MPa.

These results prove the capacity of stochastic metamodeling to produce estimations of probabilistic risk measures on the basis of numerical modeling. The main observations are too high fluid overpressure and high variability of the maximal radius for chosen input characteristics. In what follows we present the ways of verification of the metamodel results.

Verification of the meta-model

As there is no Monte Carlo representative reference for verification of the meta-model, the relative errors between several meta-models are analyzed. The meta-model for which the results are presented above consists of $N = 23$ expansion coefficients. Each coefficient contributes to the fitting of the meta-model to the data. First coefficients representing linear model are the greatest; the higher the number of coefficients in a meta-model, the higher statistical moment can be captured by the meta-model. To justify the choice of the order of expansion and the number of simulations used to build a particular meta-model (the support), the author proposes following considerations.

1. Different truncation order of polynomials

The truncated expansion on a basis of n -dimensional Hermite polynomials of degree not exceeding p is:

$$\tilde{z} = \sum_{k=0}^{P-1} \beta_k \Psi_k(x), \quad (4.1)$$

where $P = \binom{n+p}{p}$. For the same support of 100 numerical simulations, several meta-models of total degrees $p = 1, \dots, 7$ have been constructed in order to determine the best truncation order. The comparison is reported in Figure 4.3. Starting from degree 4, the meta-model gives reasonable approximation of the response mean, standard deviation and exceedance probability. Therefore, it has been chosen for the study.

2. Cross-validation of the meta-model by 'leave-one-out' technique

Cross-validation consists in dividing the support into subsets: one of these subsets is supposed to be a set of known data (training set) and the other(s) - supplementary observations for testing the model (testing or validation set). Exhaustive cross-validation relies on all possible divisions of the dataset into training and testing sets of given size. The 'leave-one-out' cross-validation is an exhaustive cross-validation with only one observation left as a testing set. If \tilde{z}_{-i} is a meta-model constructed from the full experimental design excluding the i -th observation, the predicted residual is defined as the difference between the i -th observation and its value predicted by the meta-model \tilde{z}_{-i} : $\Delta^i = z(x_i) - \tilde{z}_{-i}(x_i)$.

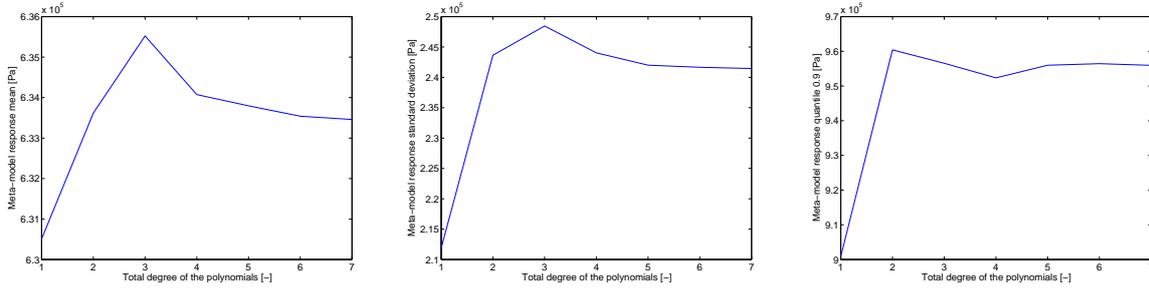


Figure 4.3: Response mean, standard deviation and exceedance probability for meta-models of different total degree ($p = 1, \dots, 7$) constructed on the same support (100-sample standard Monte Carlo simulation).

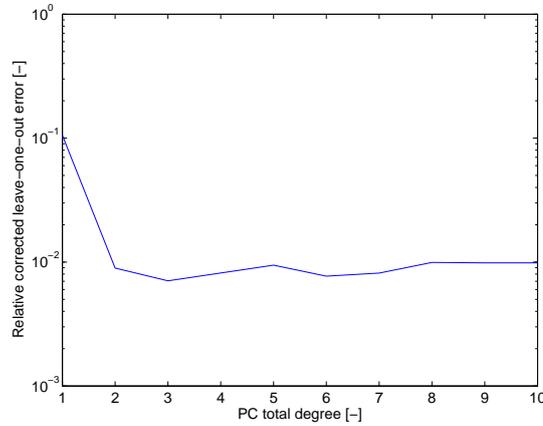


Figure 4.4: Relative corrected leave-one-out error estimate for the meta-model of the averaged overpressure on the top of the reservoir. Optimal total degree equals to 3.

The values of the relative corrected leave-one-out error for the meta-models of different truncation order are reported in Figure 4.4. With respect to this error the optimal degree is equal to 3. The order 4 chosen previously shows a low error as well, therefore it is retained for the final results.

4.1.3 Sensitivity analysis

The variance-based sensitivity analysis provides the Sobol's sensitivity indexes (contributions of input variables and their interactions to the output variance). For PCE meta-models the Sobol's sensitivity indexes can be computed from the expansion coefficients as described in Appendix C. Tables 4.2 and 4.3 report the Sobol's sensitivity indexes and total Sobol's sensitivity indexes for both variables of interest.

The sensitivity indexes are appreciably different for the two variables of interest. The intrinsic permeability takes more than 90% of sensibility for the averaged overpressure, for maximal lateral spread it takes only 41% but still stays the most important parameter. Figure 4.5 illustrates that the maximal lateral spread is positively correlated to the intrinsic permeability whereas the averaged pressure has a negative correlation.

For the averaged overpressure the sensitivity indexes are different from total sensitivity indexes (almost 20% of relative difference for the injection debit). It means that coupled terms are more significant for the averaged pressure than for the lateral spread.

Surprising results are the insignificance of the damaged zone radius and extremely low contribution of the porosity for both variables of interest.

Table 4.2: Sensitivity indexes for the integrated overpressure computed through PCE

Input variable	Sobol' index	Total Sobol' index
kr_{max}	$4 \cdot 10^{-3}\%$	$7 \cdot 10^{-2}\%$
ϕ	1%	1%
H	$2 \cdot 10^{-1}\%$	$2 \cdot 10^{-1}\%$
Snw_{max}	1%	1%
K_{int}	93%	94%
EDZ	$< 10^{-6}\%$	$3 \cdot 10^{-2}\%$
Q	5%	6%

Table 4.3: Sensitivity indexes for the maximal lateral spread computed through PCE

Input variable	Sobol' index	Total Sobol' index
kr_{max}	10%	10%
ϕ	$< 10^{-6}\%$	$< 10^{-6}\%$
H	14%	14%
Snw_{max}	32%	32%
K_{int}	40%	41%
EDZ	$< 10^{-6}\%$	$< 10^{-6}\%$
Q	3%	3%

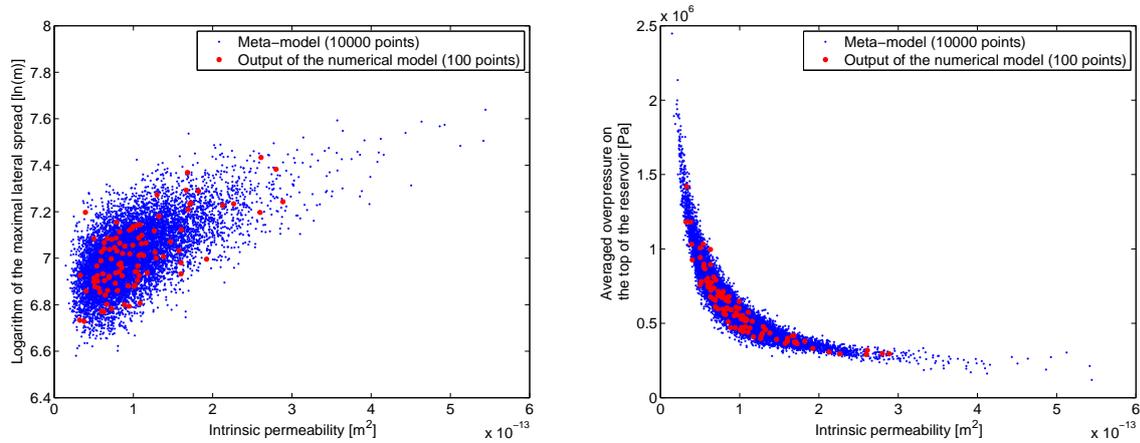


Figure 4.5: Visualization of the simulation results: both variables of interest and their meta-models as functions of the most influential input variable - the intrinsic permeability. To view the dependencies on the rest of the input variables the reader can refer to Appendix D.

4.2 Level-2 uncertainty propagation using meta-model and Taylor approximation

In Section 3.5.2, an uncertainty propagation with level-2 probabilistic setting has been described and applied to a simplified physical model of CO₂ spread. The level-2 risk measures have been reached through double Monte Carlo sampling. With numerical model such an approach is not accessible as it demands thousands of model runs. Nevertheless, in the present section we propose that even in the case of time-consuming numerical models the second level quantities of interest can be estimated through metamodeling and Taylor approximation.

As it has been shown in Section 4.1 the maximal lateral spread of the CO₂ cloud after 10 years of injection is positively correlated with all the input variables (see Figure D.2). Therefore, we suppose that the response threshold exceedence probability is monotonous with respect to the mean values of the input PDFs. Assuming additionally that the epistemic deviations are limited (uncertainties are supposed small enough for Taylor approximation), a significant simplification can be done in introducing the level-2 computations. The principle is to undertake parameter perturbation for the characteristics $\theta_{\mathbf{X}}$ of the chosen PDFs (see Table 4.1) to estimate the level-2 variation through first-order Taylor decomposition for variance (Equation B.1). The function f on which Taylor method is applied is:

$$\theta_{\mathbf{X}} \rightarrow C_z(\theta_{\mathbf{X}}, \mathbf{d}),$$

where $\theta_{\mathbf{X}}$ is the set of parameters defining the PDFs of \mathbf{X} components, C_z is the quantity of interest, \mathbf{d} is a vector of fixed parameters of the model.

Ignoring the second order contributions, the parameter perturbation can be performed in 'one at a time' setting. To introduce level-2 variability the mean values of input PDFs have been considered uncertain and perturbed by 10% the coefficient of variation in the direction of least values.

Table 4.4: Computation of the sensitivity of the r_{max} variance ($C_z = \text{Var}(r_{max})$) towards the level-2 perturbations; computation of the importances assuming that the coefficient of variation at levels 1 and 2 are equal.

Parameter $\theta_i = \mathbb{E}(X_i)$	Nominal value of θ_i θ_{i_0}	Perturbation of θ_i : -0.1 CV θ_{i_0}	Sensitivity (S) $\left(\frac{\partial f}{\partial \theta_i}(\theta_{i_0})\right)^2$	Variance $(\text{CV } \theta_{i_0})^2$	Importance (I) $S \cdot \text{Var}(\theta_i)$	Normalized importance $\frac{I}{\text{Var}(C_z)}$
θ_{krnw}	0.20 [-]	-0.005	$2.9 \cdot 10^{-2}$	$2.5 \cdot 10^{-3}$	$7.4 \cdot 10^{-7}$	3%
θ_ϕ	0.15 [-]	-0.002	$5.4 \cdot 10^{-2}$	$4 \cdot 10^{-4}$	$2.2 \cdot 10^{-5}$	1%
θ_H	120 [m]	-0.52	$6 \cdot 10^{-6}$	27	$1.6 \cdot 10^{-4}$	6%
$\theta_{Snrw_{max}}$	0.67 [-]	-0.009	$3.8 \cdot 10^{-5}$	$8.1 \cdot 10^{-3}$	$3 \cdot 10^{-7}$	$10^{-2}\%$
$\theta_{K_{int}}$	10^{-13} [m ²]	$-5 \cdot 10^{-15}$	$1.2 \cdot 10^{20}$	$2.5 \cdot 10^{-27}$	$3 \cdot 10^{-7}$	$10^{-2}\%$
θ_{EDZ}	10 [m]	-0.5	$8.9 \cdot 10^{-5}$	25	$2.2 \cdot 10^{-3}$	82%
θ_Q	1 [Mt]	-0.009	$2.8 \cdot 10^{-2}$	$8.1 \cdot 10^{-3}$	$2.3 \cdot 10^{-4}$	8%
					$\text{Var}(C_z) =$ $2.7 \cdot 10^{-3}$	

In order to avoid the impact of the statistical fluctuations, the seed of pseudo-random computer sampling is fixed. For each perturbation the same type of meta-model as described in the previous section is constructed and analyzed. As a result we obtain one CCDF curve per parameter set $\{\theta_i\}$. The corresponding CCDFs of the maximal lateral extent are presented in Figure 4.6.

The quantitative values of the normalized importance indexes are obtained for the quantity of interest of the variance of r_{max} : $C_z = \text{Var}(r_{max})$ (see Table 4.4).

Comparing the obtained importances to the level-1 sensitivities (see Table 4.3) one would notice an inverse hierarchy of parameter importancies. Such result is counter-intuitive and can be influenced by the hypothesis of equivalence of coefficients of variation at levels 1 and 2.

Generalizing the present analysis, it should be noticed that eliminating least important level-1 variables for reducing the probabilistic model may not be conservative if the subjective uncertainty on this variables is high.

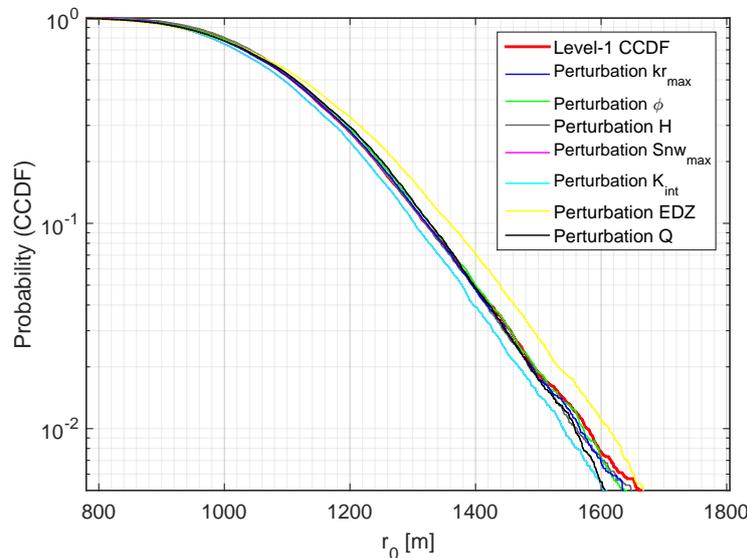


Figure 4.6: The maximal lateral spread CCDF for the level-2 perturbations of the input variables (one by one) compared to the initial level-1 CCDF.

4.3 Results for homogeneous permeability

Taking into account the spatial variability or considering homogeneous variation of the properties does not give the same results for uncertainty propagation. In order to be able to compare these approaches for the permeability, we provide hereafter the reference analysis with homogeneous permeability as the only input variable.

Figure 4.7 shows the response variables as functions of intrinsic permeability. Quasi-linear positive correlation can be observed for the logarithm of r_{max} , and hyperbolic negative correlation for the averaged pressure on the top of the reservoir. In such representation it can be noticed that the numerical instability affects significantly the maximal lateral spread. It can be explained by the great sensibility of the front position in multi-phase flow problems. The results for the overpressure are more stable, as they are averaged over time and space.

Verification of the meta-model

As already discussed for the case of multivariate entries, there is no reference for the meta-model verification, therefore results for different truncation order and support are analyzed. For the one-variable meta-model, the results for different truncation order is shown in Figure 4.8. Above order 3 all three quantities are stabilized. Therefore, the meta-model of the order 3 is retained for the study.

The relative error of the meta-model is reported in Figure 4.9. The optimal degree with respect to this error is 3 or 4.

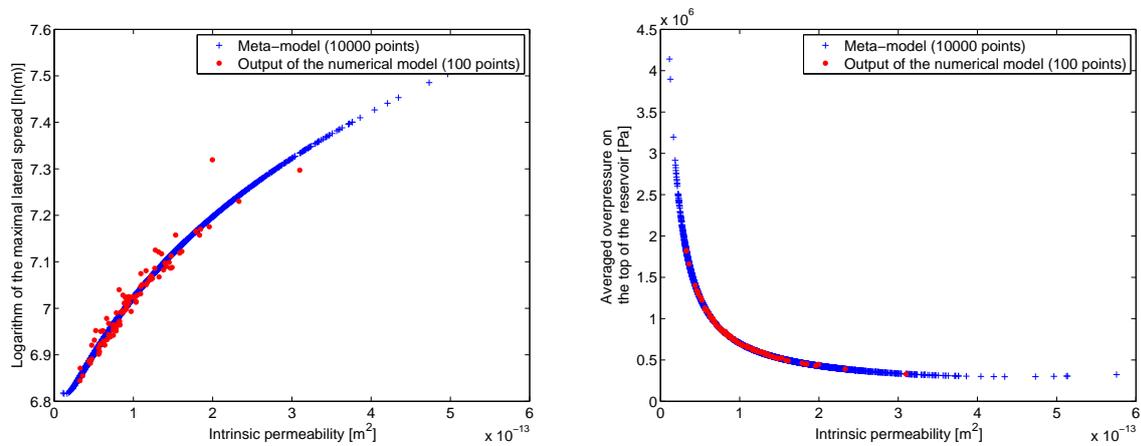


Figure 4.7: Visualization of the simulation results: both variables of interest and their meta-models as functions of the only variable permeability

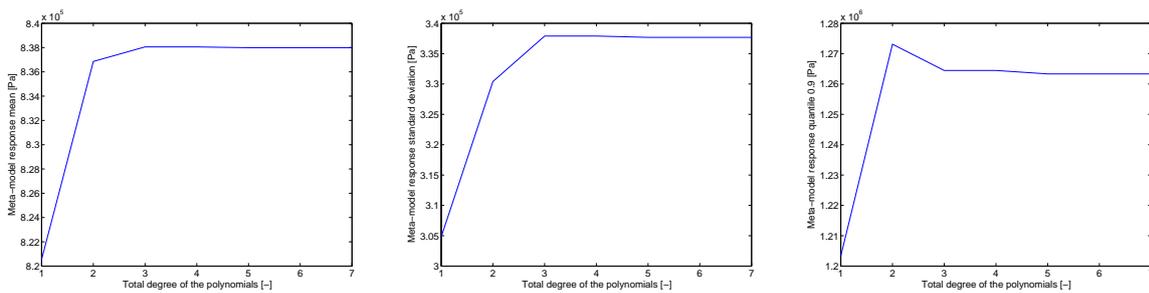


Figure 4.8: Response mean, standard deviation and exceedance probability for meta-models of different total degree of polynomials ($p = 1, \dots, 7$) constructed on the same support (100-sample standard Monte Carlo simulation)

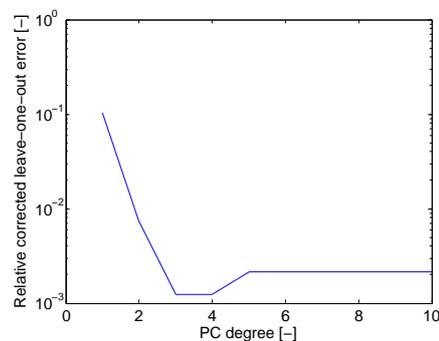


Figure 4.9: Relative corrected leave-one-out error of the meta-model for the only variable of intrinsic permeability.

4.4 Concluding remarks

The meta-models constructed using the numerical model gives access to the approximation of the cumulative complementary distribution function (or threshold exceedence probability curve). The CCDFs are evaluated for two variables of interest: the maximal lateral extent of the CO₂ cloud and the averaged overpressure on the top of the reservoir contributing to the quantification of the scenarii of the excessive lateral spread and the leakage through the caprock. In terms of risk assessment the obtained CCDFs represent probability versus intensity curves, which can be coupled with vulnerability curves resulting in the final risk curve.

The polynomial chaos expansion has been used for meta-modeling. The great advantage of this approach is that it allows the computation of the global sensitivity indexes directly from the meta-model coefficients. The sensitivity indexes calculated for both variables of interest indicate that the most influential input variable is the intrinsic permeability. It is important to notice that rather limited variation has been attributed to the intrinsic permeability (the bilateral interval 99.7% includes less than 2 orders of magnitude). According to (Bear, 1979) hydraulic conductivity can possibly vary over 13 orders of magnitude. Even with such limited variability the intrinsic permeability contributes by 90% to the averaged overpressure variance. Therefore, a particular care should be taken in describing its uncertainty. This task is impeded by the presence of heterogeneities. The uncertainty of the intrinsic permeability represented as spatial variability is considered in Chapter 5. An uncertainty propagation study with homogeneous permeability as the only input variable has been run in order to provide a reference for comparison with the spatial variability case.

The level-2 probabilistic study has been performed on the numerical model indicating that the level-2 importances do not coincide with the level-1 importances. Considering the level-2 quantity of interest (such as the variance of the mean value or the variance of the variance of the maximal lateral extent), we obtain that the biggest contribution comes from the input variable which is insignificant in level-1 (thickness of the damaged zone around the injection well). Therefore, this parameter needs to be considered in spite of small level-1 contribution.

Chapter 5

Spatial variability

Contents

5.1 Geological heterogeneity	78
5.1.1 A matter of scale	78
5.1.2 Spatial variability and uncertainty related to its representation as a random field	79
5.2 Random tensor field of permeability: modelling and simulation results	80
5.2.1 Theoretical basis of the generation of anisotropic random field of permeability	80
5.2.2 Simulated anisotropic random field of permeability	84
5.3 Probabilistic study for random tensor field of permeability .	86
5.3.1 Results for media with heterogeneous orthotropic and anisotropic permeability	86
5.3.2 Statistics on the maximum radius with variability	88
5.4 Permeability uncertainty: homogeneous versus heterogeneous variation	97
5.4.1 Comparison of the results	98
5.4.2 Inference of the homogeneous equivalent permeability distribution	101
5.5 Concluding remarks	104

Geological heterogeneities are recognized as a major feature and constraint of quantitative hydrogeology especially for model calibration or flux prediction (Eaton, 2006). They are ubiquitous due to the diversity of natural processes at stake in the genesis of the underground media.

The geological spatial heterogeneities (which can be presented as 'natural spatial variability' uncertainty type mentioned in the Introduction 1.2) add a degree of complexity to risk assessment. The existence of facies spatial distribution, natural variability within facies, structural features such as faults, folds, fractures makes it difficult to assess the suitability of sites for long-term CO₂ storage.

The uncertainty about the underground properties combines geological spatial heterogeneities, incomplete information based on measurements of variable quality, subjectivity of information interpretation. Thus, theoretically this uncertainty can be both represented as a lack of knowledge (as we are not able to make enough measurements, these measurements being moreover imprecise and indirect) and as stochastic uncertainty (we could never know the media properties at each point corresponding to a REV of the scale of interest, and modelling them as random is the only option left).

In this chapter it is demonstrated by numerical simulations that accounting or not for heterogeneity of the medium leads to results differing by an order of magnitude for certain variables of interest. Therefore, the way of dealing with the uncertainty related to the underground properties matters to uncertainty propagation, and an expert performing risk assessment should pay attention to how spatial heterogeneity is taken into account.

5.1 Geological heterogeneity

5.1.1 A matter of scale

Even if heterogeneities in the properties of natural medium is commonly acknowledged, the way of describing them in quantitative way is a subject of wide discussion in the scientific literature. The main stumbling point of heterogeneity description when modelling flows in porous media is the scale. As we are not able to include all the physical phenomena at all scales for quantifying the processes, the model equations often refer to a certain characteristic length, whereas the natural media is heterogeneous at all scales. Therefore, the challenge consists in obtaining the characteristics at the right scale relying on finer or coarser description locally available. For underground media we will distinguish two scales of measurements: one macro scale (e.g. measurements of type well-test) and one corresponding to measurements on core samples and well logging data.

In order to include fine scale information into coarser model either the data should be averaged or the problem should be homogenized. The difference between averaging and homogenization is clarified by Wood (2009). The homogenization theory can be found in the following works: Allaire (1992); Olla (1994); Jikov et al. (1994). The particular case of flow through highly heterogeneous media was described by Panfilov (2000). The case of double-porosity media was also considered by Arbogast et al. (1990); Bourgeat et al. (1998); Panfilov (2000). The upscaling becomes even more complex in presence of multi-phase flow (Ewing, 1997; Das and Hassanizadeh, 2005).

The link between uncertainty propagation and model upscaling/downscaling should be understood to justify the use of spatial models as decision support systems. Only few works are available on the subject. The relative contribution of uncertain model inputs to the variance of aggregated model output under a change of spatial support has been studied by [Saint-Geours et al. \(2012\)](#) and [Saint-Geours \(2012\)](#) for global sensitivity analyses.

The influence of heterogeneities and upscaling on flow predictions in the context of CO₂ storage was studied by [Bouquet et al. \(2013\)](#). A short term study for 2D model was performed using industrial flow simulator and several upscaling techniques. The main results indicate that the choice of an upscaling technique is less influential for the results dispersion than the spatial variability itself (especially the correlation length of permeability heterogeneities). Even if for 3D model the choice of an upscaling technique could be more influential than for 2D formulation, in the present work the upscaling is considered out of scope, and the effect of heterogeneity is studied at a fixed scale.

