POLITECNICO DI MILANO

Scuola di Ingegneria Industriale

Corso di Laurea Specialistica in Ingegneria Energetica, Ind. Oil and Gas Production

An Approach to Model Reduction and Sensitivity Analysis for Preliminary Risk Assessment of groundwater bodies associated with Hydraulic Fracturing.

Relatore: Prof. Alberto Guadagnini

Correlatore: Prof. Monica Riva, Aronne dell' Oca

Tesi di laurea di:
Giovanni Calanchi
Matricola 838136

Anno accademico 2015/2016

ACKNOWLEDGEMENTES

To my family, supporting and standing always behind me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

KEYWORDS: Global Sensitivity Analysis; Polynomial Chaos Expansion; Sobol Indices; Morris Indices; Probabilities Density Functions.

The content of this thesis is mainly concerned with the exploration of two different techniques of Global Sensitivity Analysis (GSA) group in a context of gas intrusion coming from a reservoir into an aquifer due to fracking operations. The process of intrusion has been described by two target quantities defined for the system under analysis. Two GSA techniques are represented by Sobol and Morris indices and indices principally state the level of influence of input parameters of the problem on the output target quantities. Sobol indices are easily computed through the use of a surrogate model of the full model built with the Polynomial Chaos Expansion (PCE) technique. The surrogate model also allows to achieve a Monte Carlo Analysis of global target quantities, defining their probability density functions and as a consequence giving some information about response surfaces.

PAROLE CHIAVE: Analisi di Sensitività Globale; Espansione in Caos Polinomiale; Indici di Sobol; Indici di Morris; funzioni di distribuzione di probabilità.

Il contenuto di questa tesi riguarda principalmente l'esplorazione e l'utilizzo di due differenti tecniche di Analisi di Sensitività Globale nel contesto di un'intrusione di un gas proveniente da un giacimento e entrante in un acquifero a cause di operazioni di fracking. Il processo di intrusione è stato descritto da due quantità di interesse definite per il sistema sotto analisi. Le due tecniche di sensitività sono rappresentate dagli indici di Sobol e Morris e gli indici determinano principalmente il livello di influenza dei parametri di ingresso del problema sulle quantità di interesse in uscita. Gli indici di

Sobol sono facilmente calcolati attraverso l'utilizzo di un modello surrogato del modello completo costruito con la espansione di tipo Polynomial Chaos Expansion. Il modello surrogato permette inoltre di ottenere un'analisi Monte Carlo delle quantità globali di interesse, definendo le loro funzioni di distribuzione di probabilità e di conseguenza fornendo informazioni riguardo alle superfici di risposta.

INTRODUCTION

The study of the contamination of groundwater aquifers caused by the intrusion of substances coming from reservoirs is a big environmental issue: reserves of drinkable water are decreasing but at the same time the global need of potable water is increasing and so the preservation of aquifers is a fundamental goal for the entire mankind.

From a general point of view this thesis aims to identify which are the critical parameters related to the intrusion of reservoir substances into aquifers and which is the grade of uncertainty of these processes.

From a specific point of view, the study analyses a specific case: a gas reservoir with an aquifer above is modelled and it has been supposed for the sake of simplicity that reservoir contains only hydrogen. Aquifer and reservoir are separated by a cap rock and there is a fault which links the aquifer to the reservoir.

In this work principal parameters affecting the response (which is expressed through global target quantities) have been identified through two different techniques of Global Sensitivity Analysis (GSA). GSA methods also provide the uncertainty associated to principal parameters for a certain response The investigation relies upon these techniques because it has been demonstrated the efficacy of methods based on variance analysis, which analyse the interaction between multiple parameters affected by uncertainty and also they analyse the propagation of this uncertainty on the considered process.

Two GSA techniques are the calculation of Morris indices and the calculation of Sobol indices. Morris's analysis gives back which parameters influence mostly the mean of

the response, while Sobol's analysis provides which parameters mostly influence the variance of the response.
For each of the two quantities of interest several Probability Density Functions (PDF) have been built (with the PCE method that will be explained afterwards) and matched one with the other in order to discover some correlations and obtain information about structures of responses and their interaction. This probabilistic approach defines levels of risk, predictions and evaluations about values of global quantities starting from a certain level of uncertainty of input parameters.
A peculiarity to remark is that the calculation of Sobol indices has been done with the aid of PCE (*Polynomial Chaos Expansion)* decomposition, which calculates indices in an analytical way through the polynomial coefficients obtained previously from the decomposition which approximates the model.
The utilization of the PCE has two advantages: (1) it provides a surrogate model, which is useful for other operations like Monte Carlo simulations and (2) it reduces of many orders the time needed for the calculation of the indices.
Sparse grids collocation method has been chosen to build up the PCE surrogate model. If compared to traditional Monte Carlo methods this choice allows to reduce appreciably the computational cost associated to the construction of the surrogate model and as a consequence the computational cost of Sobol indices. The procedure for the Sobol indices calculation, the definition of the PCE surrogate model and *Sparse grids* collocation method have been executed by means of calculation codes already developed during past studies conducted by MOX of the Politecnico of Milan.
In chapter 1 the general contest of this study is introduced by starting from a general view of the energetic panorama. Then a brief introduction to the new horizons of improvement regarding oil and gas world is presented and the focus is set on new zones of exploration, on new drilling techniques and on new techniques aimed to improve production. As last the goal of the thesis is presented and collocated in the global contest.
In chapter 2 the methodologic approach is illustrated, in other words it is explained how to pursue the goal. So in this chapter firstly it is shown briefly what full model is (the configuration of the problem, which are input parameters and their domain, which are target quantities of interest) and how its numerical code works. Secondly it is

displayed the construction of the surrogate model.
In chapter 3 theories of Sobol and Morris are introduced, which are essentially the contribute to the analysis of sensitivity.
In chapter 4 results are presented. There is a part of results that is called “Preliminary Results”, which is a recap of all simulations that were useful to understand the right way to approach to the problem, although their results weren’t significant if taken by themselves. The second part of the chapter is dedicated to the final results which try to answer to the initial question: which are the most influent parameters that govern the intrusion of gas in the aquifer for our specific case? Although limited to a very specific case, these results could be used as a starting point for future analysis or for comparisons with other cases.
Chapter 5 is for the conclusions, for some suggestions about new ANOVA (ANalysis Of VAriance) techniques that could be integrated with those present in this study and some considerations for future analysis.

1 GENERAL OVERVIEW

Nowadays the global request of Oil and Gas is continuously increasing, in 2015 was around 562 million barrels for oil and about 3469 billion cubic meters for what concerns gas (OPEC Annual Report, 2016; Bp Review, 2016).
The new economic powers of Asia like China, India, Indonesia and Brazil will play a fundamental role in this growth, but also other big countries like USA and Russia will probably increment their energy demand.
Most of statistics state energy requirements will increase by 34% through the next 20 years. This growth will be generated by an increment of the global population of about 1.5 billion people. Oil and Gas will remain the energy keys, providing about 60% of the increase on energy and accounting for almost 80% of total energy supplies in 2035 (especially demand for natural gas will grow strongly). Along with fossil fuels Renewables Energies will improve dramatically, thanks to better competitive costs and they are expected to act a fundamental role in the second part of the century (BP Energy Outlook, 2016).