5.1.2 Spatial variability and uncertainty related to its representation as a random field

Simulation of equiprobable realizations of the media properties (samples of a random field) allows the analysis of the impact of the spatial variability on the long-term system evolution predictions. A random field expresses at a time the natural stochastic variability and the lack of knowledge that we have on the spatial repartition of the natural medium properties.

Nevertheless, it is important to remember that several sources of errors arise when using such approach for representation of the spatial variability. First of all, as for any sampling technique, there would be a ‘propagation error’ related to the fact that the number of generated samples is not infinite. Indirectly, the propagation error can be appreciated through the convergence rate of the estimators of the quantities of interest.

Secondly, if the assumptions made about the field structure are erroneous or/and the model is constrained to data of insufficient quality and quantity, such a stochastic model would not be representative of the natural spatial variability and would include itself an imprecision. This is an uncertainty of the second order as it concerns the imprecision in the model of the sources of uncertainty. The major challenge is to correctly constrain the stochastic model to the available information accounting for all types of uncertainties. Such an inverse problem in the case of independent identically distributed random variable is described in ([De Rocquigny, 2012](#), page 225).

In the case of spatial variability of permeability the problem is two-fold: first, it is necessary to have enough data that experimentalists claim to be hard or even impossible to obtain (such as the correlation structure); and second, permeability measurements (both on core samples and *in-situ*) are indirect and need an interpretative model (including inversion). The result of the interpretation of a well-test is the apparent value of the permeability in a certain zone around the well. In the proximity of a well the flow is radial. Several authors have proposed analytical formula to find apparent permeability around a well for radial flow conditions, e.g. [Desbarats \(1994\)](#). The existence of effective permeability (or conductivity) for radial flow in two-dimensional bounded domain has been studied by [Noetinger and Haas \(1996\)](#); [Franzetti et al. \(1996\)](#).

These concerns are discussed here to underline the importance of addressing the propagation error and second level uncertainty when using a stochastic field representation for spatial variability. The latter needs the creation of a conceptual framework for quantification of the complex of uncertainties related to random field identification and data assimilation. It remains a scientific topic of actuality which to our knowledge has not been developed enough or numerically tested. For the numerical part the obstacle is evidently the great computational resources needed to properly perform at least one first-level probabilistic study including spatial variability (nevertheless, it can become accessible with the growth of computational powers).

In the present work a random field of permeability with minimal parametrization (inspired from the literature) is chosen to study the influence of spatial variability on the two-phase flow characteristics. The structure of the random field and the way of its sampling are presented hereafter.

5.2 Random tensor field of permeability: modelling and simulation results

This section deals with the modelling of a positive-definite matrix-valued random field of permeability. The field is lognormal, constrained by its mean value, dispersion, correlation structure and anisotropy index. The field generation procedure and results are communicated in (Okhulkova et al., 2014).

5.2.1 Theoretical basis of the generation of anisotropic random field of permeability

The preliminary studies (Section 3.5) have shown that permeability is the most important parameter. Several authors also state that hydraulic conductivity (permeability) is the primary control in fluid displacements in water and gas drives (Rehfeldt et al., 1992; Lake and Carroll, 1986; Lake et al., 1991).

It is worth mentioning that the porosity, assumed as a deterministic quantity in the present chapter, could have been modelled as a random field using the same technique. Still there are two arguments in favor of modeling permeability rather than porosity, which are: permeability changes both the magnitude and direction of flow, whereas porosity only changes the magnitude of the velocity; hydraulic conductivity can vary over 13 orders of magnitude, whereas effective porosity varies in the limits of less than 2 orders of magnitude (Bear, 1979).

For these reasons numerous studies attempt to quantify the influence of permeability heterogeneity on the flow characteristics and to propose a methodology of permeability upscaling from the scale of geological description to the scale of reservoir simulation (Durlafsky et al., 1996). Most of the studies address the problem in two-dimensional configuration for single fluid flow and assume statistical homogeneity or isotropy.

The first consistent attempt in treating key problems of heterogeneous reservoir description dates back to the work of Warren and Price (1961). The authors of the paper affirm that it is possible to obtain the qualitative measures of the degree and the scale of

the heterogeneity and its spatial configuration if core analysis and pressure build-up data are available. Nevertheless, the authors underline the restrictive assumptions that both the core and the build-up information should represent valid measures of the reservoir properties (that is to say, the indirect measurements are correctly interpreted).

The early study where the hydraulic conductivity is modeled as a spatial random field with a prescribed marginal probability distribution and correlation structure was performed by [Hamed et al. \(1996\)](#) for two-dimensional case.

In the present work a 3D field is generated for permeability of anisotropic nature.

Anisotropic nature of permeability

The Darcy's law in its initial form is only valid for isotropic media. Nevertheless, several approaches show the possibility of generalization of the Darcy law for three-dimensional flow in anisotropic media (see [\(Bear, 1972\)](#) section 5.10).

According to generalized form of the Darcy law the hydraulic conductivity K relates the specific discharge $q(q_x, q_y, q_z)$ to the hydraulic gradient $J(J_x, J_y, J_z)$ through the following matrix equation:

$$\begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix} = \begin{bmatrix} K_{xx} & K_{xy} & K_{xz} \\ K_{yx} & K_{yy} & K_{yz} \\ K_{zx} & K_{yz} & K_{zz} \end{bmatrix} \cdot \begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} \quad (5.1)$$

Therefore, in three-dimensional space nine components $K_{ij}(i, j = x, y, z)$ form the hydraulic conductivity tensor, which is a symmetric, positive-definite second-order tensor. Due to symmetry it only contains six different components. The cross-component K_{xy} (as an example) can be interpreted as the weight of the contribution of the component J_y of the hydraulic gradient to the specific discharge in x direction (q_x).

The hydraulic conductivity expresses the capacity of a fluid to be transported through a porous media. As a consequence it includes both properties of the fluid and the medium. The influencing fluid properties are density ρ [kg/m³] and viscosity μ [Pa.s]. The influencing matrix properties (such as shape of the pores and their size distribution tortuosity, porosity) are regrouped into intrinsic permeability k_i [-]. Equation 5.2 gives the relation between the hydraulic conductivity K , effective permeability K_α , intrinsic permeability k_i and relative permeabilities $k_{r\alpha}$ for multi-phase flow in isotropic media:

$$K = \frac{K_\alpha \rho g}{\mu} \quad (5.2a)$$

$$K_\alpha = k_i k_{r\alpha} \quad (5.2b)$$

[Bear \(1987\)](#) states that the concept and definition of relative permeability can not be extended to the general case of anisotropy when x , y and z are not principal directions. Nevertheless, it is possible to define the relative permeability as $\mathbf{K}_\alpha \mathbf{k}_i^{-1}$. Several authors have shown that upscaling can lead to anisotropic relative permeability defined in this way (e.g. [Keilegavlen et al., 2012](#)).

Returning to the flow equations of the numerical model (2.16, 2.17) it can be noticed that the relative permeability is only present in product with the intrinsic permeability.

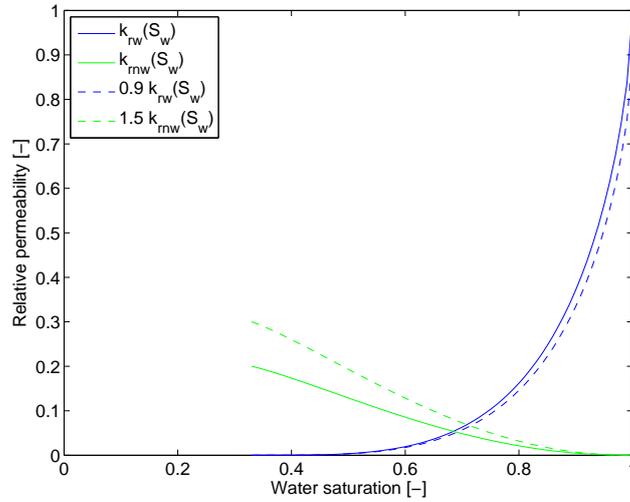


Figure 5.1: Relative permeability curves and their possible perturbation

Therefore, the analysis presented further in Section 5.2.2 can be interpreted as a study for a constant scalar intrinsic permeability and the relative permeability curves weighted by an anisotropic coefficient (this interpretation is only valid for the case of lognormal fields). An example of perturbed curves of relative permeability is given in Figure 5.1. Both the relative permeability endpoint and the entire curve are weighted by a sample of a random coefficient.

In this work the functional dependence of the relative permeability on saturation is assumed to be independent of direction. Therefore, the relations 5.2 for anisotropic media become:

$$\mathbf{K} = \mathbf{K}_\alpha \frac{\rho g}{\mu} \quad (5.3a)$$

$$\mathbf{K}_\alpha = \mathbf{k}_i k_{r\alpha} \quad (5.3b)$$

The simplification made on the relative permeability allows to restrain the anisotropy to the intrinsic permeability (hereafter referred to as permeability).

An important property of a second-rank tensor:

Given the components K_{ij} of a tensor \mathbf{K} in a coordinate system (x, y, z) , the principal directions (x', y', z') can always be found for which the tensor would take a diagonal form.

In light of the property mentioned above the introduction of anisotropy into a random field is equivalent to rotation of the principal axes. Therefore, the way of generating of heterogeneous and anisotropic permeability field proposed in this work consists in defining an orthotropic permeability field and then rotating the principal axes by means of an anisotropy kernel.

Random orthotropic permeability tensor field

The natural deep aquifers targeted to be CO₂ storage sites are geological formations of sedimentary origin. Sediments are commonly deposited in such a way that the permeability of the flow is greater along the planes of deposition than across them. That is why even within a homogeneous facies the permeability should globally be seen as an orthotropic property (eg. if (x,y) is the sedimentation plane and z is the normal direction then $K_x = K_y > K_z$).

Let us discuss how to construct numerically an orthotropic random permeability field. At each point the permeability tensor \mathbf{K}_o has the diagonal form and consists only of horizontal and vertical components (here assumed independent). So that in a point the random orthotropic permeability tensor \mathbf{K}_o is:

$$\mathbf{K}_o(\mu, \sigma) = K_h(\mu, \sigma) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + K_v(\mu, \sigma) \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.4)$$

where K_h and K_v are horizontal and vertical permeability scalar random variables respectively. The parameters μ and σ stand for the mean value and the standard deviation chosen for the horizontal permeability. The parameters for vertical permeability are drawn out of μ and σ by their multiplication by a constant ratio.

After introducing the spatial component with a vector of correlation lengths in the three spatial directions $\boldsymbol{\ell} = (\ell_1, \ell_2, \ell_3)$ the random orthotropic permeability tensor field $\mathbf{K}_o(\mathbf{x}; \mu, \sigma; \boldsymbol{\ell})$ can be written as follows:

$$\mathbf{K}_o(\mathbf{x}; \mu, \sigma; \boldsymbol{\ell}) = \begin{pmatrix} K_h(\mathbf{x}; \mu, \sigma; \boldsymbol{\ell}) & 0 & 0 \\ 0 & K_h(\mathbf{x}; \mu, \sigma; \boldsymbol{\ell}) & 0 \\ 0 & 0 & K_v(\mathbf{x}; \mu, \sigma; \boldsymbol{\ell}) \end{pmatrix} \quad (5.5)$$

where K_h and K_v are independent scalar lognormal random fields. Hence $\log(K_h)$ and $\log(K_v)$ are scalar Gaussian random fields with identical correlation functions that are supposed to be separable:

$$R(\boldsymbol{\eta}; \boldsymbol{\ell}) = \rho(\eta_1; \ell_1) \cdot \rho(\eta_2; \ell_2) \cdot \rho(\eta_3; \ell_3) \quad (5.6)$$

and where $\rho(\eta; \ell)$ is chosen as a squared cardinal sine.

$$\rho(\eta; \ell) = \frac{4\ell^2}{\pi^2\eta^2} \sin^2\left(\frac{\pi\eta}{2\ell}\right) \quad (5.7)$$

More complex correlation structure could be defined provided that additional information is available.

Random anisotropic permeability tensor field

In order to add some random anisotropy the permeability tensor is written as follow:

$$\mathbf{K}(\mu, \sigma, \delta_g) = (\mathbf{K}_o(\mu, \sigma))^{\frac{1}{2}} \times \mathbf{G}(\delta_g) \times (\mathbf{K}_o(\mu, \sigma))^{\frac{1}{2}}, \quad (5.8)$$

where \mathbf{K}_o is the random orthotropic permeability tensor, \mathbf{G} is the anisotropy kernel (Soize, 2005), δ_g is the dispersion parameter of the anisotropy kernel.

The anisotropy kernel \mathbf{G} The anisotropy kernel is a normalized, symmetric, positive-definite real random matrix defined on the probability measure space $(\mathcal{A}, \mathcal{F}, P)$. The derivation of the form of \mathbf{G} is based on the entropy optimization principle (Shannon, 1948; Jaynes, 1957). Using this principle the probability density function of \mathbf{G} is constructed under the constraints of positivity, symmetry, normality and finite Frobenius norm.

In order to be able to perform a Monte Carlo numerical simulation of the random matrix \mathbf{G} it is proposed to write the anisotropy kernel in the following algebraic representation (the expression is completed with the case $(\delta_g = 0)$):

$$\mathbf{G} = \begin{cases} \mathbf{I}, & \delta_g = 0 \\ \mathbf{L}^T(\delta_g)\mathbf{L}(\delta_g), & \delta_g > 0 \end{cases} \quad (5.9)$$

where \mathbf{I} is the identity matrix, \mathbf{L} is an upper triangular matrix with its elements defined as follows:

$$L_{ij}(\delta_g) = \begin{cases} \frac{\delta_g}{2} N_{ij}, & j > i \\ \frac{\delta_g}{2} \sqrt{2h(N_{ij}, a_i)}, & j = i \end{cases} \quad (5.10)$$

N_{ij} - six independent copies of a normalized centered Gaussian random variable, $h(N_{ij}, a_i)$ - a non-linear isoprobabilistic transformation that converts N_{ij} into a Gamma distributed scalar random variable, a_i - parameters of the transformation (shape parameter of the Gamma distribution) computed as (Ta et al., 2010):

$$a_i = \frac{2}{\delta_g^2} - \frac{i-1}{2}. \quad (5.11)$$

$$\mathbf{K}(\mathbf{x}; \mu, \sigma, \delta_g; \boldsymbol{\ell}) = (\mathbf{K}_o(\mathbf{x}; \mu, \sigma; \boldsymbol{\ell}))^{\frac{1}{2}} \times \mathbf{G}(\mathbf{x}; \delta_g; \boldsymbol{\ell}) \times (\mathbf{K}_o(\mathbf{x}; \mu, \sigma; \boldsymbol{\ell}))^{\frac{1}{2}} \quad (5.12)$$

5.2.2 Simulated anisotropic random field of permeability

Conformly to the procedure described in the previous section a set of 100 random tensor 3D fields of permeability have been simulated. The steps and the choices made for the present particular study are the followings:

1. Choose the characteristics of the scalar random fields K_h and K_v :
 - the distribution law: lognormal
 - the correlation function : equations 5.6, 5.7
 - the correlation lengths $(\ell_1, \ell_2, \ell_3) = (50 \text{ m}, 50 \text{ m}, 20 \text{ m})$
 - mean value and standard deviation: 10^{-13} m^2 and $4.9 \cdot 10^{-27} \text{ m}^2$ for K_h ; 10^{-14} m^2 and $4.9 \cdot 10^{-28} \text{ m}^2$ for K_v

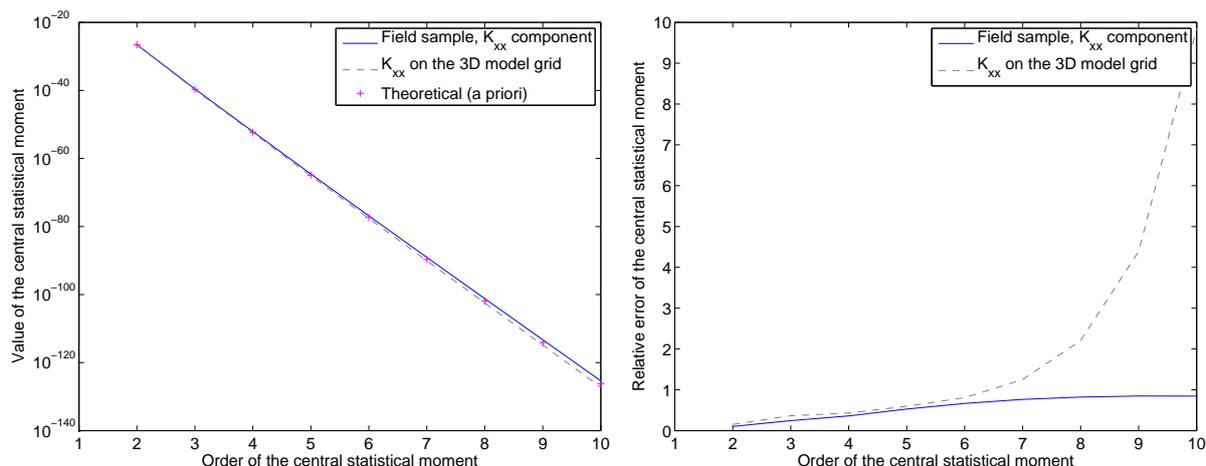


Figure 5.2: Higher statistical moments of the generated field of permeability (K_{xx}) on a regular grid and on the grid of the 3D numerical model of the reservoir compared to the moments of the a priori PDF

- the coordinates of the grid points in which the field should be evaluated: regular grid with a step of 5 m in each direction
2. Generate a realization of K_h , K_v and combine the orthotropic tensor field of permeability $\mathbf{K}_o(\mathbf{x}; \mu, \sigma; \ell)$ as in equation 5.5
 3. Choose the value of δ_g and construct the Cholesky factorization matrix \mathbf{L} of the anisotropic kernel \mathbf{G} in every point:
 - the diagonal elements are Gamma distributed scalar random fields with the same correlation structure as K_h (equations 5.10 and 5.11)
 - the upper extra-diagonal elements are normalized centered Gaussian random variables weighted by $\delta_g/2$
 4. Use the equation 5.12 to obtain the final stochastic permeability tensor field.

To evaluate the quality of the simulated orthotropic field we first verify that the PDF chosen for the scalar random fields K_h and K_v are properly sampled.

For the anisotropic field the pdf does not need to be equal to the input PDF as it is modified by the anisotropy kernel.

Figure 5.2 presents the central statistical moments of higher orders for the anisotropic field in comparison with the input PDF moments.

In order to analyze the simulated field the semivariogram of the permeability field at the distance of mean lateral extent is computed. To do so the values of K_{xx} and K_{zz} components are read on a circle with the radius 1062 m. The computed angular semivariogram of $K_{xx}(\text{angle})$ is shown in the Figure 5.3 (left).

The angle which corresponds to the correlation length at the distance of 1062 m from the well is 2.7° . This angle is quite visible on the semivariogram. The mean semivariogram tends to its sill - the variance value which was used to simulate the random field of permeability ($\text{Var}(K_{xx}) = 2.5 \cdot 10^{-27} \text{ m}^4$). It proves the stationarity and ergodicity of the

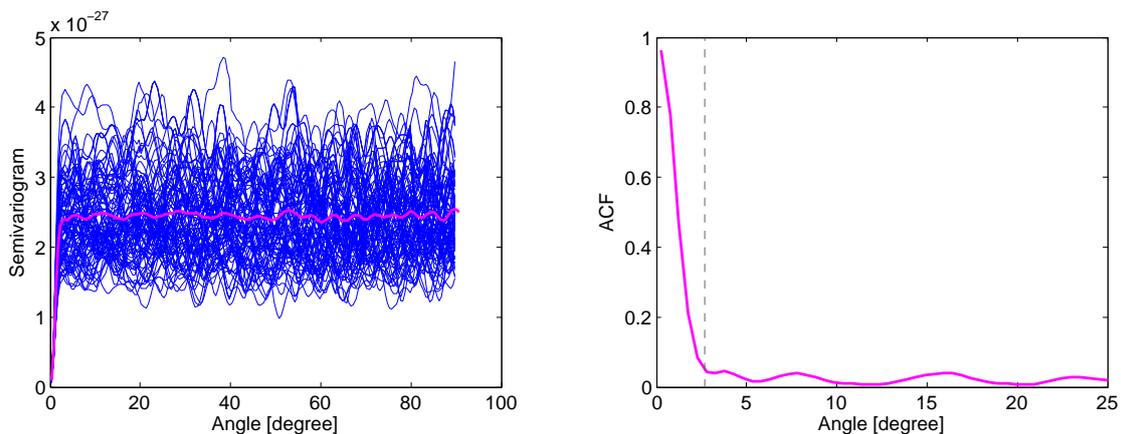


Figure 5.3: Angular correlation of the horizontal permeability at the distance of 1062 m from the injection well. Right: angular semivariograms of 100 K_{xx} realizations (blue lines) and the mean semivariogram (magenta line). Left: angular auto-correlation function corresponding to the mean semivariogram.

simulated field. The mean semivariogram is used to compute the auto-correlation function (Figure 5.3 (right)).

Similar curves are obtained for the vertical permeability. The correlation angle read from the semivariogram of K_{zz} is the same as for the horizontal permeability (2.7°). The semivariogram sill for the vertical permeability also verifies its a priori variance. Thus, the semivariogram is shown to provide correct characteristics for the permeability field.

5.3 Probabilistic study for random tensor field of permeability

Using the coupled 3D-2D axisymmetric model described in section 2.4.2 the flow simulation has been run for 100 samples of the random tensor field of permeability. The mean simulation time on a personal computer (4 CPU Xeon E5630 2.53GHz, RAM 23.4GiB) for a prediction till 10 years reaches 6.5 hours. Each solution contains 146332 degrees of freedom with 15256 elements in the mesh.

The same variables of interest as for the study described in the previous chapter were retained: the maximal lateral spread of the CO_2 after 10 years of injection and the maximal and averaged pressures on the top of the reservoir. As the problem has a radial character, the maximal lateral spread of the injected gas is considered as a function of the angle. An example of an output for a 0.99 contour of water saturation is shown in the Figure 5.4. If a non-convex form is encountered in the front the external edge is taken.

5.3.1 Results for media with heterogeneous orthotropic and anisotropic permeability

Heterogeneous permeability influences the spatial repartition of the injected fluid. In order to characterize the main features we analyze one sample. Intuitively the changes would be

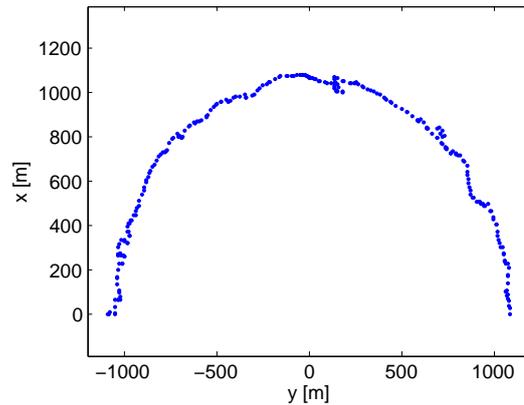


Figure 5.4: The maximal lateral spread of CO_2 after 10 years of injection (view from above on a half-reservoir with the injection well at the position $(0,0)$).

an oscillating (distorted) front behavior, different ratio between the gravity and horizontal pressure gradient effects. Such effects can be observed on a vertical map of saturations presented in Figure 5.6. Of course the difference between homogeneous and heterogeneous cases should not be judged by a single realization. The results with probabilistic dimension are described in the next section (5.4)

What is the front in our numerical model? When the gas flows through the media there is a rapid decrease of water saturation from 1 to 0.5 (as shown on Figures 5.5 and 5.6). At the beginning of the injection the slope of this decrease is a quazi-step function, therefore we could speak about the front in its classical meaning. But with time the slope flattens due to the presence of relative permeability curves and buoyancy.

For the sake of simplicity let us attribute the notion of front in the numerical model to the 0.99 water saturation contour.

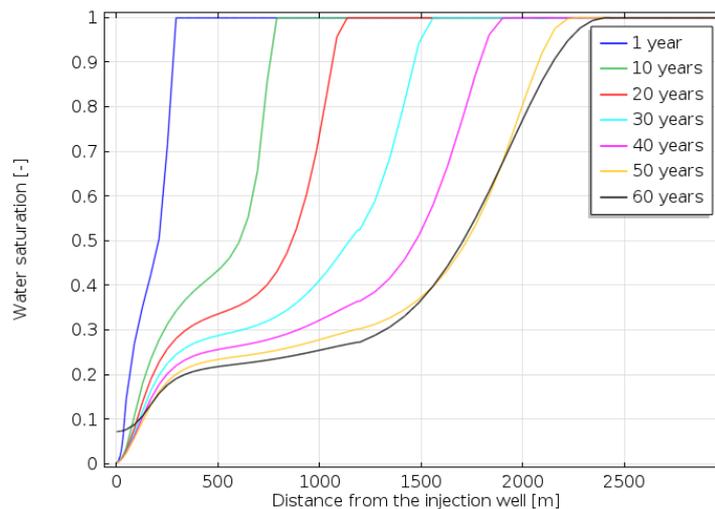


Figure 5.5: Shape of the front ($S_w = 0.99$) in the middle of the reservoir during the injection (till 50 years) and 10 years after injection. Homogeneous permeability 10^{-13} m^2 .