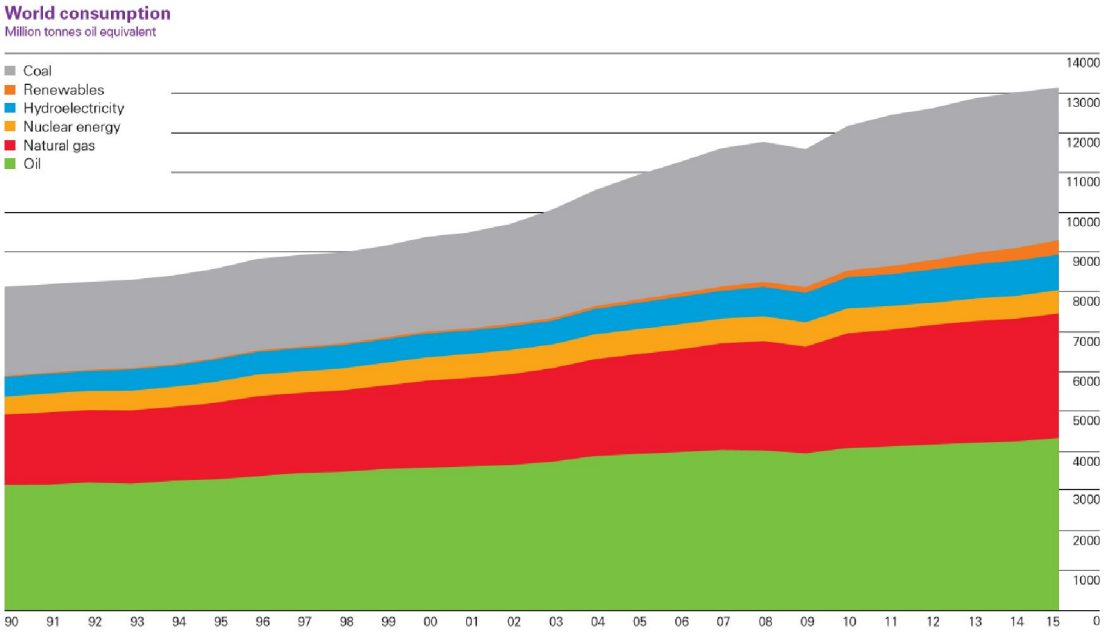


Figure 1.1. World primary energy consumption in 2015.

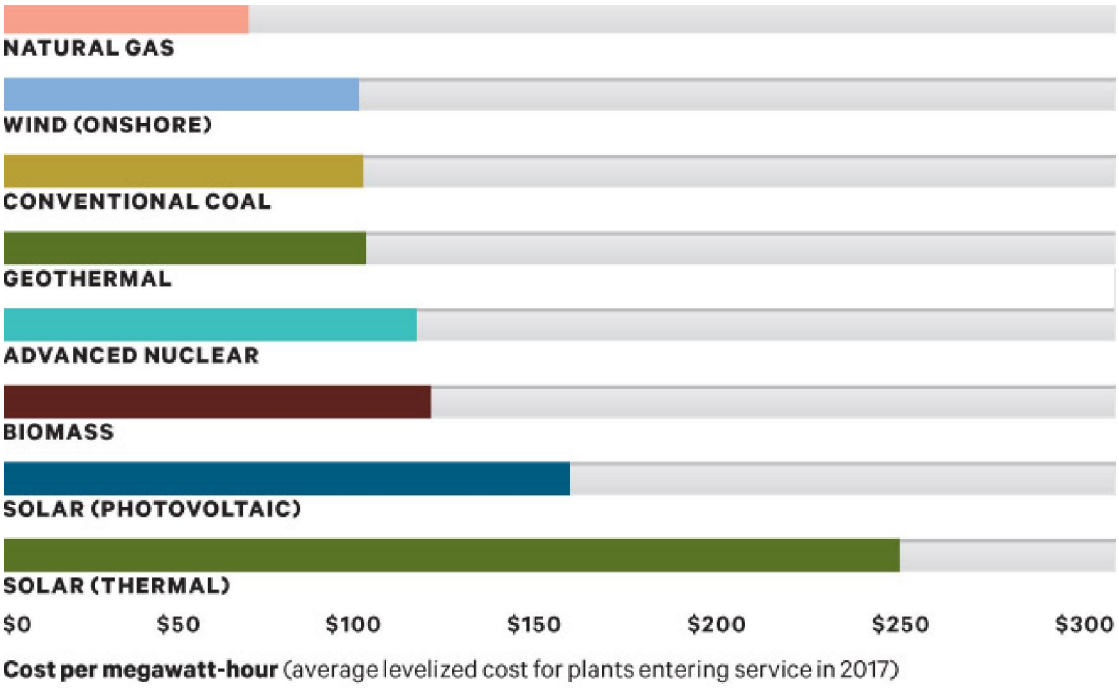


Figure 1.2. Actual cost of electric energy for different renewables sources.

In the Oil & Gas world plenty of techniques have been developed during past years to supply this increasing demand. Surely refining techniques have been developed in order to increase the fraction of most valuable products coming from oil and in order to improve efficiency of processes and minimize production wastes.

But most of research and development has been applied on O&G exploration and O&G production sectors. Indeed, the new horizon of exploration is the off-shore exploration, where the extraction has been limited in the past years mostly because of higher costs and because of lack of suited technologies. To have an idea in 1970 450 meters was the maximum deepness for exploration operations, nowadays (2014) the new limit for exploration is 3174 meters, made by ONGC company in the Indian sea. This impressive growth was promoted like just said by the increasing demand and at the same time by the fact that the accessible reserves on-shore were draining away. The other principal sector of research and development is oil production phase. Oil production techniques include a wide range of operations, from methods used to build wells extractions methods to the methods that boost extractions (tertiary production). In fact there is always a trading between costs and benefits for every choice. To have an example, a vertical well will be less costly (up to ten times) than a horizontal well, but a horizontal well will ensure (if done in a proper way) more production in its life (this can be easily explained by the fact that normally reservoir profiles follow a horizontal development which means they are 10-20-30 wider than high). Moreover, it has to be considered the number of wells, their disposition etc…

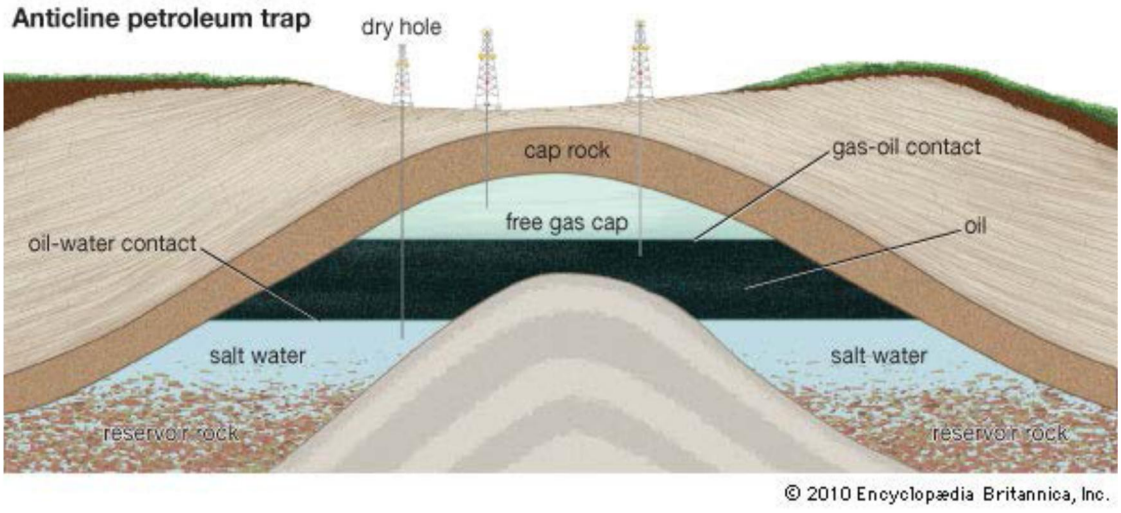


Figure 1.3. General representation of a reservoir.