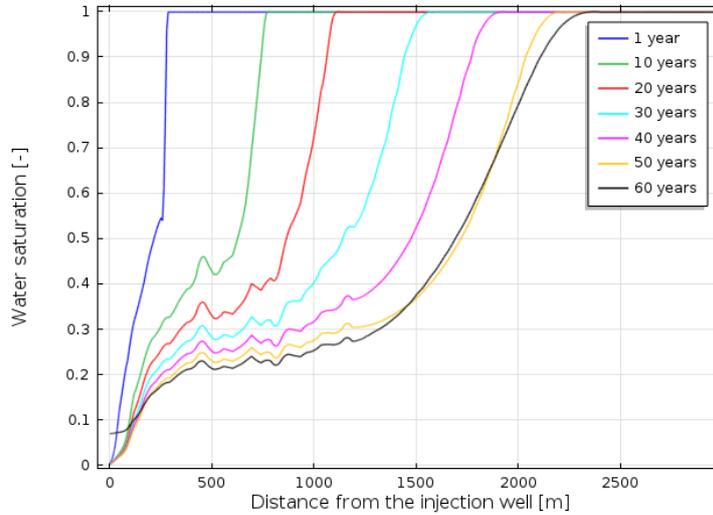


Figure 5.6: Shape of the front ($S_w = 0.99$) in the middle of the reservoir during the injection (till 50 years) and 10 years after injection. Anisotropic permeability field: $\mu = 10^{-13} \text{ m}^2$, $\sigma = 4.9 \cdot 10^{-27} \text{ m}^2$, $\delta_g = 0.2$

Saffman and Taylor (1958) distinguished two regimes of front behavior depending on the mobility ratio difference of present fluids: *stable* regime for which the heterogeneities slightly disorder the front which still appears to be statistically well defined (in this case the apparent front thickness does not increase with time), *unstable* regime characterized by presence of viscous fingering, front thickness grows linearly with time. This definition was proposed for sharp interfaces, when the saturation after the front is maximal without any transition zone. Although in the numerical model under consideration the front is not sharp it would be interesting to notice if the contours of fixed saturation are stable or not.

As the reader can see on Figure 5.7 for high gas saturations the contours are not stable even for a heterogeneous permeability field with low contrast (one and a half times difference in order of magnitude between 0.15% and 99.85% quantiles of permeability).

Remarks

- The front $S_w = 0.99$ is stable with rare locations of further lateral extent.
- The contours of lower water saturations (eg. $S_w \in (0, 0.4]$) can be discontinuous.

5.3.2 Statistics on the maximum radius with variability

In order to quantify in a probabilistic sense the spread of the injected gas in the case of heterogeneous permeability, we develop statistical estimators allowing to obtain such characteristics as the mean value, the standard deviation and the cumulative density function or complementary cumulative density function of the maximal lateral spread r_{\max} . The key point of the proof is that the spatial averaging gives access to the estimates listed above for even small number of 3D simulations.

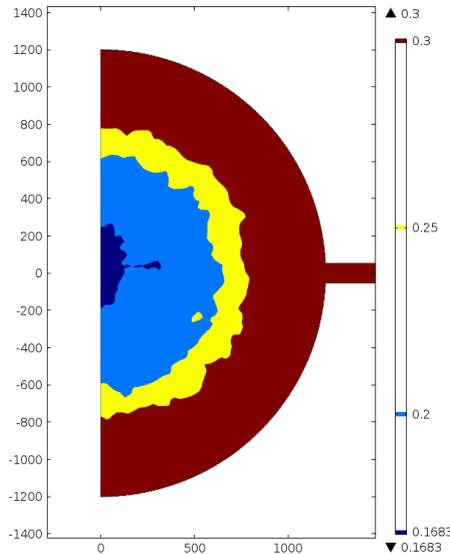


Figure 5.7: Water saturation on the top of the reservoir after 10 years of injection: 0.2 0.25 and 0.3 contours. Both pistons of further local extent and trapped zones of lower gas saturations can be observed

It is assumed that the field of random permeability is statistically invariant for any rotation around the axis of the well^a. As a consequence, the random field $r_{\max}(\theta)$ is statistically invariant with respect to any rotation of the horizontal axes. Hence all its marginal PDFs are invariant with respect to any rotation around the vertical axis. As a consequence, $p(r_{\max})$, the first marginal PDF of $r_{\max}(\theta)$ is independent of angle θ and any statistical moment (including the mean and the standard deviation) are also independent of θ .

Now let us consider the average radius $R_{\max}(\omega)$ as a random variable defined as the spatial average:

$$R_{\max} = \frac{1}{2\pi} \int_0^{2\pi} r_{\max}(\theta) d\theta$$

Its mean (ensemble average) denoted \bar{R}_{\max} is obviously equal to the mean of $r_{\max}(\theta)$. Hence any statistical estimator of $\mathbb{E}(R_{\max})$ is also an estimator of $\mathbb{E}(r_{\max})$ but is likely to converge much more rapidly due to space averaging.

Actually this can be extended to estimators of any quantity derived from the first order marginal PDF of $r_{\max}(\theta)$, including standard deviation or quantiles. Let us denote by (Ω, \mathcal{E}, P) the abstract underlying probabilistic set and define any statistical quantity \bar{q} :

$$\bar{q} = \mathbb{E}(q) = \int_{\Omega} q(r_{\max}(\theta, \omega)) dP(\omega) = \int_0^{+\infty} q(r) p_{r_{\max}}(r) dr$$

^aMeaning that all the marginal PDF are invariant with respect to any rotation of the horizontal axis, the property being satisfied when the permeability field is statistically orthotropic.

with $q(r)$ any function of the maximum radius. Due to rotation-invariance this quantity does not depend on the angle θ at which r_{\max} is observed and $p_{r_{\max}}(r)$ the first marginal pdf of the maximum radius is also independent of the observation angle.

Let us now define the spatial average Q defined for any event $\omega \in \Omega$ by:

$$Q(\omega) = \frac{1}{2\pi} \int_0^{2\pi} q(r_{\max}(\theta, \omega)) d\theta$$

and let \bar{Q} be its mean value which coincides with \bar{q} as:

$$\bar{Q} = \mathbb{E}(Q) = \int_{\Omega} Q(\omega) dP(\omega) = \frac{1}{2\pi} \int_0^{2\pi} \left(\int_{\Omega} q(r_{\max}(\theta, \omega)) dP(\omega) \right) d\theta = \bar{q}$$

We can now define the two Monte-Carlo estimators of these two quantities as:

$$\hat{Q} = \frac{1}{N} \sum_{i=1}^N Q(\omega_i), \quad \hat{q}(\theta) = \frac{1}{N} \sum_{i=1}^N q(r_{\max}(\theta, \omega_i))$$

which will only coincide for very large values of N . However for a fixed N we can easily show that:

$$\hat{Q} = \frac{1}{2\pi} \int_0^{2\pi} \hat{q}(\theta) d\theta$$

However \hat{Q} is a much better estimator as its variance $\sigma_{\hat{Q}}^2$ is much smaller than $\sigma_{\hat{q}}^2$ the variance of \hat{q} . Indeed one can easily show that^b:

$$\frac{\sigma_{\hat{Q}}^2}{\sigma_{\hat{q}}^2} = \frac{1}{2\pi} \int_0^{2\pi} \gamma_q(\theta) d\theta$$

where γ_q is the angular correlation function of the $q(\theta)$ random field. It is worth noticing that this angular-correlation function is only dependent on θ thanks to the rotation-invariance hypothesis. It is also important to notice that the detailed γ_q function is not required in the estimate of the standard deviation of the estimator \hat{Q} but only its sum over $[0, 2\pi]$ which can be defined as the correlation angle θ_{qc} :

$$\theta_{qc} = \int_0^{2\pi} \gamma_q(\theta) d\theta$$

Following this analysis, an estimator of $\sigma_{\hat{Q}}^2$ the square of the standard deviation on \bar{Q} is built using an estimator on $\sigma_{\hat{q}}^2$ and an estimator on θ_{qc} as follows:

$$\widehat{\sigma_{\hat{Q}}^2} = \frac{N}{N-1} \widehat{\sigma_{\hat{q}}^2} \frac{\widehat{\theta_{qc}}}{2\pi}$$

^b

$$\sigma_{\hat{Q}}^2 = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2\pi} \int_0^{2\pi} (\mathbb{E}(\hat{q}(r_{\max}(\theta))\hat{q}(r_{\max}(\theta')))) - (\mathbb{E}(\hat{Q}))^2) d\theta d\theta'$$

and

$$\sigma_q^2 \gamma_q(\theta - \theta') = \mathbb{E}(\hat{q}(r_{\max}(\theta))\hat{q}(r_{\max}(\theta')))) - (\mathbb{E}(\hat{q}))^2$$

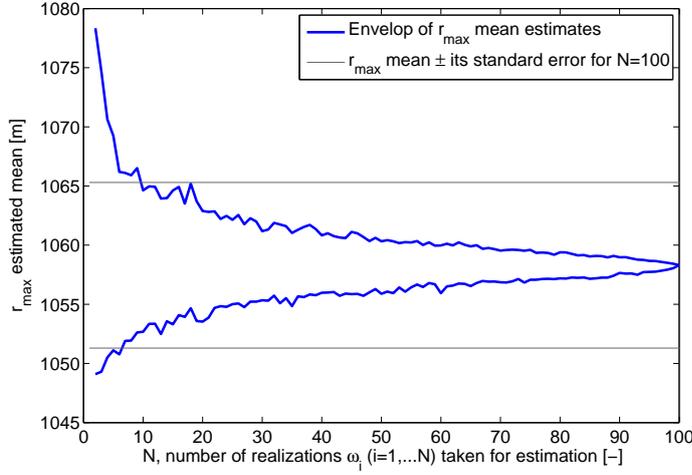


Figure 5.8: Dependence of the estimate of the R_{max} mean with the number N of realizations: minimal and maximal values for 10^3 randomly chosen subsets of N values. The estimated mean and the standard deviation for 100 realizations are 1058 m and 7 m respectively.

where

$$\widehat{\sigma}_q^2 = \frac{N}{N-1} (\widehat{Q}_2 - \widehat{Q}^2)$$

with $q_2(r) = q(r)^2$ and:

$$\widehat{\theta}_{qc} = \frac{1}{2\pi N \widehat{\sigma}_q^2} \sum_{i=1}^N \int_0^{2\pi} \int_0^{2\pi} (q(r_{\max}(\theta, \omega_i))q(r_{\max}(\theta', \omega_i)) - \bar{q}^2) d\theta d\theta'$$

Figure 5.8 illustrates the dependence of the estimator \widehat{Q} (for $q = r_{max}$) mean and standard deviation ($\sqrt{\widehat{\sigma}_Q^2}$) on the number N of realizations ω_i , $i = 1, \dots, N$. Instead of showing the estimate for 1 subset of $N = 1, \dots, 100$ realizations, 10^3 subsets from the totality of $\binom{100}{N}$ combinations are randomly chosen. It gives an approximation of the envelop for all possible combinations. For each subset the mean value is computed. The minimal and maximal mean value of the subsets are represented in Figure 5.8 together with the standard deviation estimate ($\sqrt{\widehat{\sigma}_Q^2}$) and the maximal standard deviation between the 10^3 estimates of the mean value.

When considering the function q in the form $q = (r_{max} - \bar{r}_{max})^2$, the estimator \widehat{Q} gives an approximation of r_{max} variance. Figure 5.9(left) illustrates the dependence of such estimator mean on the number N of samples ω_i , $i = 1, \dots, N$. The Figure 5.9(right) represents the square root of the variance envelop, giving envelop for the standard deviation, the standard error is evaluated as described hereafter.

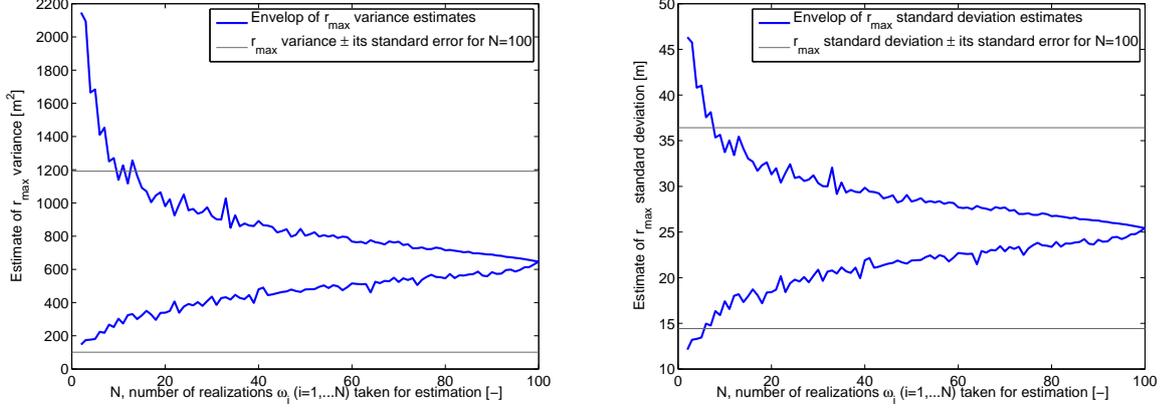


Figure 5.9: Dependence of the estimate of the r_{max} variance (left) and standard deviation (right) with the number N of realizations: envelop for 10^3 randomly chosen subsets of N values. The values of the estimator and its standard error for 100 samples for variance are 646 m^2 and 545 m^2 (for standard deviation 25 m and 11 m respectively).

Knowing the estimated variance (646 m^2) the standard deviation is computed resulting in 25 m . Now let us find an approximation of the standard error of the standard deviation knowing the standard error of the variance. The standard error of the variance is commonly found with the help of chi-squared distribution as it represents a sum of squares of Gaussian standard distributions. In this case the uniformly minimum variance unbiased estimator S^2 of the variance has a standard error (Lehmann and Casella, 1998):

$$\sigma_{S^2} = \sigma^2 \sqrt{\frac{2}{(n-1)}},$$

where σ^2 is the estimated variance, $(n-1)$ is the number of degrees of freedom of the chi-squared distribution. As we already have a standard deviation of the variance estimator ($\sqrt{\sigma_Q^2} = 545 \text{ m}^2$), we can find a degree of freedom of a chi-squared estimator which would give the same standard error. We find $n = 1 + 2 \frac{\sigma^2}{\sigma_{S^2}^2}$ equals to 4. The uniformly minimum variance unbiased estimator of the standard deviation is $K_n S$ (Lehmann and Casella, 1998, p.92), where $K_n = \sqrt{\frac{n-1}{2} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}}$. The standard error of $K_n S$ is:

$$\sigma_{K_n S} = \sigma K_n \sqrt{\frac{V_n}{n-1}},$$

V_n being the variance of chi distribution with $(n-1)$ degrees of freedom: $V_n = 2 \left(\frac{n-1}{2} - \frac{\Gamma^2(\frac{n-1}{2})}{\Gamma^2(\frac{n}{2})} \right)$. For $n = 4$ the factor $K_n \sqrt{V_n}$ takes the value of 0.73. Therefore, the standard error $\sigma_{K_n S}$ can be estimated to 11 m .

The final goal being the evaluation of the CCDF of the variable of interest, let us now

define an estimator of the CDF of the maximum radius. To do so, the function q can be expressed as an indicator function h which equals to 0 for radius exceeding r_0 :

$$h(\omega, r_0 - r_{max}(\theta)) = \begin{cases} 1, & r_{max} \leq r_0 \\ 0, & r_{max} > r_0 \end{cases} \quad (5.13)$$

The statistical quantity $\bar{h}(r_0)$ representing the ensemble average of the function $h(r_0)$ would give the cumulative probability at radius r_0 :

$$\bar{h}(r_0) = \mathbb{E}(h(\omega, r_0 - r_{max})) = \int_{\Omega} h(\omega, r_0 - r_{max}) dP(\omega) = \int_0^{+\infty} h(r_0 - r) p_{r_{max}}(r) dr = P(r_{max} < r_0)$$

Introducing the spatial average $H(\omega, r_0)$ allows an estimate of the cumulative probability in the following way:

$$P(r_{max} < r_0) = \bar{H}(r_0) \approx \hat{H}(r_0) = \frac{1}{N} \sum_{i=1}^N H(\omega_i, r_0)$$

In order to get an idea of the convergence rate of $P(r_{max} < r_0)$ with the number N of samples taken for evaluation, the curves for the quantile 80% ($r_0 = 1071$ m) are shown in Figure 5.10. The representation is similar to Figure 5.8: the envelop of results for 10^3 subsets is shown. In the case of homogeneous permeability the exceedence probabilities for $r_0 = 1500$ m and $r_0 = 2000$ m have been evaluated. In the case of heterogeneous permeability no sample reaches the radius of 1300 m. That is why the radius $r_0 = 1071$ m is chosen corresponding to the quantile 80%.

Computing $P(r_{max} < r_0)$ for each r_0 gives the whole CDF curve of the maximal lateral spread r_{max} . The resulting complementary cumulative distribution function ($1 - \hat{H}(r_0)$) is shown in Figure 5.11 together with the $N=100$ samples of $1 - H(\omega_i, r_0)$ ^c.

The asymptotic standard deviation of the estimator $\hat{H}(r_0)$ is expressed as (De Rocquigny, 2012, p.290):

$$\sigma_{\hat{H}} = \sqrt{\frac{1}{N} \bar{H}(r_0)(1 - \bar{H}(r_0))} \approx \sqrt{\frac{1}{N} \hat{H}(r_0)(1 - \hat{H}(r_0))}.$$

Several curves $1 - H(\omega_i, r_0)$ (Figure 5.11) are outside of the 95% asymptotic confidence interval, it means that the values are far from being distributed according to a Gaussian distribution. At both extremities certain curves reach the Chebyshev 95% bound.

In Figure 5.12 the resulting complementary cumulative distribution function ($1 - \hat{H}(r_0)$) is compared to a lognormal CCDF having the same first two statistical moments. At logarithmic scale a lognormal CCDF behaves as a logarithm of the complementary error function with an argument $(\ln(x) - \mu)/\sqrt{2}\sigma$:

^cIt is important to note that the variability in Figure 5.11 illustrates the computational procedure with no relevance of data epistemic uncertainty unlike Figures 3.15 and 3.16 in Section 3.5.3. The curves $1 - H(\omega_i, r_0)$ are smoothed by considering the front as a smoothed Heaviside in the form of Gaussian CDF with a standard deviation of 5 m.

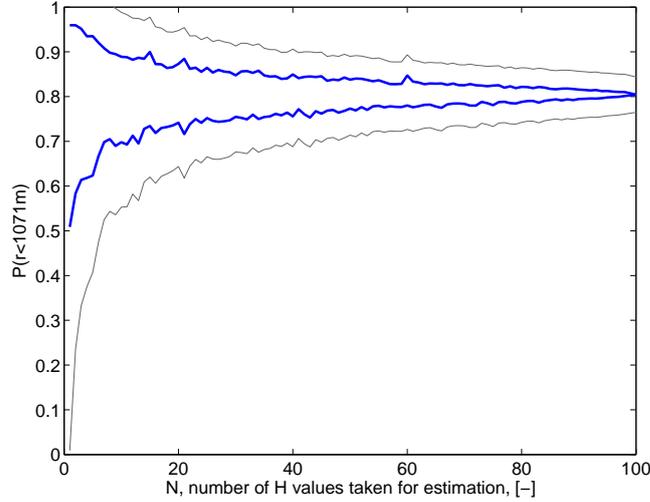


Figure 5.10: Estimation of $P(r_{max} < 1071 \text{ m})$: minimal and maximal values for 10^3 randomly chosen subsets of N samples of H . The error bound corresponds to the asymptotic standard deviation computed as $\sqrt{P(1-P)/N}$. For 100 samples the probability is estimated to 0.8 with a standard deviation of 0.04.

$$f = \ln\left(\frac{1}{\sqrt{\pi}} \int_{-\frac{\ln(x)-\mu}{\sqrt{2}\sigma}}^{+\infty} e^{-t^2} dt\right)$$

As it can be noticed the lognormal CCDF with the same mean and variance underestimates the probabilities higher than 0.5 and lower than $7 \cdot 10^{-2}$. The main difference of the estimated CCDF with respect to a lognormal CCDF is the change of slope occurring for probabilities less than $9 \cdot 10^{-2}$ (or for radius over 1090 m). A possible explanation of this change is the physical phenomenon of faster advancement of the parts of the front where the fingering has started due to permeability heterogeneity. Nevertheless, to prove the physical origin of this observation it should be demonstrated that the low probability estimates are based on sufficient sampling. Another reason could be the proximity of the transition zone from 3D to 2D rotation invariant formulation which is situated at 1200 m from the injection well. Therefore, further analysis is needed to confirm the origin of the slope change in the empirically found CCDF.

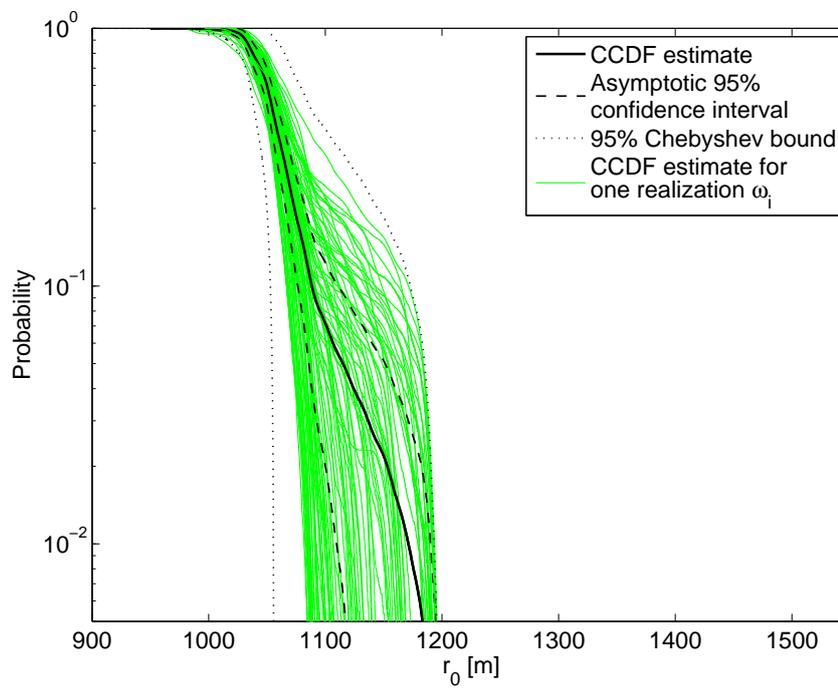


Figure 5.11: Estimate of the CCDF of the maximal lateral spread (in black) for the case of heterogeneous permeability, the curves $1-H(\omega_i, r_0)$ for each model output^c: 100 samples (in green), the theoretical 95% confidence interval is computed as $\sqrt{P(1-P)/100}$ and Chebyshev 95% bound is computed as $\pm k\sigma$, where $k = 1/\sqrt{1-0.95}$ and σ is the standard deviation of the values $1-H(\omega_i, r_0)$ for each given r_0 .

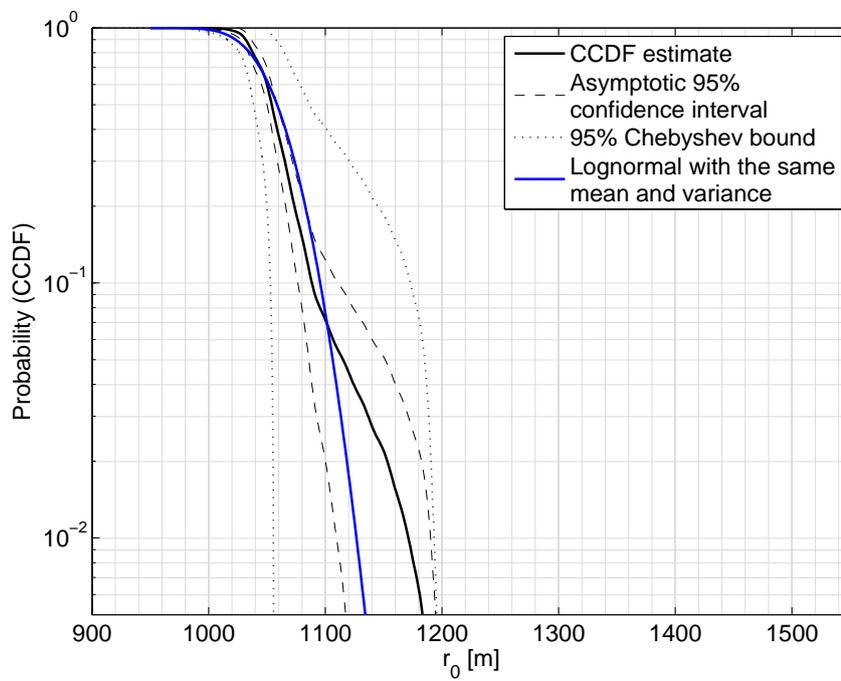


Figure 5.12: Estimate of the CCDF of the maximal lateral spread with logarithmic scale and the asymptotic confidence interval 95%.