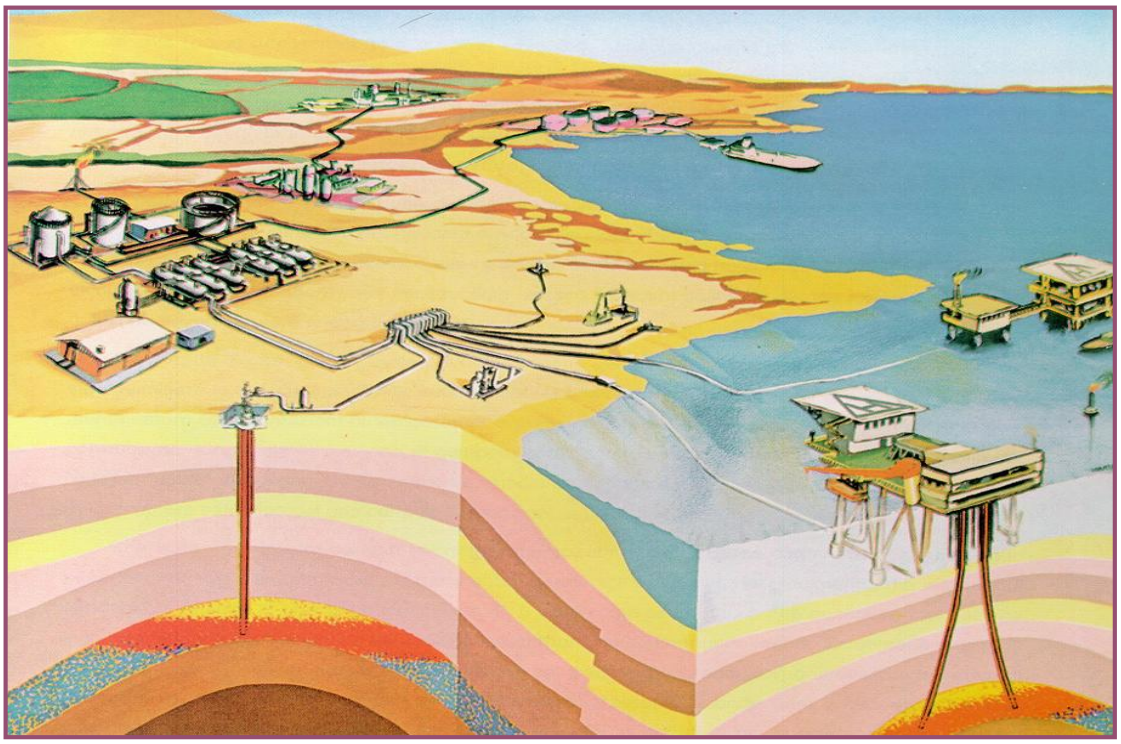


Figure 1.4. General representation of offshore and onshore developments.

Another example about the choice of wells regards the internal structure of the wells: which type of well completion should be installed? More expensive solutions (like single intelligent completion, dual completion, single horizontal completion) permit a better control of the extraction and as consequence they permit to have an optimal production, in contrast to solutions like single completion which is less expensive but also less efficient and flexible.