5.4 Permeability uncertainty: homogeneous versus heterogeneous variation

Often for parametric uncertainty propagation studies the input variables are considered homogeneous for the sake of simplicity. In this section we study the influence of the choice to represent the inputs as homogeneous or heterogeneous random quantities on the resulting uncertainty of the variable of interest. With that end in view, the impacts of the homogeneous variation and the spatial variation of permeability for the same input probability density function are compared. In the second part an equivalent homogeneous variation is inferred reproducing heterogeneous results.

First, let us discuss the relation of one of these media descriptions with respect to potential available information. The question of how to represent the uncertainty of the underground medium properties depends a lot on the quantity and scale of data available for its description. What do we know about the permeability?

- (a) We can have a set of local measurements (on the laboratory scale: maximal size of samples being less than 1m) corresponding to the description of samples from a well. We can acquire some statistics on this set of measurements. So we know a spatial mean and standard deviation of laboratory scale permeabilities (for a given geological formation).
- (b) Or we can have a set of field-scale measurements for different fields [L.W.Gelhar 1992]. So for a given field we know one or luckily several values of the field-scale permeability (or hydraulic conductivity or transmissivity). Such a test requires a lot of time and its precision is probably lower than laboratory measurements.

For data of type ‘a’ we can construct a probabilistic model of the heterogeneous permeability field. The key hypothesis to make in this case is that the number of samples taken in different locations is sufficient to represent the marginal distribution. The variability of the response should be interpreted as due to the unknown spatial repartition of the media properties between the points of observation. To introduce the second level uncertainty we should consider how far the experimental marginal distribution is from the real one. But there is no data which could provide quantitative base for answering this question. To summarize the advantages and disadvantages of introducing the uncertainty on the underground medium properties in the form of random field constrained to data of type ‘a’ (local small-scale measurements in wells):

- + Possibility to introduce spatial variability.
- + Possibility to predict the large scale behavior of the system taking into account the spatial variability.
- The marginal distribution is not fully known.
- The spatial correlations determine the response variability, but the correlation length is poorly known as well; subjective choice of the correlation function.
- The global response variability can be under- or over-estimated because of the non-consideration of the second level uncertainty (the error in the marginal distribution and the correlation structure)

For data of type ‘b’ the uncertainty arises from the imprecision of the test and its quantitative interpretation model. To evaluate the impact of this uncertainty on the response variability by attributing to the homogeneous permeability a probability density function centered on the measured value and with standard deviation corresponding to the measurement error.

Taking the mean value of the measurements of type ‘a’ and attributing it to the whole domain would correspond to the procedure of ‘averaging’ (Wood, 2009). This value is not equal to the equivalent permeability. The measurement of type ‘b’ gives directly the equivalent permeability. What we could mean by ‘homogeneous case’ is that the permeability has been homogenized or averaged up to macro scale of the problem. So we are actually speaking about ‘equivalent’ or ‘apparent’ permeability.

A precaution should be taken when comparing the response variability resulting from the two probabilistic studies described above as they account for different kinds of uncertainty (unknown spatial repartition and imprecision in measurements).

5.4.1 Comparison of the results

The results obtained in Section 4.3 for homogeneously varying permeability and the results for the permeability spatial variability discussed in Section 5.3.2 are summarized in Table 5.1. Analyzing these results we can notice a slight retention of the front in the heterogeneous case: the mean spread is 50 m less than for homogeneous variation. The standard deviation is reduced by more than 4 times in presence of heterogeneities. It can be explained by the spatial averaging. Both studies give a possibility to evaluate the quantiles and even their confidence intervals. The Chebyshev conservative bounds show the intervals in which the quantiles would lie if no information on the character of the input random variable were available. The estimate that we obtain are more precise. In the homogeneous case the quantiles have higher values in accordance with higher mean value and standard deviation. The standard error of the estimates is rather small for the output mean values (less than 1%), for the output standard deviation the bounds are larger (7% in homogeneous case and over 40% in heterogeneous case).

Comparing the results for pressure gives a similar perspective. We observe a significant reduction of the variance when the uncertainty is represented through spatial variability. The coefficient of variation changes by two orders: from 40% to only 0.4%. Such result could be expected as a consequence of the definition of the output variable in the form of a spatial average. The last two sections of the comparative table refer to the maximal fluid overpressure on the top of the reservoir and in the entire reservoir during the whole considered time period. The maximal overpressure on the top of the reservoir does not seem to be influenced by local effects (the mean value in the homogeneous and heterogeneous cases coincide). The maximal overpressure in the entire reservoir corresponding to the proximity of the injection zone is subjected to the local changes in pressure in the heterogeneous case, which can be demonstrated by the reduction of the estimated mean value from 7.2 to 6.8 MPa. The coefficient of variation for this variable of interest reduces by two orders.

Table 5.1: Comparison of the results for probabilistic studies on homogeneously and spatially varying permeability (with a lognormal PDF $\mu=-30.05$, $\sigma=0.47$)

		Homogeneous case	Heterogeneous field	
Maximal lateral spread	Mean	1118 m \pm 11 ^d m	1058 m \pm 7 m	
	Std	113 m \pm 8 m	25 m \pm 11 m ^e	
	Median	1097 m \pm 14 m	1059 m \pm 3	
	Quantiles	90%	1268 m \pm 19 m ^f	1096 m \pm 4 m
		95%	1330 m \pm 24 m	1107 m \pm 5 m
		99%	1470 m \pm 43 m	1127 m \pm 10 m
	Chebyshev intervals	90%	760-1475 m	979-1139 m
		95%	613-1623 m	945-1173 m
		99%	0-2248 m	805-1313 m
	Average overpressure (top)	Mean	$8.4 \cdot 10^5$ Pa	$11.4 \cdot 10^5$ Pa
CV		40%	0.4%	
Maximal overpressure (top)	Mean	$1.8 \cdot 10^6$ Pa	$1.8 \cdot 10^6$ Pa	
	CV	36%	2.1%	
Maximal overpressure (reservoir)	Mean	$7.2 \cdot 10^6$ Pa	$6.8 \cdot 10^6$ Pa	
	CV	40%	0.5%	

The resulting CCDFs for the maximal lateral extent and the averaged overpressure on the top of the reservoir are presented in Figures 5.13 and 5.14 respectively. These figures illustrate the effect of reduced variance when the uncertainty is attributed to the spatial variability compared to the case of homogeneous permeability.

^dThe standard errors reported here are computed for 100 Monte Carlo samples before the treatment by a metamodel, they correspond to the asymptotic standard deviation such as presented in Table 3.8.

^eThe standard error for the standard deviation is estimated from the standard error of variance through an equivalent χ^2 estimator (details in Section 5.3.2).

^fThe standard error of quantile estimations is computed as (Kendall and Stuart, 1977): $\frac{\hat{\sigma}\sqrt{P(1-P)}}{\sqrt{N}U(z(P))}$, where $\hat{\sigma}$ is the standard deviation estimate, P is the percentile of the quantile to estimate, N is the number of samples, $U(z(P))$ is the standard normal density of P^{th} quantile of the standard normal distribution.

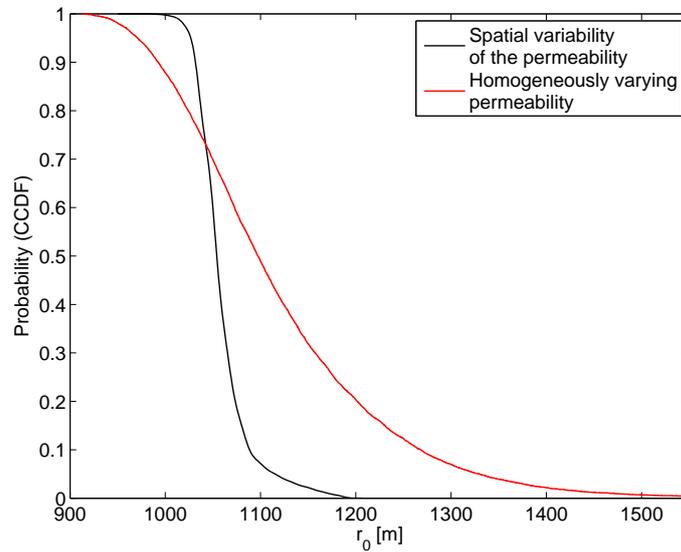


Figure 5.13: Comparison of the empirical CCDFs of the maximal lateral extent of the CO₂ cloud for 2 cases: homogeneous and heterogeneous variation of permeability based on the same PDF (lognormal $\mu=-30.05$, $\sigma=0.47$).

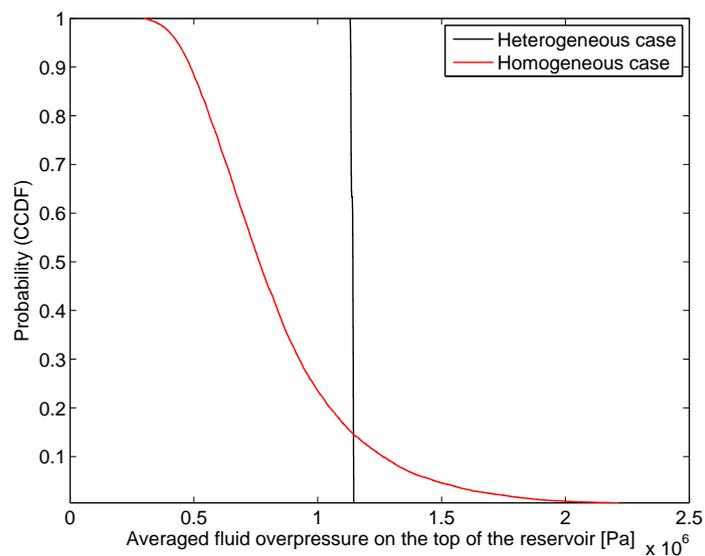


Figure 5.14: Comparison of the empirical CCDFs of the averaged fluid overpressure for 2 cases: homogeneous and heterogeneous variation of permeability based on the same PDF (lognormal $\mu=-30.05$, $\sigma=0.47$).

5.4.2 Inference of the homogeneous equivalent permeability distribution

Now let us find what should be the variability in the homogeneous permeability so that after its propagation through a homogeneous model we could obtain the same uncertainty on the maximal lateral spread as in the heterogeneous case. In other words, the challenge here will be to find an equivalent homogeneous permeability PDF whose propagation through 2D rotation-invariant model would give the same CCDF as for the heterogeneous permeability case (such as shown in Figure 5.11).

To do so, let us notice that the sampling of the meta-model (see Figure 4.7) represents a monotonically strictly increasing (and therefore invertible) function, mapping the variable of interest to the permeability for a sufficiently large range of values. Therefore, having a set of values of r_{max} we can compute the corresponding set of values of permeability through interpolation between the points of the meta-model sampling (no matter the probabilistic density). Hence, a probabilistic inversion of the CDF for the heterogeneous case (that we would like to reproduce) would give a random set of maximal radius values that can be converted into a random set of permeability values through the interpolated meta-model (the reader is invited to refer to Appendix F for details). The obtained random set of permeabilities would give an approximate characteristics of the permeability variability to put into homogeneous uncertainty propagation.

To summarize, the main steps for the inference of the homogeneous equivalent permeability distribution are:

- Determination of the CDF of the variable of interest for the heterogeneous case to be reproduced (the empirical CDF or its fit/extrapolation by a known law - both are considered here)
- The inverse transform sampling of the determined CDF. It gives a random set of the maximal radius values which approximates the CDF in question (details in Appendix F). This step is needed because the CDF is deduced through an estimator different from a simple sampling, therefore no representative random set of maximal radius values is available.
- The computation of the random set of permeabilities corresponding to the random set of the maximal radius from the previous step. Each value of the maximal radius is recalculated into a permeability value through a linear interpolation of the meta-model (Appendix F).
- Finding characteristics of a lognormal distribution representing the mean value and the variance of the random set of permeabilities.

The obtained equivalent homogeneous permeability PDF is compared to the initial permeability PDF in Figure 5.15. In order to reproduce the heterogeneous permeability results the log standard deviation should be divided by 3 (its value equals to 0.47 for the initial PDF and 0.15 for the retrofitted distribution) and actual standard deviation should be divided by 4.5 (the respective values are $5 \cdot 10^{-14} \text{ m}^2$ and $1.1 \cdot 10^{-14} \text{ m}^2$). The mean value shifts slightly to the left from 10^{-13} m^2 to $7.5 \cdot 10^{-14} \text{ m}^2$. The conspicuous fact is

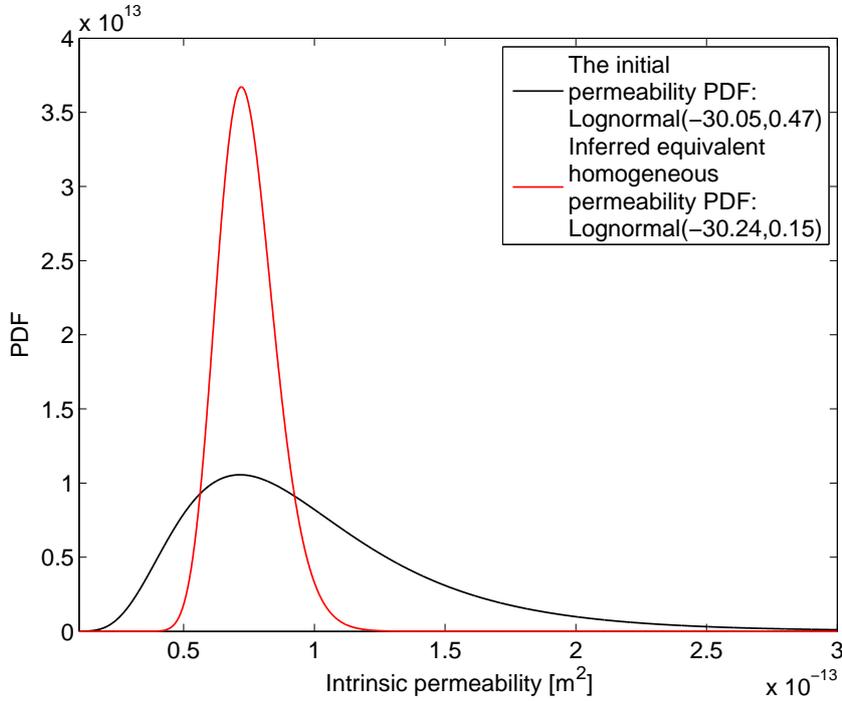


Figure 5.15: The permeability probability density functions: the initially considered PDF: lognormal with $\mu=-30.05$, $\sigma=0.47$ and inferred homogeneous equivalent permeability PDF: lognormal with $\mu=-30.24$, $\sigma=0.15$. The modes of both PDFs are equal to $7.2 \cdot 10^{-14} \text{ m}^2$

that the mode value remains invariable (the modes of both PDFs are equal to $7.2 \cdot 10^{-14} \text{ m}^2$).

After the propagation of the deduced equivalent homogeneous permeability distribution through PCE metamodel for 2D rotation-invariant numerical model (such as described in section 4.3), we obtain a retrofitted CCDF curve for the maximal lateral extent of the CO_2 cloud.

Following the same procedure not for the empirical CCDF of the heterogeneous case but for its lognormal fit reproducing the two central moments, we obtain the second retrofitted CCDF (see Figure 5.16). Performing a two-sample Kolmogorov-Smirnov test leads to a conclusion that both retrofitted curves follow the same distribution for a significance level of 10^{-3} , the maximal difference between the curves being 10^{-2} . The test is performed directly on the output 10^4 metamodel samples.

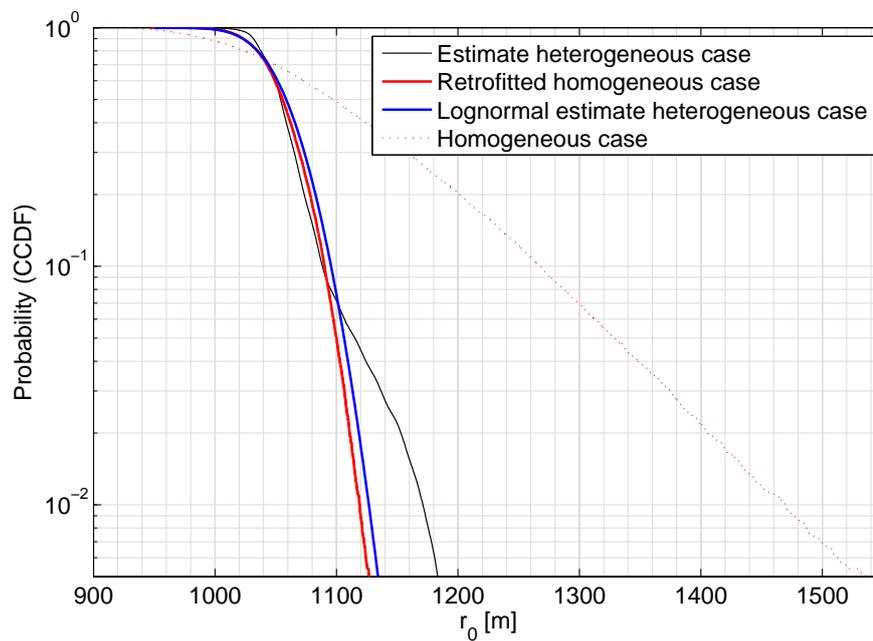


Figure 5.16: Comparison of the empirical CCDFs of the maximal lateral extent of the CO_2 cloud. The estimated CCDF for heterogeneous case (the black line, permeability follows a lognormal PDF with $\mu=-30.05$, $\sigma=0.47$), its lognormal fit with the same mean and variance (the blue line); and the inferred equivalent homogeneous case (the red line, the characteristics of input lognormal PDF found by retrofitting are $\mu=-30.24$, $\sigma=0.152$). The initial homogeneous case CCDF is reminded in red dotted line.

Table 5.2: Comparison of the results for probabilistic studies on heterogeneous permeability (with a lognormal PDF $\mu=-30.05$, $\sigma=0.47$, corresponding to the mean value and standard deviation of $1 \cdot 10^{-13} \text{ m}^2$ and $5 \cdot 10^{-14} \text{ m}^2$) and homogeneous permeability (with lognormal PDF $\mu=-30.24$, $\sigma=0.148$ corresponding to the mean value and standard deviation of $7.5 \cdot 10^{-14} \text{ m}^2$ and $1.1 \cdot 10^{-14} \text{ m}^2$)

		Retrofitted homogeneous case	Heterogeneous field
Maximal lateral spread	Mean	1056 m \pm 3 m	1059 m \pm 7 m
	Std	26 m \pm 2 m	25 m \pm 11 m
	Median	1056 m \pm 3 m	1059 m \pm 3 m
	Quantiles 90%	1090 m \pm 4 m	1096 m \pm 4 m
	95%	1100 m \pm 5 m	1107 m \pm 5 m
	99%	1119 m \pm 10 m	1127 m \pm 10 m
	Chebyshev 90%	973-1138 m	979-1139 m
	intervals 95%	940-1172 m	945-1173 m
	99%	796-1316 m	805-1313 m

5.5 Concluding remarks

The present study shows a tremendous variance reduction of both chosen variables of interest when the input uncertainty in intrinsic permeability is attributed to the spatial variability. Compared to an equivalent study where the same PDF is attributed to homogeneous permeability, the variance of the maximal lateral extent is reduced by factor more than 20 and for the averaged fluid overpressure by 75. The mean value of the maximal lateral extent of the CO_2 in the heterogeneous case is smaller by 5.4% compared to the homogeneous variation. So a slight retention of the front in presence of heterogeneities is observed. It is inferred that in order to reproduce the heterogeneous permeability results in terms of the whole CCDF, the log standard deviation should be divided by 3 (its value equals to 0.47 for the initial PDF and 0.15 for the retrofitted distribution) and actual standard deviation should be devised by 4.5. It is remarkable that the value of the permeability mode does not change.

Conclusions and perspectives

The geological storage of CO₂ in deep aquifers is a promising technology for reducing the CO₂ concentration in the air. Nevertheless, in most current projects even short term predictions of spatial distribution of CO₂ during injection differ from the monitoring results. It can be related to the lack of knowledge of the underground medium properties such as heterogeneity characteristics. Therefore, risk analysis based on a proper uncertainty treatment should be put in the first place in the decision making process concerning CO₂ storage viability.

The results of risk analysis for CO₂ storage in deep aquifers are site specific. The main reasons are the particular for each site configuration of vulnerable elements, determining the ‘Exposure’ component of risk, and breaches of integrity of the recovering geological layers (conductive faults or wells), determining the leakage intensity. Therefore, no general conclusion can be made on the feasibility or performance of the geological CO₂ storage.

Nevertheless, for the non-altered evolution (with no external events or breaches of caprock integrity) the intensity of the underground flows during and after the CO₂ injection can be estimated by a predictive model adjustable for different particular sites, by a proper choice of representative input parameter descriptions and geometry. For rapid screening and ranking of scenarios a simplified analytical model such as proposed by [Nordbotten and Celia \(2006\)](#) can be used. However, a numerical model is needed for more detailed and precise flow estimation.

In the present work, the Dogger aquifer of the Paris sedimentary basin was chosen to represent a potential CO₂ storage site. The analytical model is out of validity limits for the characteristics of the Dogger aquifer, as the gravity component is not negligible due to the high aquifer thickness (120 m). Therefore, several numerical models predicting the flow of the CO₂ and the brine during and after the injection have been developed. The numerical 2D rotation-invariant model created at INERIS prior to this work has been reduced to a closed system of differential equations and optimized to be included in probabilistic studies. A model including a full 3D zone coupled with 2D rotation-invariant formulation far from the well has been elaborated in order to study the effects of heterogeneities on the flow.

The subject of the present work stands at the junction of such disciplines as the numerical modeling of physical phenomena, the risk analysis and the uncertainty treatment (including the spatial variability). The important part of the work consists in uniting these branches in the context of CO₂ storage. The risk analysis part is the most general term including the others, that is why it is discussed first.

Most risk analysis techniques are adapted to binary events (of something happening or not happening) more than processes. The use of events is quite convenient for distinction

of possible scenarios and their visualization in the form of Event trees or Fault trees. When it comes to quantification of slow underground processes, an event actually can be represented as a condition on the state of the system. The intensity and probability of each event is related to the intensities and probabilities of other events in the global context of system evolution. Therefore, the classical Bayesian approach of computing the probabilities through an Event or Fault tree becomes impossible due to unknown conditional probabilities. Furthermore, a new possibility arises to express the probability of events through the probability of corresponding substates of the system given the input uncertainty.

Pursuing the goal of finding the most adapted CO₂ storage risk assessment methodology, which includes and is based on uncertainty propagation through predictive models, we suggest that it should be oriented on processes and system states rather than events because of the particularities of CO₂ storage highlighted in Chapter 3. In this sense, the expertise in the domain of chronic risk predictions (such as water or soil contamination) can provide useful tools and ideas. First, the uncertainty sources should be thoroughly quantified as this step would correspond to the cause identification and description. Then, the most complete predictive model should be used to propagate the input uncertainty. Beforehand, the criteria of undesirable system states should be defined. Among these criteria, the acceptable ones would represent conditions or filters on the predictive model intermediate outputs, the non-acceptable ones would stop the simulation registering the input parameters and encountered conditions which lead to such system state. In both cases the quantitative estimates of the intensity and the probability of each dangerous phenomenon would be accessible without neglecting their relation with the intensity and the probability of other phenomena. The probable inconvenience of such an approach is that it would be less obvious to couple its results with other more technical (accidental type) risks for the steps of capturing, transport and injection.

The literature review of risk assessment methodologies applied so far to the CO₂ storage (the reader can refer to Appendix A for concise descriptions) shows that certain tools such as CO2PENS or ESL allow evaluating parts of a Risk model using the approach based on system modeling and uncertainty propagation described above. The present work makes the link between risk quantification through uncertainty propagation for a system model and risk analysis (in the form of events regrouped in scenarios) by formulating the quantities of interest in the form of probabilities of threshold exceedence. Three levels of uncertainty treatment are described with corresponding quantities of interest and critical thresholds: deterministic, probabilistic and double probabilistic. The key notions, one-input example and literature references were given for critical thresholds computation corresponding to each level.

The scenario of the lateral leakage out of the storage complex was quantified through the system modeling approach and the CCDF of the maximal lateral spread of a CO₂ cloud after ten years of injection was obtained. An intermediate step for the quantification of the scenario of caprock fracturing provided probability of exceeding a critical fluid overpressure on the top of the aquifer.

The system modeling approach opens the door to the application of different techniques of parameter uncertainty propagation within risk assessment procedure. In order to explore the particularities of most commonly used techniques of uncertainty propagation and

sensitivity analysis, a simplified physical model with an analytical solution (Nordbotten and Celia, 2006) has been used. The comparison of the results shows that the first order Taylor approximation gives the fastest and quite precise appreciation of central moments of the variable of interest as well as the importance indexes for each input variable. When estimating variance and exceedence probability, a stochastic metamodel gave better estimates than the classical Monte Carlo simulation, demanding 1000 times less model runs. With the number of model runs required for converged estimates through the standard Monte Carlo simulation, a stochastic metamodel coupled with Taylor approximation for the epistemic uncertainty can give access to the second level probabilistic risk measures. It highlights the utility of using stochastic metamodels for slow flow problems.