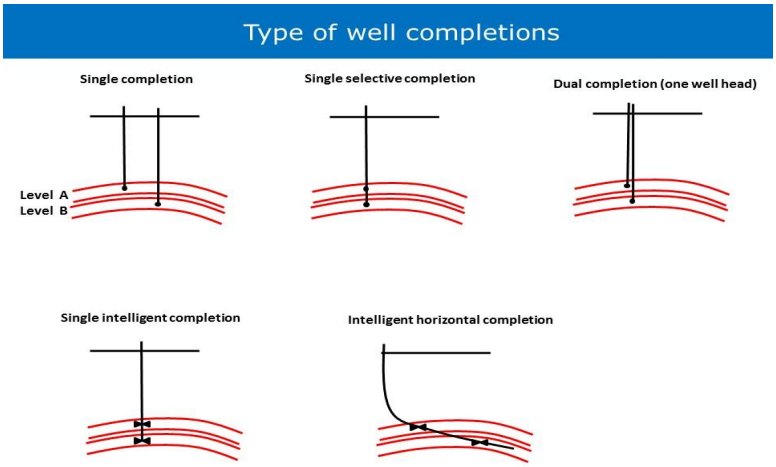


Figure 1.5. Several types of well completions available nowadays.

If primary (well self-production) and secondary production methods (injection of water and gas to maintain high the reservoir pressure) are well established, tertiary methods are continuously updated and improved. Solvents, chemicals, high pressure and temperature vapour are injected by wells into the reservoir, in order to create new paths for the trapped oil and in order to increment permeabilities and so the rate of production. Advantages are very clear considering that reservoirs productions and life wells could increase up to 40 % and over!

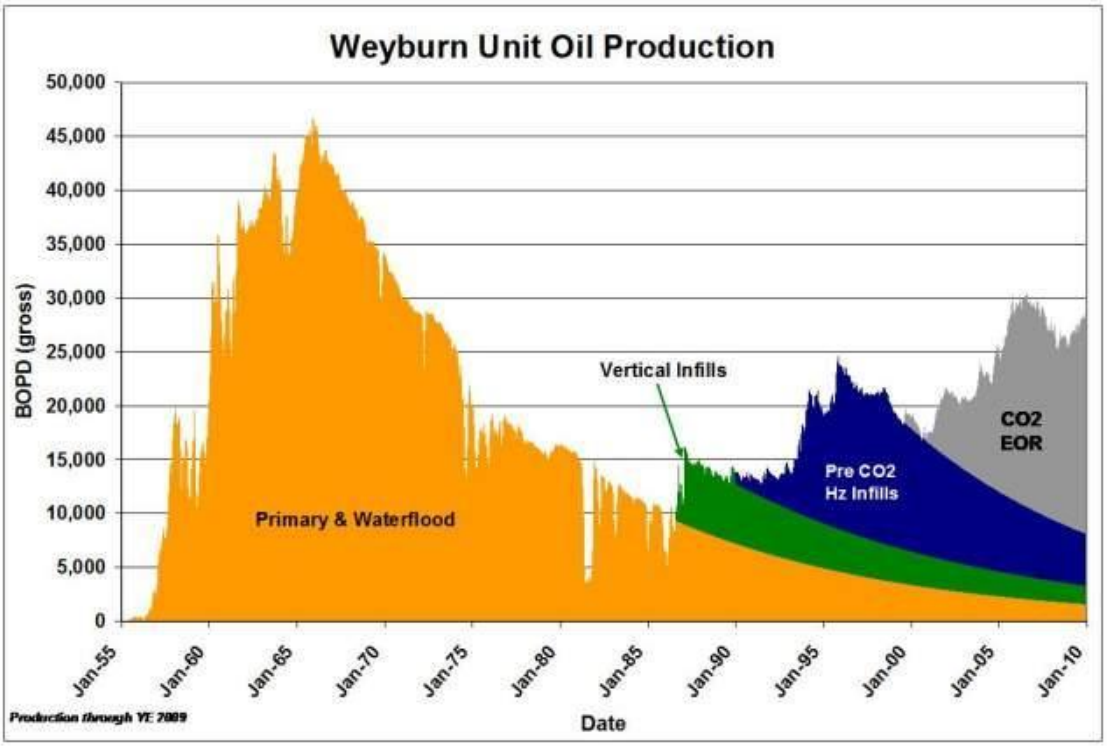


Figure 1.6. An example of tertiary recovery by means of CO2 injection.