Given the results of the preliminary comparison of uncertainty propagation techniques on the simplified physical model, a stochastic metamodel obtained by means of polynomial chaos expansion (PCE) was applied to the 2D rotation-invariant numerical model. Based on numerical experiments, subject to the limitations implied by the assumptions that were made, the following conclusions were inferred: the mean lateral spread of the CO₂ cloud after 10 years of injection is 1.1 km (bilateral 95% confidence interval: 0.9-1.5 km). The probability that the CO₂ cloud exceeds 1.5 km after 10 years of injection is of order of 10⁻², the probability of exceeding 2 km is of order of 10⁻⁴.

Among the input variables, the influence on the maximal lateral spread is principally distributed among five input variables starting from the intrinsic permeability, maximal gas saturation and thickness of the reservoir and finishing by the gas relative permeability endpoint and injection rate. The surprising result is the irrelevance of the total porosity and the thickness of the damaged zone compared to other inputs.

In terms of the average fluid overpressure on the top of the reservoir, the mean value is estimated to 0.6 MPa (bilateral 95% confidence interval: 0.3-1.2 MPa). The sensitivity analysis for the average overpressure on the top of the reservoir suggests that the 94% of influence is controlled by the intrinsic permeability and the remaining 6% can be attributed to the injection rate.

The maximal fluid overpressure on the top of the reservoir is observed straight above the injection zone. The obtained mean overpressure above the injection zone is 1.6 MPa, which is higher than the supposed critical mechanical resistance of the caprock. The corresponding bilateral 95% confidence interval is 0.7-3.4 MPa. The mean overpressure in the middle of the injection zone (which corresponds to the zone of maximal fluid overpressure inside the reservoir) is estimated to 4.9 MPa (bilateral 95% confidence interval: 1.8-9.6 MPa). These results indicate that the injection rate of 1 Mt/year is too high for the chosen characteristics of the aquifer and fracturing of the matrix can become a potential problem decreasing the containment capacity of the complex and favoring the induced seismicity. Therefore, for similar aquifers it could be preferable to reduce the injection debit and to control the injection well by pressure instead of the injection rate. A typical power plant produces 10 Mt of CO₂ per year.

An important part of this work was devoted to the question of spatial variability as a source of uncertainty. The results reported above were obtained for the homogeneous case, that is to say all the input uncertain variables including the properties of the underground medium were assumed to be spatially homogeneous. Neglecting the presence of heterogeneities is quite common in practical risk evaluation studies (particularly for

rapid screening and ranking). However, the probability to encounter a natural geological medium with homogeneous properties at a given scale is almost zero. Nonetheless, a parametric uncertainty propagation study with homogeneous variables has a meaning when interpreted as the representation of the lack of knowledge on the equivalent properties of the medium.

In this context, we tried to compare with the help of numerical simulations two probabilistic uncertainty propagation studies. In the first, the fixed PDF was sampled attributing each value to a homogeneous equivalent property. In the second, the same fixed PDF was reproduced at each sample of a random field. It was found that the intrinsic permeability is the most influential factor for both maximal lateral extent and the fluid average overpressure according to the performed numerical simulations. Therefore, it was chosen to conduct the comparison of probabilistic studies where the same PDF was attributed to homogeneous and heterogeneous intrinsic permeability. For the heterogeneous representation, a numerical code was written for generating samples of anisotropic tensor-valued random fields. The uncertainty propagation for the homogeneous case is performed through a metamodel (by polynomial chaos expansion). For the heterogeneous case each sample of the permeability random field serves as input to the coupled 3D/2D rotation-invariant model. For both cases one hundred dynamic flow simulations is run serving in the homogeneous case as support for the construction of the metamodel and in the heterogeneous case providing rich information owing to the spatial repartition. The conclusion of the comparative study is the following: for chosen variables of interest the uncertainty propagation in homogeneous case gives a significantly larger variability than the heterogeneous case (by factor more than 20 for the variance of the maximal lateral spread and 75 for the averaged fluid overpressure). The satisfactory result is that considering the uncertainty on the equivalent properties (homogeneous case) shifts the estimation in the direction of higher security: larger predicted output variance and higher exceedence probabilities (although it could be inverse for some other variable of interest). Nonetheless, such overestimation can be inconvenient for decision making, when a more precise estimation is needed.

Following this analysis, the next arising question is how much we should reduce the input homogeneous variability to reproduce the results of a study with heterogeneous variation. An approximate answer to this question is found owing to the inverse transform sampling. Reproducing the CCDF of the maximal lateral spread has given a factor twenty one for the variance of the intrinsic permeability, which means that in this particular study the output variance is proportional to the input variance.

Whence it follows that the subjective decision to represent the input uncertainty with or without spatial variability changes significantly the results.

In the light of the obtained results the following further works and improvements are suggested:

The numerical model can be further enhanced by including the mechanical effect on the matrix and chemical reactions, which would allow to consider the entire system evolution till more advanced time steps (thousands of years, when dissolution and chemical reactions become dominating). The injection controlled by pressure needs to be numerically evaluated.

Further quantification of the chosen scenarios. The results presented in Chapter 4 contribute to the evaluation of the scenarios of leakage by lateral migration and leakage through the caprock. In order to estimate the probability of leakage by lateral migration, the results for the maximal lateral spread of the CO₂ cloud should be coupled with the probability of reaching a fault at a given distance. Discrete faults of fixed or random dimensions could be generated through Poisson process of given density. If the leakage event is defined as

$$E_{leakage} = \{r_{max}(T) \geq r_f\},$$

then the probability of leakage is:

$$P(E_{leakage}) = \int_{r=0}^{+\infty} P(\exists \text{ a fault with } r_f \leq r) \cdot p(r_{max}(T)) dr.$$

For little values of radius, the study could be completed with 3D computations.

For the second scenario of leakage through the caprock, the results of fluid overpressure on the top of the reservoir could be coupled with a 1D vertical solution in order to estimate the leakage intensity. To compute the equivalent vertical permeability of the caprock, it is proposed either to perform a numerical integration of $1/K$ of a heterogeneous block, or to perform a 3D simulation with heterogeneous caprock permeability and numerically find an equivalent value.

To fulfill the risk quantification in the framework of CCS risk assessment elaborated at INERIS (to which the present work contributes), all the identified central events should be considered.

Regarding the metamodel quality optimal basis functions can be found depending on the input PDF type. In the present application Hermite polynomials provide a sufficient precision. If the statistical characteristics of input variables are constrained to data (e.g. the input PDFs have a kernel representation), arbitrary PCE can be applied such as proposed by [Ashraf et al. \(2013\)](#). Furthermore, it is straightforward to evaluate time-dependent risk measures as well as time-dependent sensitivity indexes when using PCE decomposition. To do so, a meta-model at each time step should be built (the computation resources needed for this procedure are quite modest, insignificant compared to the flow simulation itself).

Concerning the two-level setting study (Sections 3.5.2 and 4.2), it would be interesting to compute the contribution of level-2 variability to the output variance such as proposed by [Sankararaman and Mahadevan \(2013\)](#).

Concerning the spatial variability analysis different correlation functions (such as exponential) and higher fluctuation levels could be considered for the random tensor-valued permeability field, as it can have a significant impact on the flow. It is also important to perform a sensitivity analysis with respect to the anisotropy factor and correlation lengths to determine their influence on the flow characteristics.

In a more general sense, it seems that the question of uncertainty quantification related to field identification has not yet been studied enough. In the light of the conclusions of

the present work, which emphasize the importance of taking into account the spatial variability, it would be necessary to quantify the impact of the second level uncertainty related to the imprecision of field identification given a limited quantity of indirect and imprecise data.

It is presumed that two-level uncertainty studies for homogeneous and heterogeneous cases would give closer results than the one level studies reported in the present work.

Appendices

Appendix A

Review of risk assessment methodologies and tools applied to CCS

FEP An approach of risk assessment with the use of catalogs of **Features** of the geologic system which impacts the behavior of the storage, discrete **Events** and continuous **Processes** which can influence the storage behavior as well. TNO (Netherlands Organization for Applied Scientific Research) has worked on a FEP database specific to CO₂ storage and this work led to the creation of a new framework for faster and more coherent analysis: Carbon Storage Scenario Identification Framework (CASSIF) (Yavuz et al., 2009). To the knowledge of the author, there exist two generic FEP databases for geologic storage of CO₂. One was created by TNO and the other by Quintessa (Savage et al., 2004). This quite exhaustive list serves to derive site-specific lists of FEPs allowing further comprehensive evaluations of each site's particularities. This method can be used as a preliminary study complemented afterwards by figuring out the relations between FEPs and by quantifying possible scenarios.

Scenario approach based on FEP This approach consists in regrouping the FEPs to form long-term evolution scenarios and quantifying them with the help of conceptual models, running deterministic or probabilistic simulations. This approach was recommended by (Wildenborg et al., 2004, 2005). An application for In Salah site is described by (Paulley et al., 2011). The same path is followed in the ICARAS framework (Wollenweber et al., 2013): scenarios defined with the help of FEPs are investigated in a quantitative way, by fast mathematical models or by numerical codes. The tools give access to probabilistic sensitivity analysis.

FEP + discrete time Markov chains The paradigm of Markov chains allow to overcome the limitation of dividing all the possible states of the system into a set of scenarios. Instead, the states of the system become the main actors. Any combination of Events and Processes is a possible state. Of course the states for a description of the subsurface system should be simplified as well. In order to limit the number of states only principal events and processes are chosen. The system can switch from one state

to another. The probabilistic evolution of the system can be followed in terms of visit times of each state. Therefore, the advantage of such an approach is that it allows to compute relevant evolutionary timescales of the system. A demonstrative application to CO₂ storage was performed by (Nepveu et al., 2009). The main difficulty of the method is that it needs the transition probabilities from one state to another which would be left to the experts estimation. Further theoretical information on absorbing Markov chains can be found in (Grinstead and Snell, 1997).

The state space becomes too large for calculation with Markov models when the number of nodes in a fault tree increases. (Durga Rao, 2009) proposed a dynamic fault tree analysis using Monte Carlo simulation, which could become interesting for implementation for CO₂ storage with the development of computational resources.

VEF : The vulnerability evaluation framework is designed as a conceptual framework to the service of regulators and technical experts for identifying areas that could require specific risk assessment, monitoring, and management. It is a qualitative systematic method which identifies conditions which could influence the susceptibility to consequence. VEF has some similarities to the Certification Framework Approach (CFA).

Delphi Delphi is a classical team-oriented method of risk analysis for hazard identification by means of brain-storming.

SWIFT : The Structured What-If Technique (SWIFT) considers deviations from normal evolution identified by brainstorming , with questions beginning 'What if...?' and 'How could...?'. It is a form of Delphi risk analysis for systematic qualitative hazard identification. This method was developed as an alternative to the Hazard and Operability (HAZOP) technique and to the Failure Modes and Effects Analysis (FMEA). SWIFT is flexible and can be modified to suit each individual application. SWIFT review of CO₂ sequestration in geological structures was performed by (DNV, 2010). As FEP, it can be followed by a more detailed stage of analysis and quantification: (Vendrig et al., 2003) tried to develop a generic quantitative risk assessment for CCS on the basis of SWIFT. The authors came to a very important conclusion that due to the presence of different types of uncertainties, the risk levels can only be site-specific.

SRF Screening and ranking framework is implemented as a spreadsheet where user enters the scores representing expert opinions and information available on the site as well as uncertainty estimations in order to evaluate three basic characteristics of a site: potential for primary containment, potential for secondary containment if the primary leaks, potential of attenuation if both containments fail. SRF is quite similar to CFA but it is based on the assumptions about the three basic characteristics of a site cited above.

RISQUE The Risk Identification and Strategy using Quantitative Evaluation (RISQUE) is a systematic approach which uses expert panel judgments to identify risk events and evaluate them in terms of likelihood, consequences and time scale of occurrence. RISQUE

uses the acceptability criteria based on six performance indicators: containment, effectiveness, self-funding potential, wider community benefits, community safety, and community amenity. The methodology has been applied to several CO₂ storage sites in Australia (Bowden and Rigg, 2004)

CO2PENS This tool uses system-modeling approach and aims at integrating in a system-level model a number of process-level models such as the behavior of the injected fluid in the reservoir and caprock, potential release mechanisms, transport of CO₂ from the reservoir, release on surface (Stauffer et al., 2009; Viswanathan et al., 2008). A system-level model englobes the entire system from the moment of capture of CO₂, through transportation by pipeline, to injection and storage. The economic aspects can also be considered. The advantage of this tool is that it allows to choose the processes the user wants to include in a study and choose any kind of description for chosen processes (simplified analytical calculation or a variety of numerical simulators). The tool was originally designed to perform probabilistic simulations for the whole CCS chain, therefore it favors the use of probabilistic methods for uncertainty propagation.

CFA The certification framework approach also decomposes the system into process-level models. For quantification of risk, the system is divided into compartments. Conduits for leakage from source to compartments or from one compartment to another may be wells or faults. The likelihood of leakage is evaluated by estimating the probability that a conduit for leakage encounters the CO₂ plume and a target at the same time. The CO₂ flux across the pathway is simulated through deterministic simplified models, and the impacts of the release compared to acceptable thresholds. A level of risk is obtained by the product of the values of the probability and the consequences. CFA is similar to the VEF with the difference that it adds values for the leakage probability.

MCA is a tool at the service of social consideration. It delivers a rich profile of the views and preferences of participants (stakeholders) and enables to overview key issues that affect the prospects for further development of a project. It covers a variety of evaluation techniques sharing a basic framework under which the alternatives can be scored against a set of defined criteria. This list of criteria is proposed according to the fundamental goals of the geological carbon storage (Gough and Shackley, 2006; Jakobsen et al., 2013). A similar method is the Multi-Attribute Utility Theory (MAUT) which furthermore assumes a dependency of preferences of criteria.

ESL The evidence support logic is designed to identify the amount of uncertainty involved in sub-decisions. It addresses uncertainty that arises from lack of knowledge or scarcity of data.

MOSAR Organized and Systemic Method of Risk Analysis (MOSAR) is designed for analyzing the technical risks of a human plant and for identifying prevention means to neutralize them. The MOSAR method relies on a step by step method. It was applied to CO₂ storage in a generic approach.

P&R The Performance and Risk assessment for well integrity is based on the classical definition of risk (likelihood versus severity). The notion of likelihood (probability) is associated with system uncertainties and the notion of severity is associated with the mass of leaked CO₂. The method includes the definition of a Risk Acceptance Limit (RAL), which gives the operators a support for demonstrating safety to regulators. This methodology focuses on the risks of contamination of subsurface formations and hazardous releases on surface.

PRA The PRA approach ([Rish, 2005](#)) was developed for UIC Class I hazardous waste injection wells (who tells that CO₂ is a class I hazardous waste?). In this method probabilities of events and distributions of porous media and well properties are used as input for probabilistic calculations of the likelihood of events. In the next phase of the assessment, the consequences of a scenario or of an event are expressed in terms of impact of long-term high concentrations of CO₂ on key receptors. The consequences are evaluated through simulation.

MANAUS The MANAUS method is a full workflow that covers all steps of the risk assessment from the initial feedback analysis to the final evaluation of the safety barriers. The core of the method is the risk analysis, that includes 3 major steps: 1) definition of scenarios with a systematic analysis (similar to "What-if") and with the help of keywords that are derived from the FMEA method (in the specific underground context), 2) event tree representation with the OSQAR software tool, 3) quantification of scenarios with numerical modelling tools (including Monte Carlo simulations to estimate uncertainty).

CQUESTRA A risk and performance assessment code for geological sequestration of carbon dioxide is a computationally efficient semi-analytical code which is used for the probabilistic risk assessment and rapid screening of potential sites for geological sequestration of carbon dioxide. The model takes into account advection, dispersion, diffusion, buoyancy, aquifer flow rates and local formation fluid pressure.

Logic tree [Gerstenberger et al.](#) proposes a risk assessment methodology that relies technical, social, policy, safety and economic risk factors for each of the four main components of the CCS process (i.e., capture, transport, injection and storage). The methodology uses a logic tree approach coupled with the cost of avoided CO₂ and the HSE costs. Each of the four parts of the tree (corresponding to the four main components of the CCS process) rely five sub-parts (corresponding to the nature of a risk factor). A probability is assigned to each branch (YES/NO choice in each of five sub-parts) of the logic tree. The methodology highlights risk factors that may hamper a successful sequestration (it may be a sequence of risk factors of different nature). The methodology also helps to identify risk factors that may have unacceptably high uncertainties.

A systemic risk management approach for the whole chain of CO₂ Capture-Transport-Storage projects has been recently studied by ([Samadi, 2012](#)). The proposed approach integrates the legal and financial risks.

Appendix B

Deterministic and local probabilistic results for the simplified physical model

B.1 Deterministic response surface

The construction of a deterministic response surface consists in mapping the variable of interest at a certain number of points in the space of input parameters and then interpolating between the points to obtain the values over the entire domain. It can also be seen as a probabilistic response surface with uniformly distributed input variables, but few natural parameters have a uniform distribution, furthermore, the goal of a deterministic response surface is to quantify the response itself independently from the probability density. An example of response surfaces (plotted for two varying parameters at a time) is given on Figure [B.1](#). The interpolation can be performed through polynomial regression ([Box and Draper, 1986](#)), kriging, heuristic laws, artificial neural networks, etc ([Kleijnen and Sargent, 2000](#)).

The interest of such an analysis is that it captures the global trends of the function over the domain of the input variables. Of course, the quality of the approximation of the actual function by a response surface depends on its regularity and the size of the chosen support (characteristic distance between 2 points at which the function is evaluated). If the support is small enough with respect to the non-linearities of the function, the constructed response surface can be used as a computationally lighter version of the true function (it is also called a surrogate model or a meta-model). This type of analysis gives access to local sensitivity indexes in each point. They are computed by finite differences for the neighboring points.

Difficulties arise when a growing number of input variables are at stake and when the number of combinations for multiple variables becomes computationally expensive to evaluate.

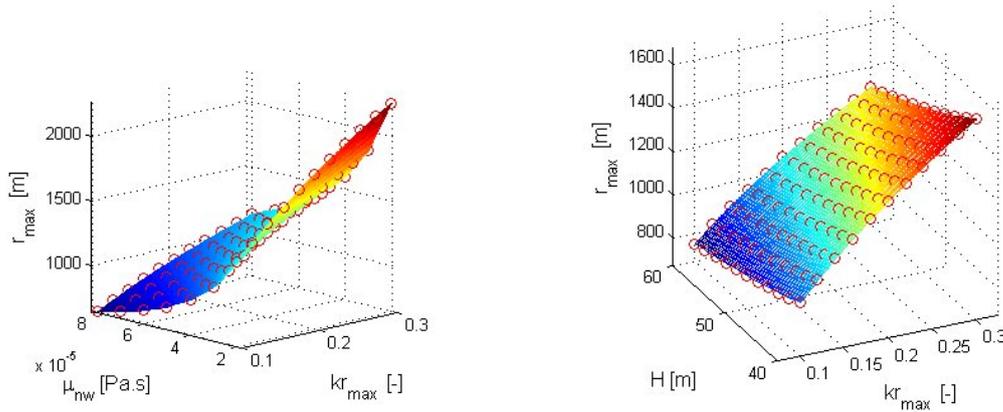


Figure B.1: Deterministic response surface of the furthest spread of CO_2 cloud for two parameters: gas viscosity and relative permeability endpoint (left) and reservoir thickness and relative permeability endpoint (right)

B.2 Local probabilistic sensitivity analysis

Local analysis consists in perturbing the vector of input parameters in order to capture the behavior of the variable of interest around a given point (called nominal hereafter). In this case, the uncertainty of the output can be estimated through the partial derivatives of the function relying the input parameters and the output. According to ([Joint Committee for Guides in Metrology, 2008](#)) the variance can be decomposed in the following way:

$$\text{Var}(z) = \sum_{i=1}^N \sum_{j=1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \text{Cov}(x_i, x_j) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 \text{Var}(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \text{Cov}(x_i, x_j) \quad (\text{B.1})$$

where $\text{Cov}(x_i, x_j) = \text{Cov}(x_j, x_i)$ is the covariance associated with the two input variables x_i and x_j . This approximation can be used in cases where the uncertainties are small with respect to non-linearities of the function f .

The general remarks concerning the local approach can be made:

- Local sensitivity analysis allow to rank the input parameters by sensitivity or importance around a point.
- Sensitivity depends on the chosen physical units. Therefore to avoid this dependency the parameters should be normalized by their initial values.
- The importance of a parameter is more significant than the sensitivity of the function with respect to this parameter

The expressions of sensitivities for the maximal spread of the CO_2 cloud are presented in Table B.1.

For the Nordbotten model all the parameters contribute to r_{max} with the same

Table B.1: Analytical computation of local response sensitivity with respect to each input parameter around the nominal point. Computation of the importance of each parameter. The final first order Taylor approximation for r_{\max} variance is $4.9 \cdot 10^4 \text{ m}^2$, corresponding to a standard deviation of 221 m.

Parameter	Nominal point	Partial derivative	Sensitivity (S)	Variance	Importance (I)	Normalized importance
x_i	x_{i_0}	$\frac{\partial f}{\partial x_i}$	$\left(\frac{\partial f}{\partial x_i}(x_{i_0})\right)^2$	$Var(x_i)$	$S \cdot Var(x_i)$	$\frac{I}{Var(z)}$
kr_{nw}	0.20	$\frac{r_{max}}{2kr_{nw}}$	$8.5 \cdot 10^6$	$2.5 \cdot 10^{-3}$	21329	0.44
H	50	$-\frac{r_{max}}{2H}$	$1.37 \cdot 10^2$	$2.5 \cdot 10^1$	3413	0.07
ϕ	0.15	$-\frac{r_{max}}{2\phi}$	$1.5 \cdot 10^7$	$2 \cdot 10^{-4}$	3413	0.07
$1 - sr_{nw}$	0.33	$-\frac{r_{max}}{2(1-sr_{nw})}$	$7.8 \cdot 10^5$	$4.4 \cdot 10^{-3}$	3413	0.07
μ_w	0.35	$\frac{r_{max}}{2\mu_w}$	$2.7 \cdot 10^{12}$	$1.2 \cdot 10^{-9}$	3413	0.07
μ_{nw}	0.05	$-\frac{r_{max}}{2\mu_{nw}}$	$1.4 \cdot 10^{14}$	10^{-10}	13651	0.28
					$Var(z) = 4.9 \cdot 10^4$	$\sum = 1$

power (+0.5 or -0.5). Consequently:

- r_{max} has the same sensitivity for all normalized input parameters
- The importance of any parameter depends only on the coefficient of variation and the nominal output value. It is invariant to the input parameters normalization, but changes when the output is normalized.
- The normalized importance depends only on the coefficient of variation. It is invariant with respect to both the input parameters and the output normalization.

Normalized importance is equal to $CV(x_i)^2 / \sum_{i=1}^N CV(x_i)^2$

For the Nordbotten's model and the coefficients of variation chosen in this study the input parameter which contributes the most to the variance of the output is the gas relative permeability endpoint (44%) followed by the gas viscosity (28%). For further comparison with other methods: nominal output value is 1168 m, standard deviation estimated by a second order Taylor expansion (obtained through equation B.1) is 221 m.

B.3 Analytical computation of the response PDF characteristics.

The algebraic operations included in the analytical formula of the response computation 2.48 are the square root, multiplication and division of the input parameters. Therefore, the properties of a lognormal distribution listed hereafter

Properties of lognormally distributed random variables:

- Product of two lognormally distributed variables $x_1 \sim \text{LogN}(\mu_1, \sigma_1)$ and $x_2 \sim \text{LogN}(\mu_2, \sigma_2)$ is a lognormally distributed variable $z = x_1 x_2 \sim \text{LogN}(\mu, \sigma)$ with $\mu = \mu_1 + \mu_2$, $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.
- Product of a lognormally distributed variable $x_1 \sim \text{LogN}(\mu_1, \sigma_1)$ and a constant 'a' is a lognormally distributed variable $z = ax_1 \sim \text{LogN}(\mu, \sigma)$ with $\mu = \mu_1 + \ln(a)$, $\sigma = \sigma_1$.
- Ratio of two lognormally distributed variables $x_1 \sim \text{LogN}(\mu_1, \sigma_1)$ and $x_2 \sim \text{LogN}(\mu_2, \sigma_2)$ is a lognormally distributed variable $z = x_1/x_2 \sim \text{LogN}(\mu, \sigma)$ with $\mu = \mu_1 - \mu_2$, $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.
- Square root of a lognormally distributed variable $x_1 \sim \text{LogN}(\mu_1, \sigma_1)$ is a lognormally distributed variable $z = \sqrt{x_1} \sim \text{LogN}(\mu, \sigma)$ with $\mu = 0.5\mu_1$, $\sigma^2 = 0.5\sigma_1^2$

allow of an analytical computation of the probability density function for the maximum extent of CO₂ cloud.

The resulting PDF of r_{max} is lognormal with μ and σ that can be computed from the parameters of input lognormal PDFs. For the values of probabilistic characteristics of the input variables chosen for this study (see Table 3.7) the final PDF of the variable of interest at stake is:

$$r_{max} \sim \text{LogN}(7.063, 0.187), \quad (\text{B.2})$$

which corresponds to the characteristics reported in Table B.2 and illustrated in Figure B.2.

Table B.2: Analytically calculated quantities of interest

Quantity of interest	Expression	Value
Mode(r_{max})	$e^{(\mu(r_{max}) - \sigma^2(r_{max}))}$	1128 m
Median(r_{max})	$e^{\mu(r_{max})}$	1168 m
Mean(r_{max})	$e^{(\mu(r_{max}) + \frac{1}{2}\sigma^2(r_{max}))}$	1188 m
Var(r_{max})	$e^{(\sigma^2(r_{max}) - 1)} \cdot e^{(2\mu(r_{max}) + \sigma^2(r_{max}))}$	$5.0 \cdot 10^4 \text{ m}^2$
Std(r_{max})	$\sqrt{\text{Var}(r_{max})}$	224 m
P($r_{max} > r_0$)	$\frac{1}{2} - \frac{1}{2} \text{erf}\left(\frac{\ln(r_0) - \mu}{\sigma\sqrt{2}}\right)$ ^a	$9.02 \cdot 10^{-2}$ for $r_0 = 1500 \text{ m}$ $0.20 \cdot 10^{-2}$ for $r_0 = 2000 \text{ m}$
Quantile		
$r_0: P(r_{max} < r_0) = P^0$	$e^{\sqrt{2}\sigma \text{erf}^{-1}(2P^0 - 1) + \mu}$	1484 m for $P^0 = 90\%$ 1588 m for $P^0 = 95\%$ 1804 m for $P^0 = 99\%$

^aerf stands for the error function: $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

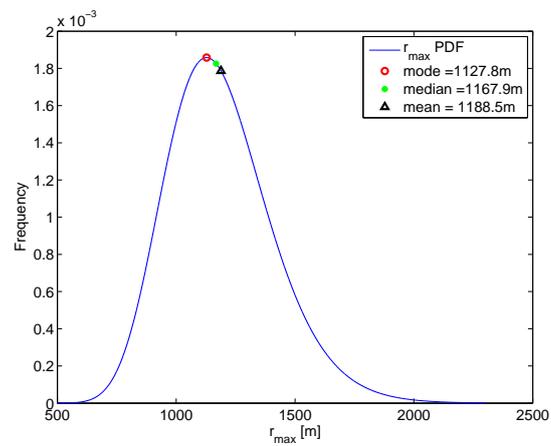


Figure B.2: Analytically calculated probability density function of the variable of interest r_{max} (maximal horizontal spread of the injected CO₂)

The characteristics of the r_{max} PDF are used as reference for the numerical uncertainty propagation by Monte Carlo method. The analytical PDF allows the quantification of the convergence of the numerical estimation.

Appendix C

Polynomial Chaos Expansion theory

Let z be the scalar output of a model that depends on the n -dimensional random vector $\mathbf{x} = (x_1, \dots, x_n)$ collecting n random parameters x_i , valued in $\Omega_i \subseteq \mathbb{R}$, $i = 1, \dots, n$:

$$z = f(\mathbf{x}), \quad \mathbf{x} \in \Omega \subseteq \mathbb{R}^n, \quad (\text{C.1})$$

where Ω denotes the Cartesian product $\Omega_1 \times \dots \times \Omega_n$.

Suppose x_1, \dots, x_n to be statistically independent and the response z to be a second-order integrable variable: the PCE of z (Le Maître and Knio, 2010; Ghanem and Spanos, 1991; Xiu and Karniadakis, 2002) is the spectral decomposition of f over a set of polynomials $\Psi_k(\mathbf{x})$, which are orthogonal with respect to the probability measure on Γ (i.e. the law of \mathbf{x}), namely (Sudret, 2008b):

$$z = \sum_{k=0}^{\infty} \beta_k \Psi_k(\mathbf{x}). \quad (\text{C.2})$$

The PCE was first introduced for Gaussian input parameters involving Hermite polynomials (e.g., (Wiener, 1938)). In the presence of non-Gaussian distributed variables, two strategies can be adopted in order to preserve the optimal exponential convergence rate: i) transform the input parameters to obtain Gaussian random entries ii) consider other families of polynomials (Xiu and Karniadakis, 2002; Soize and Ghanem, 2005).

In this work, for every input parameter x_i , $i = 1, \dots, n$, a log-normal distribution $\rho_{\Gamma_i}(x_i)$ is considered. In this case, the simplest way to reach the exponential rate of convergence is to use the Hermite basis, by transforming each input parameter x_i into a standard normal random variable ξ_i :

$$\xi_i = \Phi^{-1}(F_{x_i}(x_i)), \quad (\text{C.3})$$

where Φ and F denotes the CDF of a standard Gaussian variable and of the variable x_i respectively. Given the vector of transformed (independent) input parameters $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$, the PCE can be computed as:

$$z = \tilde{f}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} \beta_k \Psi_k(\boldsymbol{\xi}). \quad (\text{C.4})$$

For computation, expansion (C.4) needs to be truncated, obtaining finally an approximation of the response as:

$$z \approx \sum_{k=0}^{P-1} \beta_k \Psi_k(\boldsymbol{\xi}), \quad (\text{C.5})$$

where $P = \binom{n+p}{p}$, if (C.5) is the truncated expansion on a basis of n -dimensional Hermite polynomials of degree not exceeding p .

The P coefficients β_k , $k = 0, \dots, P-1$ appearing in C.5 are assumed deterministic and fully characterize the randomness of the response z . In applications, these coefficients are unknown and thus need to be properly estimated. A number of methods are presented in the literature (Sudret, 2008b; Crestaux et al., 2009); among them, in this work, the regression method is used. It consists in estimating β_k by least-squares on the basis of an experimental design $\{z^j, j = 1, \dots, N\}$. The experimental design is built by collecting N exact solutions z^j , $j = 1, \dots, N$, which are derived solving the complete mathematical model for N realizations \boldsymbol{x}^j , $j = 1, \dots, N$, of the random input \boldsymbol{x} . The least-squares estimate of $\boldsymbol{\beta} = (\beta_k) \in \mathbb{R}^P$ is found by solving the minimization problem:

$$\boldsymbol{\beta} = \underset{\beta_0, \dots, \beta_{P-1} \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{N} \sum_{j=1}^N \left\{ z^j - \sum_{k=0}^{P-1} \beta_k \Psi_k(\boldsymbol{\xi}^j) \right\}^2, \quad (\text{C.6})$$

being $\boldsymbol{\xi}^j = (\Phi^{-1}(F_{x_i}(x_i)))$. If the design matrix $\boldsymbol{\Psi} = (\Psi_j(\boldsymbol{\xi}^i)) \in \mathbb{R}^{N,P}$ is of full rank, problem (C.6) admits the unique solution $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \boldsymbol{z}, \quad (\text{C.7})$$

where $\boldsymbol{z} = (z^1, \dots, z^N)$. Techniques for an efficient computation of coefficients estimates (C.7) has been presented in the literature (e.g., (Sudret, 2008b)) and are exploited in the simulations here presented.

The utility of PCE for the uncertainty analysis of the response is two-fold:

- Response mean and variance can be directly obtained from C.5 as:

$$\mu(z) = \beta_0; \quad \operatorname{Var}(z) \approx \sum_{k=0}^{P-1} \beta_k^2 \|\psi_k\|^2.$$

Analogously, the sensitivity Sobol' indices can be computed directly from coefficients β_k , $k = 0, \dots, P-1$ (see hereafter)

- The approximation C.5 can be used as meta-model for Monte-Carlo simulations (i.e., the response is simulated through the approximate model instead of solving the PDE system). This allows to perform a much more efficient assessment of the CDF as the approximated model is easier to simulate compared to the full model C.1.

Computation of Sobol' sensitivity indices from PCE meta-model In this section the definition of Sobol' indices (Sobol, 1993) is recalled and the way their computing from a PCE meta-model is shown.

The total variance of the variable of interest can be decomposed into the effect of each random parameters x_j and the joint effects of all the subsets x_{i_1}, \dots, x_{i_s} of the input parameters x_1, \dots, x_n . As stated in (Sobol, 1993; Sudret, 2008b; Formaggia et al., 2012), this variance decomposition follows from the Sobol' decomposition of the function $\tilde{f}(\boldsymbol{\xi})$ which appears in (C.4):

$$\begin{aligned} \tilde{f}(\boldsymbol{\xi}) = & \tilde{f}_0 + \sum_{i=1}^N \tilde{f}_i(\xi_i) + \sum_{i,j=1}^N \tilde{f}_{i,j}(\xi_i, \xi_j) + \\ & + \dots + \tilde{f}_{1,2,\dots,n}(\xi_1, \xi_2, \dots, \xi_n), \end{aligned} \quad (\text{C.8})$$

where

$$\begin{aligned} \tilde{f}_0 &= \int_{\Gamma} \tilde{f}(\boldsymbol{\xi}) \rho_{\Gamma}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \\ \tilde{f}_{\mathbf{i}^{(s)}}(\boldsymbol{\xi}_{\mathbf{i}^{(s)}}) &= \int_{\Gamma_{-\mathbf{i}^{(s)}}} \tilde{f}(\boldsymbol{\xi}) \rho_{\Gamma_{-\mathbf{i}^{(s)}}}(\boldsymbol{\xi}_{-\mathbf{i}^{(s)}}) d\boldsymbol{\xi}_{-\mathbf{i}^{(s)}} - \sum_{\mathcal{I} \subseteq \mathbf{i}^{(s)}} \tilde{f}_{\mathcal{I}} \end{aligned}$$

denoting with $\mathbf{i}^{(s)}$ the multi-index $\{i_1, \dots, i_s\}$, with $\boldsymbol{\xi}_{\mathbf{i}^{(s)}}$ the collection $\{\xi_{i_1}, \dots, \xi_{i_s}\}$, with $\rho_{\Gamma} = \prod_{i=1}^n \rho_{\Gamma_i}$ the multivariate gaussian density of $\boldsymbol{\xi}$, $\Gamma_{-\mathcal{I}} = \times_{i \in \{1, \dots, N\} \setminus \mathcal{I}} \Gamma_i$ and $\rho_{\Gamma_{-\mathcal{I}}} = \prod_{i \in \{1, \dots, N\} \setminus \mathcal{I}} \rho_{\Gamma_i}$.

Notice that in the considered case $\Gamma_i = \mathbb{R}$, $i = 1, \dots, n$, but what follows remains valid for any measurable $\Gamma_i \subseteq \mathbb{R}$. In particular, decomposition (C.8) is unique whenever \tilde{f} is integrable over Γ and each term $\tilde{f}_{\mathbf{i}^{(s)}}$ in (C.8) is orthogonal with respect to the others.

To define the Sobol' indices, call V_f , $V_{\tilde{f}}$ the variance of f , \tilde{f} respectively. The Sobol' index relative to the mixed effect $\boldsymbol{\xi}_{\mathbf{i}^{(s)}}$ (or to $\mathbf{x}_{\mathbf{i}^{(s)}}$) is defined as:

$$S_{\mathbf{i}^{(s)}} = \frac{1}{V_{\tilde{f}}} \int_{\Gamma_{\mathbf{i}^{(s)}}} \tilde{f}_{\mathbf{i}^{(s)}}^2(\boldsymbol{\xi}_{\mathbf{i}^{(s)}}) \rho_{\Gamma_{\mathbf{i}^{(s)}}} d\boldsymbol{\xi}_{\mathbf{i}^{(s)}}, \quad (\text{C.9})$$

where $\Gamma_{\mathbf{i}^{(s)}} = \Gamma_{i_1} \times \dots \times \Gamma_{i_s}$. Hence, $S_{\mathbf{i}^{(s)}}$ represents the proportion of the total variance explained by $\boldsymbol{\xi}_{\mathbf{i}^{(s)}}$ (or $\mathbf{x}_{\mathbf{i}^{(s)}}$) and thus the Sobol' indices sum to 1.

Concerning the total effect of a parameter $\xi_i(x_i)$ (or of x_i) on the total variability, it can be computed directly from Sobol' indices as:

$$S_i^T = \sum_{\mathbf{i}^{(s)} : i \in \mathbf{i}^{(s)}} S_{\mathbf{i}^{(s)}}. \quad (\text{C.10})$$

The computational effort needed to compute the sensitivity indices (C.9) and (C.10) consists mainly in the cost of calculating (numerically) $2^n - 1$ integrals. To this end, Monte-Carlo quadrature schemes can be used but, due to its low rate of convergence, the computational cost might become unaffordable. However, this problem can be faced and solved by introducing the PCE of the response z , since Sobol' indices can be directly

computed from coefficients β_k , $k = 0, \dots, P - 1$. Indeed, expansion (C.5) can be reordered so that it is equivalent to (C.8):

$$\tilde{f}(\boldsymbol{\xi}) = \sum_{\mathbf{i}^{(s)} \subseteq \{1, \dots, n\}} \sum_{k \in K_{\mathbf{i}^{(s)}}} \beta_k \Psi_k(\boldsymbol{\xi}), \quad (\text{C.11})$$

where $K = \{0, \dots, P - 1\}$, $K_{\mathbf{i}^{(s)}} = \{k \in K \mid \Psi_k(\boldsymbol{\xi}) = \Psi_k(\boldsymbol{\xi} = \boldsymbol{\xi}_{\mathbf{i}^{(s)}})\}$ (Crestaux et al., 2009).

From the equivalence between (C.4) and (C.11) the bijective relation existing between Sobol' indices and coefficients β_k , $k = 1, \dots, n$ can be derived. Indeed, the *PC-based Sobol' indices* can be obtained as:

$$S_{\mathbf{i}^{(s)}} = \frac{\sum_{k \in K_{\mathbf{i}^{(s)}}} \beta_k^2 \langle \Psi_k, \Psi_k \rangle}{\sum_{k=0}^{P-1} \beta_k^2 \langle \Psi_k, \Psi_k \rangle},$$

being $\mathbf{i}^{(s)} \subseteq \{1, \dots, n\}$, while the total indices can be computed from (C.10).

Appendix D

Meta-model results for seven uncertain inputs

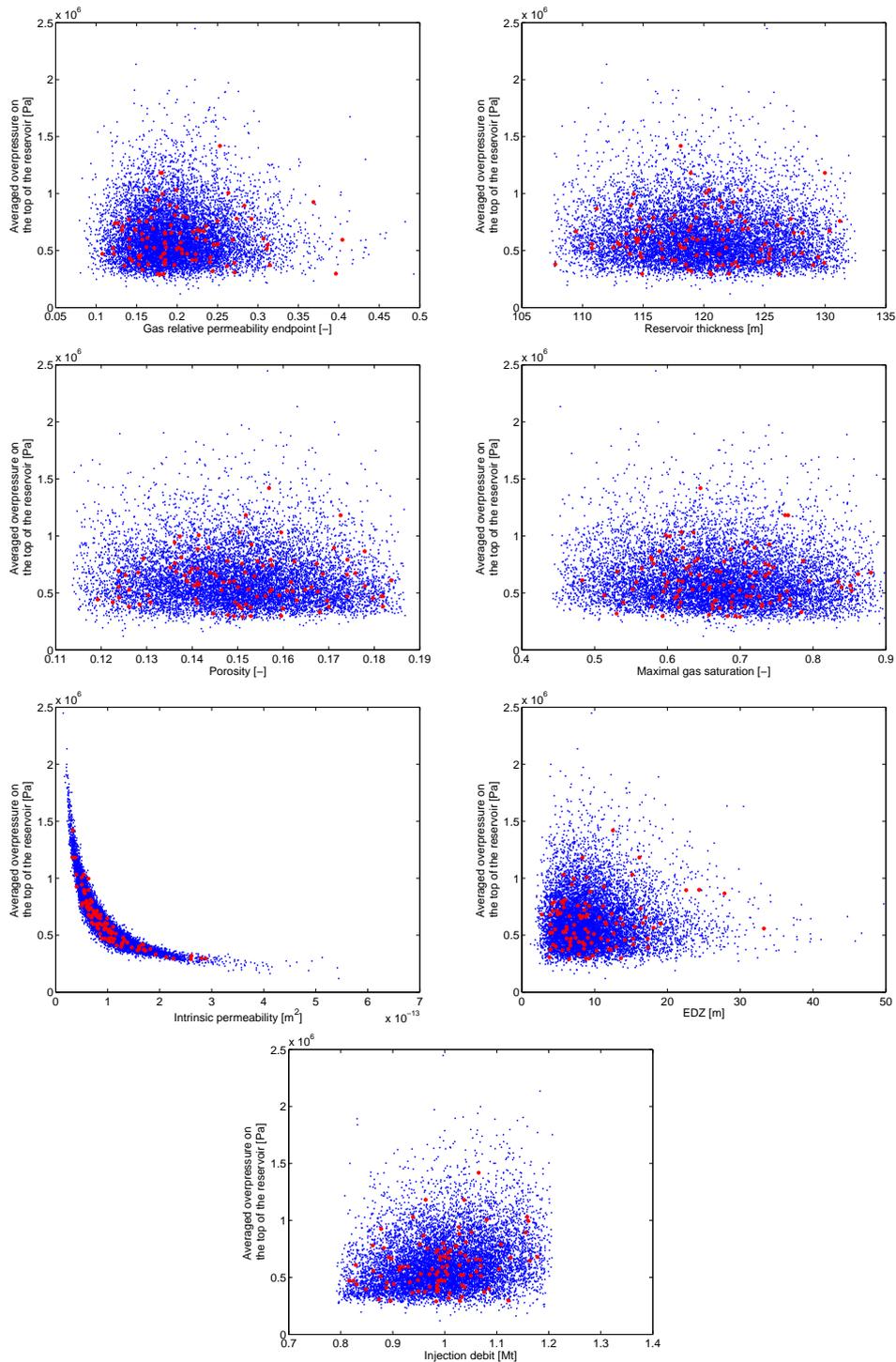


Figure D.1: Experimental design of the support of the meta-model: 100 standard Monte Carlo simulations (red points), and experimental design for meta-model analysis: 10000 samples generated through the meta-model (blue points). Variable of interest is averaged gas overpressure on the top of the reservoir.

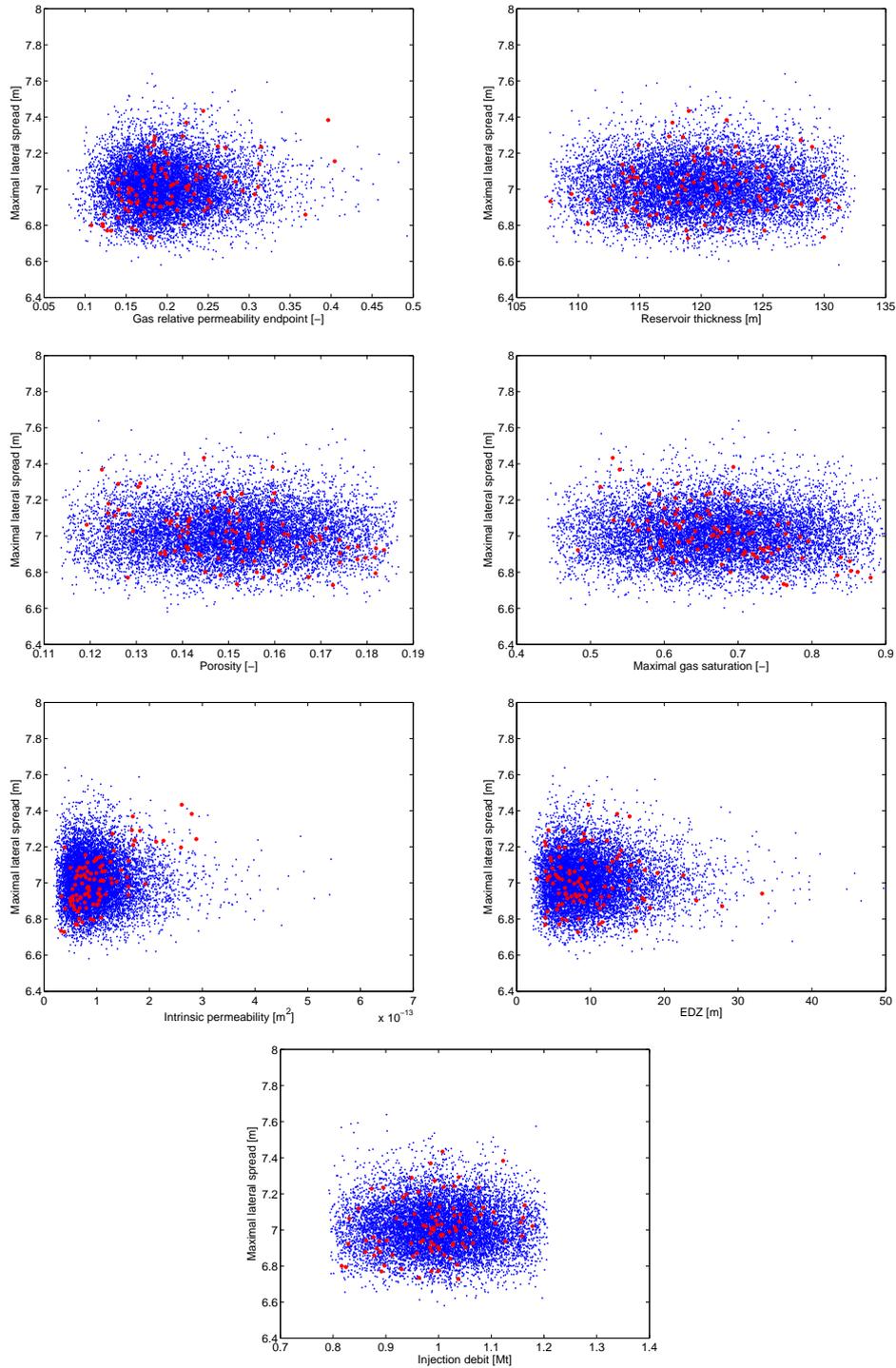


Figure D.2: Experimental design of the support of the meta-model: 100 standard Monte Carlo simulations (red points), and experimental design for meta-model analysis: 10000 samples generated through the meta-model (blue points). Variable of interest is the maximal lateral spread of CO_2 cloud.

Appendix E

The coupled model results

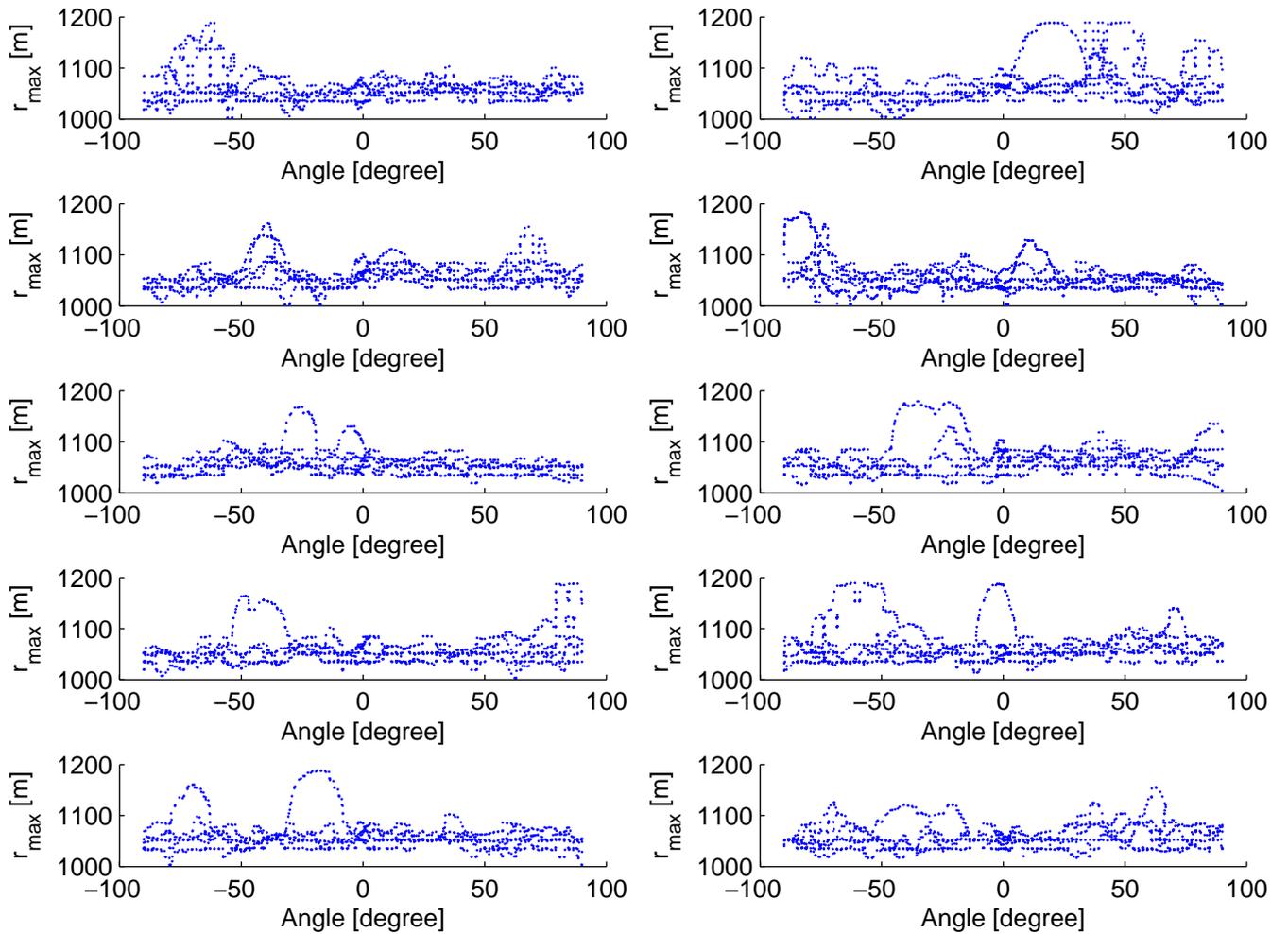


Figure E.1: The maximal lateral spread of CO₂ cloud after 10 years of injection: each panel contains 5 realizations

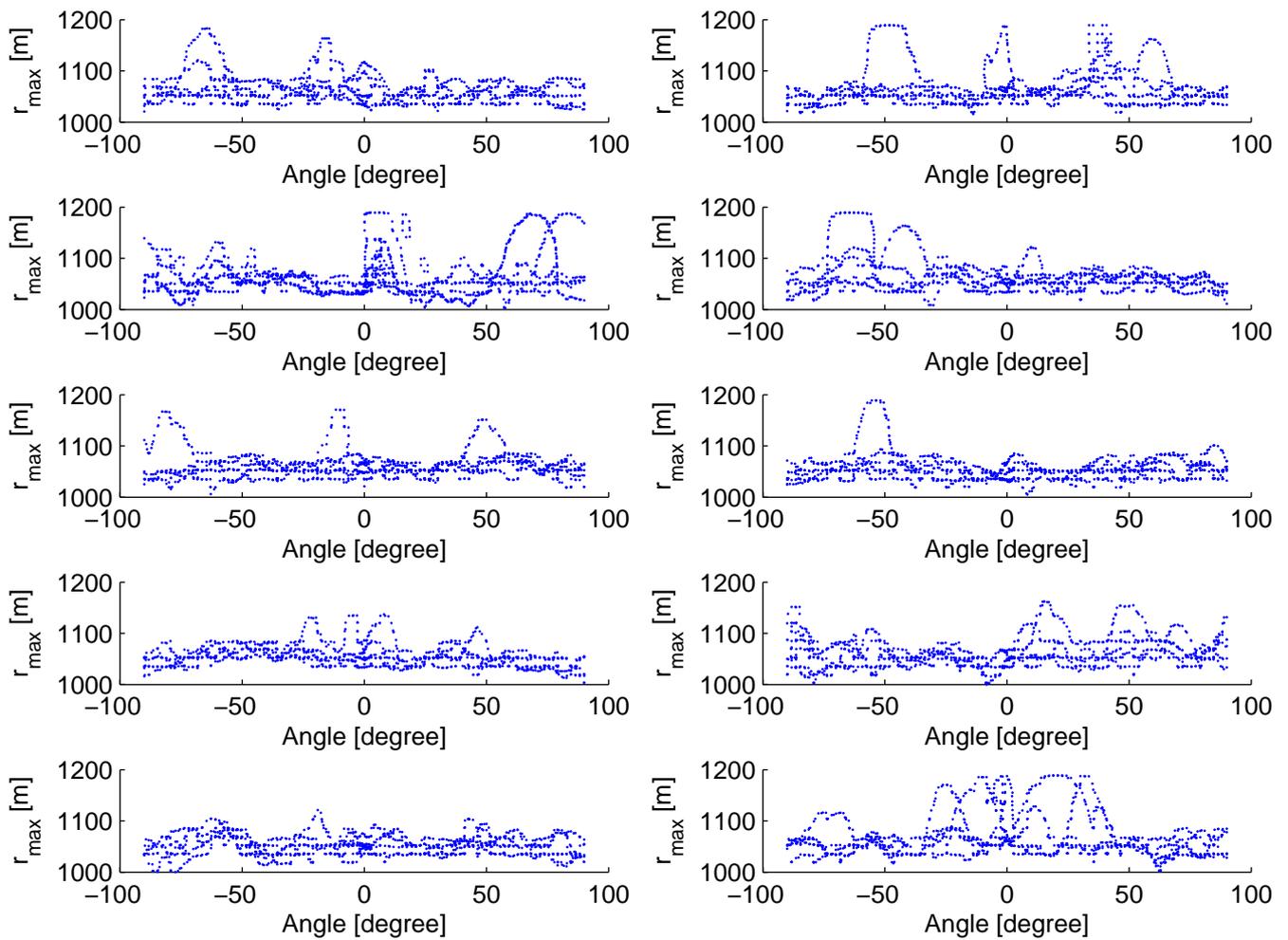


Figure E.2: The maximal lateral spread of CO₂ cloud after 10 years of injection: each panel contains 5 realizations

Appendix F

Details of the inference of the homogeneous equivalent permeability distribution

The inverse transform sampling (also called Smirnov transform (Smirnov, 1970)) is a method for generating random sample numbers given the cumulative distribution function. It consists in generating random samples from a uniform law in the interval (0,1): $u \sim U(0, 1)$ and interpreting the obtained values as probabilities. The quantiles for these probabilities are found by the inversion of the cumulative distribution function. The condition for defining the inverse distribution function (or quantile function) is that the CDF is a strictly increasing continuous function. Nevertheless, as any CDF is a non-decreasing right-continuous function, its inverse function can be defined as:

$$F^{-1}(u) = \inf\{x | F(x) \geq u\}.$$

Let us denote $F_{het}(r_{max})$ the cumulative distribution function of the maximal lateral spread when the uncertainty in the permeability values is attributed to the spatial variability (anisotropic random field described in 5.2.2 with a lognormal PDF ($\mu=-30.05$, $\sigma=0.47$), squared cardinal sine correlation structure, correlation lengths (50 m, 50 m, 20 m)). Let $\{u_m\}$ be a sampling of $u \sim U(0, 1)$, a number of $m = 10^5$ samples has been taken in the present study. The CCDF for the heterogeneous case has been evaluated for a set of thresholds r_0 , the cumulative distribution function $F_{het}(r_{max})$ is also discretized in $p - 1$ equal intervals of r_{max} with the endpoints $\{z_i\}$, $i = 1, \dots, p$ ($p = 201$ in the present study), giving a set of pairs $\{z_i, F_{het}^i\}$, $i = 1, \dots, p$. The values of a random variable z are found as values of a linear interpolation between points $\{F_{het}^i, z_i\}$ at points u :

$$\text{if } u_m \in [F_{het}^i, F_{het}^{i+1}], \text{ then } z_m = z_i + (z_{i+1} - z_i) \frac{u - F_{het}^i}{F_{het}^{i+1} - F_{het}^i}.$$

The same procedure of curve inversion through linear interpolation has been followed to obtain the values of $K_{int}(z_m)$ out of the meta-model sampling points (The interpolated function is presented in Figure F.1).

With the difference that the points of the meta-model sampling are random and not equally spaced. Let us denote $\{K_{int}^j, z_{meta-model}^j\}$, $j = 1, \dots$ the points of meta-model

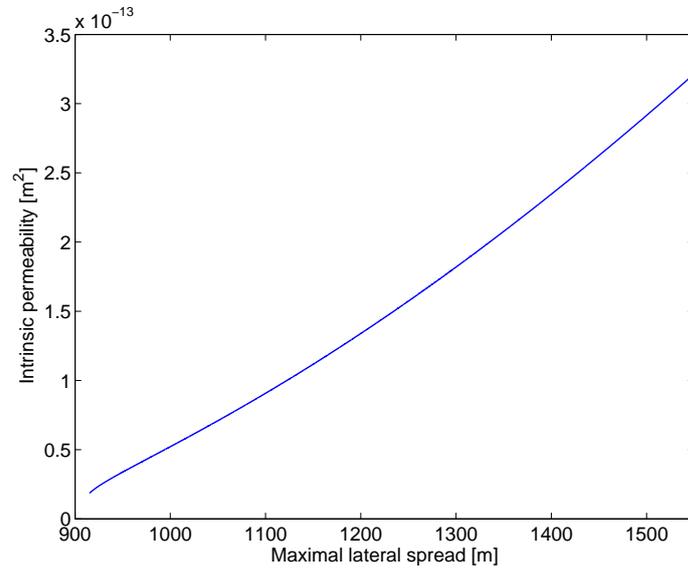


Figure F.1: The intrinsic permeability as a function of maximal radius of the CO₂ cloud: linear interpolation between the metamodel sampling points (Figure 4.7).

random Monte Carlo sampling. The values of K_{int} are found in the following way:

$$\text{if } z_m \in [z_{meta-model}^j, z_{meta-model}^{j+1}], \text{ then } K_{int} = K_{int}^j + (K_{int}^{j+1} - K_{int}^j) \frac{z_m - z_{meta-model}^j}{z_{meta-model}^{j+1} - z_{meta-model}^j}.$$

Bibliography

- Allaire, G. (1992). Homogenization and two-scale convergence. *SIAM J. Math. Anal.*, 23:1482–1518.
- Andre, L., Audigane, P., Azaroual, M., and Menjoz, A. (2007). Numerical modeling of fluid-rock chemical interactions at the supercritical CO₂ liquid interface during CO₂ injection into a carbonate reservoir, the Dogger aquifer (Paris Basin, France). *Energy Conversion and Management*, 48(6):1782 – 1797.
- Arbogast, T., Douglas, Jr., J., and Hornung, U. (1990). Derivation of the double porosity model of single phase flow via homogenization theory. *SIAM J. Math. Anal.*, 21:823–836.
- Ashraf, M., Oladyshkin, S., and Nowak, W. (2013). Geological storage of CO₂: Application, feasibility and efficiency of global sensitivity analysis and risk assessment using the arbitrary polynomial chaos. *International Journal of Greenhouse Gas Control*, 19(0):704 – 719.
- Bachu, S., Gunter, W., and Perkins, E. (1994). Aquifer disposal of CO₂: Hydrodynamic and mineral trapping. *Energy Conversion and Management*, 35(4):269 – 279.
- Baecher, G. and Christian, J. (2003). *Reliability and Statistics in Geotechnical Engineering*. John Wiley & Sons Ltd.
- Bear, J. (1972). *Dynamics of fluids in porous media*. American Elsevier, New York.
- Bear, J. (1979). *Hydraulics of groundwater*. McGraw-Hill International Book Co., London.
- Bear, J. (1987). Effective and relative permeabilities of anisotropic porous media. *Transport in porous media*, 2(3).
- Bennion, B. and Bachu, S. (2008). Drainage and imbibition relative permeability relationships for supercritical CO₂/brine and H₂S/brine systems in intergranular sandstone, carbonate, shale, and anhydrite rocks. *SPE Reservoir Evaluation & Engineering*, 11(03):487 – 496.
- Blatman, G. (2009). *Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis*. Ph.D Thesis, University Blaise Pascal, Clermont-Ferrand.

- Blatman, G. and Sudret, B. (2010). Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering and System Safety*, 95(11):1216 – 1229.
- Bouquet, S., Bruel, D., and de Fouquet, C. (2013). Influence of heterogeneities and up-scaling on CO₂ storage prediction at large scale in deep saline aquifer. *Energy Procedia*, 37(0):4445 – 4456. GHGT-11.
- Bourgeat, A., Mikelic, A., and Piatnitski, A. (1998). Modèle de double porosité aléatoire. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 327(1):99 – 104.
- Bowden, A. and Rigg, A. (2004). Assessing reservoir performance risk in CO₂ storage projects. In *Proceedings of the 7th International Conference on Greenhouse Gas Control Technologies (GHGT-7), 7-11 September 2004*, Vancouver, Canada.
- Box, G. E. P. and Draper, N. R. (1986). *Empirical Model-building and Response Surface*. John Wiley & Sons, Inc., New York, NY, USA.
- Brooks, R. and Corey, A. (1964). *Hydraulic Properties of Porous Media*. Colorado State University Hydrology Papers. Colorado State University.
- Cauvin, M. (2007). *Prise en compte des incertitudes et calcul des probabilités dans les études des risques liés au sol et au sous-sol*. Thèse de doctorat de l'Institut Nationale Polytechnique de Lorraine en Génie Civile - hydrosystèmes - Géotechnique. In French.
- Celia, M. A. and Nordbotten, J. M. (2009). Practical Modeling Approaches for Geological Storage of Carbon Dioxide. *Ground Water*, 47(5):627–638.
- Celia, M. A. and Nordbotten, J. M. (2011). How simple can we make models for CO₂ injection, migration, and leakage? *Energy Procedia*, 4(0):3857 – 3864. 10th International Conference on Greenhouse Gas Control Technologies.
- Chen, Z., Huan, G., and Ma, Y. (2006). *Computational Methods for Multiphase Flows in Porous Media*. SIAM Society for Industrial and Applied Mathematics, Philadelphia.
- Cherkaoui, A. and Lopez, P. (2009). CO₂ storage risk assessment : feasibility study of the systemic method MOSAR. In Guarascio, M., Brebbia, C., and Garzia, F., editors, *Safety and security engineering III*, volume 108 of *WIT Transactions on the Built Environment*, pages 173–184, Rome, Italie. WitPress. Southampton.
- Chorin, A. J. (1974). Gaussian fields and random flow. *Journal of Fluid Mechanics*, 63:21–32.
- Condor, J., Unatrakarn, D., Wilson, M., and Asghari, K. (2011). A comparative analysis of risk assessment methodologies for the geologic storage of carbon dioxide. *Energy Procedia*, 4(0):4036 – 4043. 10th International Conference on Greenhouse Gas Control Technologies.
- Cornet, F. H. and Buret, D. (1992). Stress field determinations in France by hydraulic tests in boreholes. *Journal of Geophysical Research: Solid Earth*, 97(B8):11829–11849.

- Crestaux, T., Maître, O. L., and Martinez, J. (2009). Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94:1161–1172.
- Croisé, J., Mayer, G., Marschall, P., Matray, J. M., Tanaka, T., and Vogel, P. (2006). Essais d'injection de gaz dans les argiles à $\frac{1}{2}$ opalinus du laboratoire souterrain du mont terri (suisse) : données et interprétation. *Oil & Gas Science and Technology - Rev. IFP*, 61(5):631–645.
- Crow, S. C. and Canavan, G. H. (1970). Relationship between a Wiener-Hermite expansion and an energy cascade. *Journal of Fluid Mechanics*, 41:387–403.
- CSLF (2009). Phase I Final Report from CSLF Risk Assessment Task Force. Technical Report CSLF-T-2009-04, Carbon sequestration leadership forum.
- Dake, L. (1978). *Fundamentals of Reservoir Engineering*. Development in Petroleum Science Series. Elsevier Scientific Publishing Company.
- Dalkhaa, C., Shevalier, M., Nightingale, M., and Mayer, B. (2013). 2-D reactive transport modeling of the fate of CO₂ injected into a saline aquifer in the Wabamun Lake Area, Alberta, Canada. *Applied Geochemistry*, 38(0):10 – 23.
- Darcy, H. (1856). *Les fontaines publiques de la ville de Dijon : exposition et application des principes à $\frac{1}{2}$ suivre et des formules à $\frac{1}{2}$ employer dans les questions de distribution d'eau*. V. Dalmont (Paris).
- Das, D. and Hassanizadeh, S. (2005). *Upscaling Multiphase Flow in Porous Media: From Pore to Core and Beyond*. Springer, New York.
- De Lucia, M., Lagneau, V., de Fouquet, C., and Bruno, R. (2011). The influence of spatial variability on 2D reactive transport simulations. *Comptes Rendus Geoscience*, 343(6):406 – 416.
- De Rocquigny, E. (2012). *Modelling Under Risk and Uncertainty: An Introduction to Statistical, Phenomenological and Computational Methods*. Wiley Series in Probability and Statistics. Wiley.
- De Rocquigny, E., Devictor, N., and Tarantola, S. (2008). *Uncertainty in industrial practice: a guide to quantitative uncertainty management*. Wiley.
- Desbarats, A. (1994). Spatial averaging of hydraulic conductivity under radial flow conditions. *Mathematical Geology*, 26(1):1–21.
- Díez, P., Egozcue, J. J., and Huerta, A. (1998). A posteriori error estimation for standard finite element analysis. *Computer Methods in Applied Mechanics and Engineering*, 163(1-4):141 – 157.
- DNV (2010). CO₂QUALSTORE - Guideline for selection and qualification of sites and projects for geological storage of CO₂. Technical Report No 2009-1425 77, Det Norske Veritas.

- Dodds, K., Waston, M., and Wright, I. (2011). Evaluation of risk assessment methodologies using the In Salah CO₂ storage project as a case history. *Energy Procedia*, 4:4162–4169.
- Dossantos-Uzarralde, P. and Guittet, A. (2008). A polynomial chaos approach for nuclear data uncertainties evaluations. *Nuclear Data Sheets*, 109(12):2894 – 2899. Special Issue on Workshop on Neutron Cross Section Covariances June 24-28, 2008, Port Jefferson, New York, USA.
- Douglas, J., Peaceman, D., and Rachford, H. (1959). A method for calculating multi-dimensional immiscible displacement. *Trans. SPE AIME*, 216:297–306.
- Durga Rao, K. (2009). Dynamic fault tree analysis using Monte Carlo simulation in probabilistic safety assessment. *Reliability engineering & system safety*.
- Durlofsky, L., Behrens, R., Jones, R., and Bernath, A. (1996). Scale up of heterogeneous three dimensional reservoir descriptions. *SPE Journal*.
- Durst, P. and Kervvan, C. (2011). Modélisation 3D couplé chimie-transport de l’injection de CO₂ à l’échelle d’un site pilote en aquifère salin profond (Dogger, bassin de Paris). Technical report, Report BRGM/RP-55667-FR. www.brgm.fr. In French.
- Eaton, T. (2006). On the importance of geological heterogeneity for flow simulation. *Sedimentary Geology*, 184(3-4):187 – 201. Heterogeneity in Sedimentary Aquifers: Challenges for Characterization and Flow Modeling, Geological Society of America Annual Meeting *GeoScience Horizons 2003*.
- Ewing, R. E. (1997). Aspects of upscaling in simulation of flow in porous media. *Advances in Water Resources*, 20(5-6):349 – 358.
- Fall, M., Nasir, O., and Nguyen, T. (2014). A coupled hydro-mechanical model for simulation of gas migration in host sedimentary rocks for nuclear waste repositories. *Engineering Geology*, 176(0):24 – 44.
- Fan, Y., Durlofsky, L. J., and Tchelepi, H. A. (2012). A fully-coupled flow-reactive-transport formulation based on element conservation, with application to CO₂ storage simulations. *Advances in Water Resources*, 42(0):47–61.
- Farret, R., Gombert, P., and Hulot, C. (2010). État de l’art et analyse des risques pour un stockage de CO₂ en aquifère salin. Technical report, INERIS. DRS-10-100887-12619A. In French.
- Farret, R. and Thoraval, A. (2012). Synthèse de l’état des connaissances sur les risques liés au stockage géologique du CO₂. rapport n°12 : les risques en phase de stockage à long terme. Technical report, INERIS. DRS-12-126009-01377A. In French.
- Farret, R. and Thoraval, A. (2013). Retour d’expérience des incidents et accidents sur des sites d’exploitation ou de stockage en milieu souterrain - application au stockage géologique du CO₂. Technical report, INERIS. DRS-12-126009-13886A. In French.

- Fenghour, A., Wakeham, W., and Vesovic, V. (1998). The viscosity of carbon dioxide. *Journal of physical and chemical reference data*, 27:31–44.
- Field, C., Barros, V., Stocker, T., and Dahe, Q. (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Formaggia, L., Guadagnini, A., Imperiali, I., Lever, V., Porta, G., Riva, M., Scotti, A., and Tamellini, L. (2012). Global sensitivity analysis through polynomial chaos expansion of a basin-scale geochemical compaction model. *Computational Geosciences*, pages 1–18. 10.1007/s10596-012-9311-5.
- Franzetti, S., Guadagnini, A., and Orsi, E. (1996). Monte carlo simulation and effective conductivity in confined radial flow fields. *Calibration and Reliability in Groundwater modelling*, (237):463–471.
- Fučík, R., Mikyška, J., Beneš, M., and Illangasekare, T. H. (2007). An improved semi-analytical solution for verification of numerical models of two-phase flow in porous media. *Vadose Zone Journal*, 6:93–104.
- Gasda, S. E., Celia, M. A., and Nordbotten, J. M. (2008). Upslope plume migration and implications for geological CO₂ sequestration in deep, saline aquifers. *The IES Journal Part A: Civil & Structural Engineering*, 1(1):2–16.
- Gerstenberger, M., Nicol, A., Stenhouse, M., Berryman, K., Stirling, M., Webb, T., and Smith, W. (2009). Modularised logic tree risk assessment method for carbon capture and storage projects. *Energy Procedia*, 1(1):2495–2502. Proceedings of the 9th International Conference on Greenhouse Gas Control Technologies (GHGT-9), 16-20 November 2008, Washington DC, {USA}.
- Ghanem, R. and Spanos, P. (1991). *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York.
- Gombert, P. and Farret, R. (2009). Adaptation des outils d’analyse de risques aux futurs complexes de stockage géologique de CO₂. *Tunnels et ouvrages souterrains*, 213:142–154. In French.
- Gough, C. and Shackley, S. (2006). Towards a multi-criteria methodology for assessment of geological carbon storage options. *Climatic Change*, 74(1-3):141–174.
- Grinstead, C. and Snell, J. (1997). *Introduction to Probability*. 2nd ed. American Mathematical Society.
- Grobe, M., Pashin, J. C., Dodge, R. L., and of Texas at Austin. Bureau of Economic Geology, U. (2009). *Carbon Dioxide Sequestration in Geological Media: State of the Science*. AAPG studies in geology. American Association of Petroleum Geologists.
- Hamed, M. M., Bedient, P. B., and Dawson, C. N. (1996). Probabilistic modeling of aquifer heterogeneity using reliability methods. *Advances in Water Resources*.

- Helton, J. and Davis, F. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23 – 69.
- Helton, J., Laboratories, S. N., and of Energy, U. S. D. (1996). *Probability, Conditional Probability, and Complementary Cumulative Distribution Functions in Performance Assessment for Radioactive Waste Disposal*. Sandia National Laboratories.
- Helton, J. C., Davis, F. J., and Johnson, J. D. (2005). A comparison of uncertainty and sensitivity analysis results obtained with random and latin hypercube sampling. *Rel. Eng. & Sys. Safety*, 89(3):305–330.
- Hudson, J., Stephansson, O., and Andersson, J. (2005). Guidance on numerical modelling of thermo-hydro-mechanical coupled processes for performance assessment of radioactive waste repositories. *International Journal of Rock Mechanics and Mining Sciences*, 42(5-6):850 – 870. Research results from the Decovalex III Benchpar projects.
- IEA GHG (2007). Environmental assessment for CO₂ capture and storage. Technical report, International Energy Agency Greenhouse Gas R&D Programme.
- IPCC (2005). *Carbon Dioxide Capture and Storage: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- ISO 31000:2009 (2009). *Risk management- Principles and guidelines*. International Organization for Standardization.
- ISO Guide 73:2009 (2009). *Risk management - Vocabulary*. International Organization for Standardization.
- Jakobsen, J. P., Roussanaly, S., Mjølhus, M. J., and Tangen, G. (2013). A standardized approach to multi-criteria assessment of ccs chains. *Energy Procedia*, 37(0):2765 – 2774. GHGT-11.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.
- Jensen, J. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193.
- Jikov, V., Kozlov, S., and Oleinik, O. (1994). *Homogenization of Differential Operators and Integral Functionals*. Springer-Verlag.
- Joint Committee for Guides in Metrology (2008). JCGM 100: Evaluation of measurement data - guide to the expression of uncertainty in measurement. Technical report, JCGM.
- Juntunen, M. and Wheeler, M. F. (2012). Two-phase flow in complicated geometries: Modeling the Frio data using improved computational meshes. *Computational Geosciences*.

- Kaplan, S. and Garrick, B. J. (1981). On the quantitative definition of risk. *Risk Analysis*, 1(1):11–27.
- Keilegavlen, E., Nordbotten, J., and Stephansen, A. (2012). Tensor relative permeabilities: Origins, modeling and numerical discretization. *International Journal of Numerical Analysis and Modeling*, 9(3):701–724.
- Kendall, M. and Stuart, A. (1977). *The Advanced Theory of Statistics*, volume 1 of *Distribution Theory*. Charles Griffin & Company Limited, London, Great Britain, 4 edition.
- Kharaka, Y. K., Thordsen, J. J., Hovorka, S. D., Nance, H. S., Cole, D. R., Phelps, T. J., and Knauss, K. G. (2009). Potential environmental issues of CO₂ storage in deep saline aquifers: Geochemical results from the Frio-I Brine Pilot test, Texas, USA. *Applied Geochemistry*, 24(6):1106 – 1112.
- Kirchsteiger, C. (1999). On the use of probabilistic and deterministic methods in risk analysis. *Journal of Loss Prevention in the Process Industries*, 12(5):399 – 419.
- Kirste, D. (2013). Mineral saturation state and the implications for reaction rates in reactive transport modeling of CO₂ storage. *Procedia Earth and Planetary Science*, 7(0):432 – 435. Proceedings of the Fourteenth International Symposium on Water-Rock Interaction, WRI 14.
- Kleijnen, J. P. and Sargent, R. G. (2000). A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, 120(1):14 – 29.
- Korre, A. and Durucan, S. (2009). A review of the international state of the art in risk assessment guidelines and proposed terminology for use in CO₂ geological storage. Technical report, IEA Greenhouse gas R&D Programme (IEA GHG).
- Lahaie, F., Farret, R., and Bumb, P. (2009). Towards a framework for CCS risk assessment. *Carbon Capture Journal*, 12:22–24.
- Lake, L. and Carroll, H. (1986). *Reservoir Characterization*. Number 1 in Reservoir Characterization. Academic Press.
- Lake, L., Fanchi, J., F, M., Arnold, K., Clegg, J., Holstein, E., and Warner, H. (2007). *Petroleum Engineering Handbook: Reservoir engineering and petrophysics*. Petroleum Engineering Handbook. Society Of Petroleum Engineers.
- Lake, L. W., Carroll, H. B. J., and Wesson, T. C., editors (1991). *Reservoir Characterization II Z*. Elsevier Science.
- Le Gallo, Y. (2009). Post-closure migration for CO₂ geological storage and regional pressure inferences. *Energy Procedia*, 1(1):3259 – 3266. Proceedings of the 9th International Conference on Greenhouse Gas Control Technologies (GHGT-9), 16-20 November 2008, Washington DC, USA.

- Le Guen, Y., Huot, M., Loizzo, M., and Poupard, O. (2011). Well integrity risk assessment of Ketzin injection well (ktzi-201) over a prolonged sequestration period. *Energy Procedia*, 4(0):4076 – 4083. 10th International Conference on Greenhouse Gas Control Technologies.
- Le Guen, Y., Le Gouevéc, J., Chammas, R., Gerard, B., Poupard, O., Van Der Beken, A., and Jammes, L. (2008). CO₂ storage - managing the risk associated with well leakage over long time scales. *SPE paper No 116424-PP*. OXAND S.A. and Schlumberger Carbon services.
- Le Maître, O. and Knio, O. (2010). *Spectral methods for uncertainty quantification*. Scientific Computation. Springer, New York. With applications to computational fluid dynamics.
- Le Neveu, D. (2008). CQUESTRA, a risk and performance assessment code for geological sequestration of carbon dioxide. *Energy Conversion and Management*, 49(1):32 – 46.
- Lehmann, E. and Casella, G. (1998). *Theory of Point Estimation*. Springer Texts in Statistics. Springer.
- Leverett, M. (1941). Capillary behaviour in porous solids. *Transactions of the AIME*, 142:159–172.
- Liu, F., Lu, P., Zhu, C., and Xiao, Y. (2011). Coupled reactive flow and transport modeling of CO₂ sequestration in the Mt.Simon sandstone formation, Midwest U.S.A. *International Journal of Greenhouse Gas Control*, 5(2):294 – 307.
- MANAUS (2011). Definition of reference risk scenarios for a CO₂ storage complex. Technical report, INERIS. Convention N 1094C0003 Ademe-GEOGREEN. In French.
- Marle, C. M. (2006). Henry Darcy et les écoulements de fluides en milieu poreux. *Oil & Gas Science and Technology - Rev. IFP*, 61(5):599–609. In French.
- Marrel, A., Iooss, B., Laurent, B., and Roustant, O. (2009). Calculations of sobol indices for the gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3):742 – 751.
- Marrel, A., Iooss, B., Van Dorpe, F., and Volkova, E. (2008). An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics and Data Analysis*, 52:4731–4744.
- Mathias, S. A., Hardisty, P. E., Trudell, M. R., and Zimmerman, R. W. (2009). Approximate solutions for pressure buildup during CO₂ injection in brine aquifers. *Transport in Porous Media*, 79(2):265–284.
- Meecham, W. C. and Jeng, D.-T. (1968). Use of the Wiener-Hermite expansion for nearly normal turbulence. *Journal of Fluid Mechanics*, 32:225–249.
- Meecham, W. C. and Siegel, A. (1964). Wiener-Hermite Expansion in Model Turbulence at Large Reynolds Numbers. *Physics of Fluids*, 7:1178–1190.

- Metcalfe, R., Paulley, A., Suckling, P., and Watson, C. (2013). A tool for integrating and communicating performance-relevant information in CO₂ storage projects: description and application to in salah. *Energy Procedia*, 37:4741–4748.
- Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *Journal of the American statistical Association*, 44(247):335–341.
- Meyer, V., Houdu, E., Poupard, O., and Gouvec, J. L. (2009). Quantitative risk evaluation related to long term CO₂ gas leakage along wells. *Energy Procedia*, 1(1):3595–3602. Proceedings of the 9th International Conference on Greenhouse Gas Control Technologies (GHGT-9), 16-20 November 2008, Washington DC, USA.
- Millard, A., Durin, M., Stietel, A., Thoraval, A., Vuillod, E., Baroudi, H., Plas, F., Bougnoux, A., Vouille, G., Kobayashi, A., Hara, K., Fujita, T., and Ohnishi, Y. (1995). Discrete and continuum approaches to simulate the thermo-hydro-mechanical couplings in a large, fractured rock mass,. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 32(5):409–434.
- Mohd Amin, S., Weiss, D., and Blunt, M. (2014). Reactive transport modelling of geologic CO₂ sequestration in saline aquifers: The influence of pure CO₂ and of mixtures of CO₂ with CH₄ on the sealing capacity of cap rock at 37 °C and 100 bar. *Chemical Geology*, 367(0):39 – 50.
- MoMas (2012).
- Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, 12(3):513–522.
- Nepveu, M., Yavuz, F., and David, P. (2009). FEP analysis and Markov chains. *Energy Procedia*, 1(1):2519 – 2523. Proceedings of the 9th International Conference on Greenhouse Gas Control Technologies (GHGT-9), 16-20 November 2008, Washington DC, USA.
- Nikolaevskij, V. (1990). *Mechanics of Porous and Fractured Media*. Series in theoretical and applied mechanics. World Scientific.
- Noetinger, B. and Haas, A. (1996). Permeability averaging for well tests in 3D stochastic reservoir models. *SPE Journal*. SPE 36653.
- Nordbotten, J. M. and Celia, M. A. (2006). Similarity solutions for fluid injection into confined aquifers. *Journal of Fluid Mechanics*, 561:307–327.
- Nordbotten, J. M., Celia, M. A., and Bachu, S. (2005). Injection and storage of CO₂ in deep saline aquifers: Analytical solution for CO₂ plume evolution during injection. *Transport in Porous Media*, 58:339–360. 10.1007/s11242-004-0670-9.
- Okhulkova, T., Clouteau, D., Cottureau, R., and Farret, R. (2014). Probabilistic modeling and simulation of CO₂ storage in heterogeneous geological formations. In *Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures*. CRC Press.

- Okwen, R., Stewart, M., and Cunningham, J. (2011). Analytical model for screening potential CO₂ repositories. *Computational Geosciences*, 15(4):755–770.
- Oladyshkin, S., Class, H., Helmig, R., and Nowak, W. (2011). An integrative approach to robust design and probabilistic risk assessment for CO₂ storage in geological formations. *Computational Geosciences*, 15(3):565–577.
- Oldenburg, C. (2005). HSE screening risk assessment for geologic CO₂ storage sites. In *Fourth Annual Conference on Carbon Capture and Sequestration*.
- Oldenburg, C. (2007). Screening and ranking framework for geologic CO₂ storage site selection on the basis of health, safety, and environmental risk. *Environmental Geology*.
- Oldenburg, C. M. (2009). *Certification Framework Based on Effective Trapping for Geologic Carbon Sequestration*. Lawrence Berkeley National Laboratory.
- Olla, S. (1994). *Homogenization of Diffusion Processes in Random fields*. C.M.A.P. - École Polytechnique, Palaiseau. Lecture Notes.
- OpenTURNS (2012).
- Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263:641–646.
- Orszag, S. A. and Bissonnette, L. R. (1967). Dynamical Properties of Truncated Wiener-Hermite Expansions. *Physics of Fluids*, 10:2603–2613.
- Panfilov, M. (2000). *Macroscale models of flow through highly heterogeneous porous media*. Theory and applications of transport in porous media. Kluwer Academic Publ., Dordrecht, Boston, London.
- Paté-Cornell, M. (1996). Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety*, 54(2-3):95 – 111. Treatment of Aleatory and Epistemic Uncertainty.
- Paulley, A., Metcalfe, R., and Egan, M. (2012). Geological disposal programme design and prioritization in the face of uncertainty: use of structured evidence support logic techniques. *Mineralogical Magazine*, 76(8):3497–3507.
- Paulley, A., Metcalfe, R., and Limer, L. (2011). Systematic FEP and scenario analysis to provide a framework for assessing long-term performance of the krechba CO₂ storage system at in salah. *Energy Procedia*, 4(0):4185 – 4192. 10th International Conference on Greenhouse Gas Control Technologies.
- Perrin, J.-C., Krause, M., Kuo, C.-W., Miljkovic, L., Charoba, E., and Benson, S. M. (2009). Core-scale experimental study of relative permeability properties of CO₂ and brine in reservoir rocks. *Energy Procedia*, 1(1):3515 – 3522. Proceedings of the 9th International Conference on Greenhouse Gas Control Technologies (GHGT-9), 16-20 November 2008, Washington DC, USA.

- Refsgaard, J. C., van der Sluijs, J. P., Brown, J., and van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, 29(11):1586 – 1597.
- Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., and Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process - a framework and guidance. *Environmental Modelling & Software*, 22(11):1543 – 1556.
- Rehfeldt, K. R., Boggs, J. M., and Gelhar, L. W. (1992). Field study of dispersion in a heterogeneous aquifer: 3-d geostatistical analysis of hydraulic conductivity. *Water Resources Research*, 28(12):3309–3324.
- Rhenals Garrido, D. R., Lafortune, S., Souli, H., and Dubujet, P. (2013). Impact of supercritical CO₂/water interaction on the caprock nanoporous structure. *Procedia Earth and Planetary Science*, 7:738 – 741.
- Rillard, J., Gombert, P., Toulhoat, P., and Zuddas, P. (2013). Kinetic rate of iron release during artificial CO₂ injection in a shallow aquifer: Preliminary results. *Procedia Earth and Planetary Science*, 7(0):742 – 745. Proceedings of the Fourteenth International Symposium on Water-Rock Interaction, WRI 14.
- Rish, W. (2005). A probabilistic risk assessment of class I hazardous waste injection wells. In Tsang, C. and Apps, J., editors, *Underground Injection Science and Technology*, Developments in Water Science, chapter 10, pages 93–135. Elsevier Science. Hull and Associates, Inc., Dublin, OH, USA.
- Roger and Ghanem (1998). Probabilistic characterization of transport in heterogeneous media. *Computer Methods in Applied Mechanics and Engineering*, 158(34):199 – 220.
- Rohmer, J. and Seyedi, D. (2010). Coupled large scale hydromechanical modelling for caprock failure risk assessment of CO₂ storage in deep saline aquifers. *Oil Gas Sci. Technol. Rev.IFP*, 65(3):503–517.
- Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition.
- Rutqvist, J. (2012). The geomechanics of CO₂ storage in deep sedimentary formations. *Geotechnical and Geological Engineering*, 30(3):525–551.
- Rutqvist, J., Birkholzer, J., and Tsang, C.-F. (2008). Coupled reservoir-geomechanical analysis of the potential for tensile and shear failure associated with CO₂ injection in multilayered reservoir-caprock systems. *International Journal of Rock Mechanics and Mining Sciences*, 45(2):132 – 143.
- Rutqvist J., Birkholzer J.T., C. F. e. T. C.-F. (2007). Estimating maximum sustainable injection pressure during geological sequestration of CO₂ using coupled fluid flow and geomechanical fault-slip analysis. *Energy Conversion and Management*, 48(6):1798–1807.

- Saffman, P. and Taylor, G. (1958). The penetration of a fluid into a porous medium of Hele-Shaw cell containing a more viscous liquid. *Proceedings of the Royal Society of London, Series A. Mathematical and Physical Sciences*, 245(1242 (Jun. 24, 1958)):312–329.
- Saint-Geours, N. (2012). *Sensitivity analysis of spatial models: application to cost-benefit analysis of flood risk management plans*. PhD thesis, Montpellier 2 Biostatistique.
- Saint-Geours, N., Lavergne, C., Bailly, J.-S., and Grelot, F. (2012). Change of support in spatial variance-based sensitivity analysis. *Mathematical Geosciences*, 44(8):945–958.
- Samadi, J. (2012). *Développement d'une approche systémique de management des risques pour les projets de captage, transport et stockage de CO₂*. PhD Thesis, ENMP, Paris.
- Sankararaman, S. and Mahadevan, S. (2013). Separating the contributions of variability and parameter uncertainty in probability distributions. *Reliability Engineering & System Safety*, 112(0):187 – 199.
- Savage, D., Maul, P., Benbow, S., and Walke, R. (2004). A generic FEP database for the assessment of long-term performance and safety of the geological storage of CO₂. Technical report, Quintessa QRS-1060A-1. (www.co2captureandstorage.info/docs/QuintessaReportIEA.pdf).
- Scalabrin, G., Marchi, P., Finezzo, F., and Span, R. (2006). A Reference Multiparameter Thermal Conductivity Equation for Carbon Dioxide with an Optimized Functional Form. *Journal of Physical and Chemical Reference Data*, 35(4):1549–1575.
- Sergienko, E., Lemaître, P., Arnaud, A., Busby, D., and Gamboa, F. (2013). Reliability sensitivity analysis based on probability distribution perturbation with application to CO₂ storage. *Structural Safety*.
- Shannon, C. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, 623–656.
- Shen, J., Dangla, P., and Thiery, M. (2013). Reactive transport modeling of CO₂ through cementitious materials under CO₂ geological storage conditions. *International Journal of Greenhouse Gas Control*, 18(0):75 – 87.
- Shi, J.-Q., Smith, J., Durucan, S., and Korre, A. (2013). A coupled reservoir simulation-geomechanical modelling study of the CO₂ injection-induced ground surface uplift observed at Krechba, in Salah. *Energy Procedia*, 37(0):3719 – 3726. GHGT-11.
- Siegel, A., Imamura, T., and Meecham, W. C. (1965). Wiener-Hermite Expansion in Model Turbulence in the Late Decay Stage. *Journal of Mathematical Physics*, 6:707–721.
- Smirnov, N. V. (1970). *Theory of probability and mathematical statistics: Selected works*. Nauka, Moscow. In Russian.

- Smith, D., Noy, D., Holloway, S., and Chadwick, R. (2011). The impact of boundary conditions on CO₂ storage capacity estimation in aquifers. *Energy Procedia*, 4(0):4828 – 4834. 10th International Conference on Greenhouse Gas Control Technologies.
- Sobol, I. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Modelinf comput. Experiment*, 1(4):407–414.
- Soize, C. (2005). Random matrix theory for modeling uncertainties in computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, 194(12-16):1333–1366.
- Soize, C. and Ghanem, R. (2005). Physical systems with random uncertainties: Chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput.*, 26(2):395–410.
- Solomon, S., Kristiansen, B., Stangeland, A., Torp, T. A., and Kørstad, O. (2007). A proposal of regulatory framework for carbon dioxide storage in geological formations. *Bellona*. Prepared for International Risk Governance Council Workshop, March 15-16 2007, Washington,DC.
- Souley, M. and Thoraval, A. (2011). Nonlinear mechanical and poromechanical analyses : comparison with analytical solutions. In *COMSOL Conference 2011, 26-28 October 2011*.
- Stauffer, P. H., Viswanathan, H. S., Pawar, R. J., and Guthrie, G. D. (2009). A system model for geologic sequestration of carbon dioxide. *Environmental Science & Technology*, 43:565–570.
- Stenhouse, M., Zhou, W., Kozak, M., and Wilson, M. (2006). Risk Assessment and Geological Storage of CO₂ Briefing Document. Technical report, Monitor Scientific Report MSC1-2512.
- Stenhouse, M., Zhou, W., Savage, D., and Benbow, S. (2005). Framework methodology for long-term assessment of the fate of CO₂ in the Weyburn Field. In *Carbon Dioxide Capture for Storage in Deep Geologic Formations - Results from the CO₂ Capture Project, Volume 2*, pages 1251–1262.
- Sudret, B. (2008a). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979.
- Sudret, B. (2008b). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979.
- Sudret, B. and Der Kiureghian, A. (2000). Stochastic finite elements and reliability: a state-of-the-art report. Technical report, University of California, Berkeley. Technical report no. UCB/SEMM-2000/08.
- Ta, Q. A., Clouteau, D., and Cottureau, R. (2010). Modeling of random anisotropic elastic media and impact on wave propagation. *European Journal of Computational Mechanics*, 19(1-3):241–253.

- Thoraval, A. (2008). Evaluation de l'impact d'un stockage de CO₂ dans un aquifère du bassin parisien sur l'intégrité mécanique des formations hôtes et du recouvrement, une contribution de l'INERIS au programme ANR GEOCARBONE PICOREF. In *Séance technique CFMR "Géomécanique et stockage de CO₂", 23 octobre 2008, Paris. In French.*
- Thoraval, A., Guglielmi, Y., and Cappa, F. (2012). 3D simulations of an injection test done into an unsaturated porous and fractured limestone. In *COMSOL Conference, 10-12 October 2012.*
- Tixier, J., Dusserre, G., Salvi, O., and Gaston, D. (2002). Review of 62 risk analysis methodologies of industrial plants. *Journal of Loss Prevention in the Process Industries*, 15(4):291 – 303.
- Tong, F., Niemi, A., Yang, Z., Fagerlund, F., Licha, T., and Sauter, M. (2013). A numerical model of tracer transport in a non-isothermal two-phase flow system for CO₂ geological storage characterization. *Transport in Porous Media*, 98(1):173–192.
- U.S.EPA (2008). Vulnerability evaluation framework for geologic sequestration of carbon dioxide. Technical report, U.S. Environmental Protection Agency. EPA430-R-08-009.
- Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, 44(5):892–898.
- Vendrig, M., Spouge, J., and Bird, A. (2003). *Risk Analysis of the Geological Sequestration of Carbon Dioxide*. Report (Great Britain. Department of Trade and Industry. Cleaner Coal Technology Programme). Department of Trade and Industry. Great Britain. Cleaner Coal Technology Programme.
- Vidal-Gilbert, S., Nauroy, J.-F., and Brosse, E. (2009). 3D geomechanical modelling for CO₂ geologic storage in the Dogger carbonates of the Paris Basin. *International Journal of Greenhouse Gas Control*, 3(3):288 – 299.
- Vidal-Gilbert, S. and Thoraval, A. (2007). Modelling the mechanical impact of CO₂ injection into a carbonate reservoir of the Paris basin. In *Abstracts book of the 1st French-German symposium on geological storage of CO₂ , 21-22 June 2007.*
- Vilarrasa, V., Bolster, D., Olivella, S., and Carrera, J. (2010). Coupled hydromechanical modeling of CO₂ sequestration in deep saline aquifers. *International Journal of Greenhouse Gas Control*, 4(6):910 – 919. CO₂ Storage at the EGU General Assembly 2009.
- Viswanathan, H. S., Pawar, R. J., Stauffer, P. H., Kaszuba, J. P., Carey, J. W., Olsen, S. C., Keating, G. N., Kavetski, D., and Guthrie, G. D. (2008). Development of a hybrid process and system model for the assessment of wellbore leakage at a geologic CO₂ sequestration site. *Environmental Science & Technology*, 42(19):7280–7286.

- Walker, W., Harremoes, P., Romans, J., van der Sluus, J., van Asselt, M., Janssen, P., and Krauss, M. (2003). Defining uncertainty. a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1):5–17.
- Warren, J. and Price, H. (1961). Flow in heterogeneous porous media. *SPE Journal*. SPE 1579-G.
- Wassermann, S., Schulz, M., and Scheer, D. (2011). Linking public acceptance with expert knowledge on CO₂ storage: Outcomes of a delphi approach. *Energy Procedia*, 4(0):6353–6359. 10th International Conference on Greenhouse Gas Control Technologies.
- WHO (2004). *IPCS risk assessment terminology*. World Health Organization. International Programme on Chemical Safety.
- Wiener, N. (1938). The Homogeneous Chaos. *American Journal of Mathematics*, 60(4):897–936.
- Wiener, N. (1962). *Nonlinear Problems in Random Theory*. MIT.
- Wilday, J. and Farret, R. (2010). Erra a2: contribution to the risk assessment framework in inttegrisk project. Technical report, INERIS.
- Wildenborg, T., Chadwick, A., Deflandre, J.-P., Eiken, O., Mathieson, A., Metcalfe, R., Hattenberger, C. S., and Wollenweber, J. (2013). Key messages from active CO₂ storage sites. *Energy Procedia*, 37(0):6317 – 6325. GHGT-11.
- Wildenborg, T., Leijnse, T., Kreft, E., Nepveu, M., and Obdam, A. (2004). Long term safety assessment of CO₂ storage: The scenario approach. In *Seventh International Conference on Greenhouse Gas Control Technologies (GHGT-7) 7-11 September 2004*.
- Wildenborg, T., Leijnse, T., Kreft, E., Nepveu, M., Obdam, A., Wipfler, L., van der Grift, B., Hofstee, C., van Kesteren, W., Gaus, I., Czernichowski-Lauriol, I., Torfs, P., Wojcik, R., and Orlic, B. (2005). Risk assessment methodology for CO₂ storage: The scenario approach. In *The CO₂ Capture and Storage project (CCP)*, volume II, pages 1283–1305.
- Wollenweber, J., Busby, D., Wessel-Berg, D., Nepveu, M., Codreanu, D. B., Grimstad, A.-A., Sijacic, D., Maurand, N., Lothe, A., Wahl, F., Polak, S., Boot, H., Gril¹/₂ver, A., and Wildenborg, T. (2013). Integrated carbon risk assessment (icaras). *Energy Procedia*, 37(0):4825 – 4832. GHGT-11.
- Wood, B. D. (2009). The role of scaling laws in upscaling. *Advances in Water Resources*, 32(5):723 – 736. Dispersion in Porous Media.
- www.carbon-dioxide-properties.com (2010).
- Wyatt, D., Bowden, A., and Simms, M. (2009). Assessing and quantifying risks in geological carbon sequestration projects using RISQUE methodology with a modified Delphi approach to reduce uncertainty. In *AAPG/SEG/SPE Hedberg Conference, Geological Carbon Sequestration: Prediction and Verification, August 16-19 2009*, Vancouver, BC, Canada.

- Xiu, D. and Karniadakis, G. (2002). The wiener-askey polynomial chaos for stochastic differential equations. *Journal of Scientific Computing*, 26.
- Xiu, D. and Karniadakis, G. E. (2003). Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, 187(1):137 – 167.
- Xiu, D., Lucor, D., Su, C.-H., and Karniadakis, G. E. (2002). Stochastic modeling of flow-structure interactions using generalized polynomial chaos. *Journal of Fluids Engineering*, 124(1):51–59.
- Xu, T. and Li, J. (2013). Reactive transport modeling to address the issue of CO₂ geological sequestration. *Procedia Earth and Planetary Science*, 7(0):912 – 915. Proceedings of the Fourteenth International Symposium on Water-Rock Interaction, WRI 14.
- Yamamoto, S., Miyoshi, S., Sato, S., and Suzuki, K. (2013). Study on geomechanical stability of the aquifer -caprock system during CO₂ sequestration by coupled hydromechanical modelling. *Energy Procedia*, 37(0):3989 – 3996. GHGT-11.
- Yavuz, F., van Tilburg, T., David, P., Spruijt, M., and Wildenborg, T. (2009). Second generation CO₂ FEP analysis: CASSIF - carbon storage scenario identification framework. *Energy Procedia*, 1(1):2479 – 2485. Greenhouse Gas Control Technologies 9 Proceedings of the 9th International Conference on Greenhouse Gas Control Technologies (GHGT-9), 16-20 November 2008, Washington DC, USA.
- Zheng, L., Apps, J. A., Zhang, Y., Xu, T., and Birkholzer, J. T. (2009). Reactive transport simulations to study groundwater quality changes in response to CO₂ leakage from deep geological storage. *Energy Procedia*, 1(1):1887 – 1894. Proceedings of the 9th International Conference on Greenhouse Gas Control Technologies (GHGT-9), 16-20 November 2008, Washington DC, USA.
- Zhou, Q., Birkholzer, J. T., Tsang, C.-F., and Rutqvist, J. (2008). A method for quick assessment of CO₂ storage capacity in closed and semi-closed saline formations. *International Journal of Greenhouse Gas Control*, 2(4):626 – 639. TCCS-4: The 4th Trondheim Conference on CO₂ Capture, Transport and Storage.