But all that glitters is not gold! These techniques are hardly manageable because it is evident that there is not the possibility to know exactly the underground conformation for kilometres and as a consequence the precise effects range of the injections cannot be predicted. In addiction these techniques have a giant power in terms of pressure and temperature and rate and so once they are injected in wells they are mostly unstoppable. As a result, many environmental problems can arise, like instability phenomena of the ground (like earthquakes) and contamination of groundwater aquifers (intrusion of various type of reservoir substances into the aquifer due to the

weakening of the substrate which divides the reservoir from the aquifer).

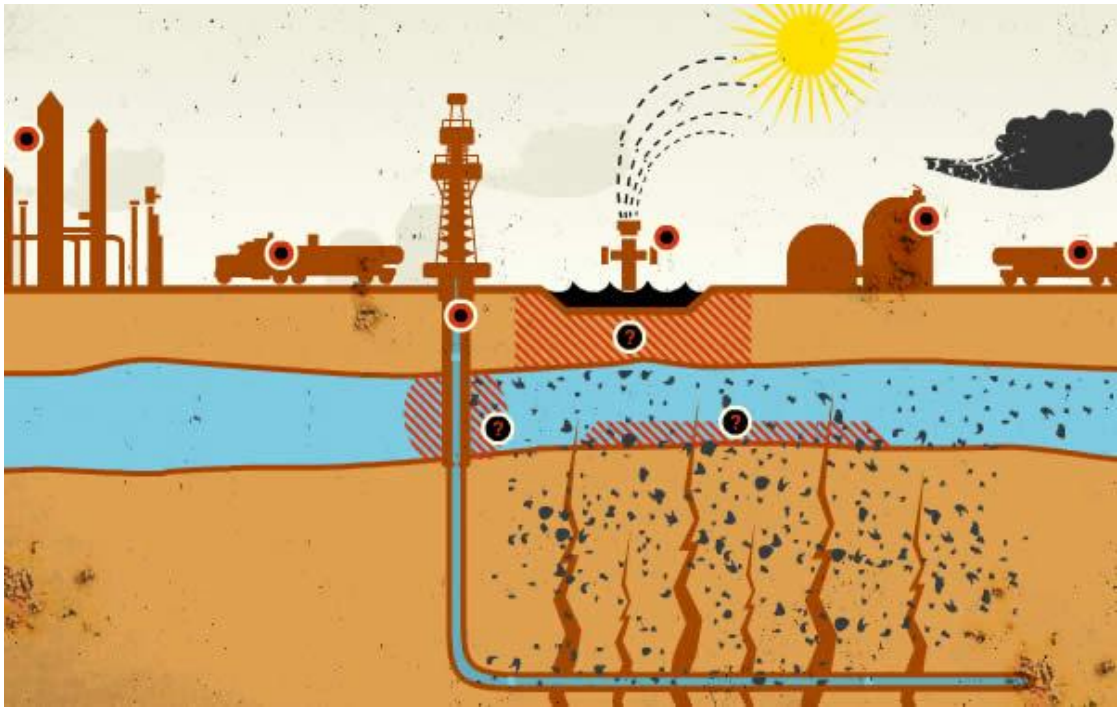


Figure 1.7. Generic representation of possible aquifer contamination deriving from steam injection operations.

This work of thesis analyses this last aspect: by starting from a model of a gas reservoir yet developed the aim is to understand which are the most influent parameters which control the seal of the aquifer respect to the reservoir in order to prevent the poisoning of the aquifer but at the same time keeping the production rate highest as possible. There is also the need of quantifying the uncertainty that affects simulations in order to have an idea of the reliability of results.

1.1 PROCESS EQUATIONS

For the phenomena under study there are four fundamental equations:

1 The Darcy Law:

A spatial domain is *continuous* if variables of state and parameters, which describe the material inside the domain, can be defined in every point of the domain. Considering a *porous medium* two components can be identified: the solid ground matrix and the fluid inside pores. The observation scale which distinguishes the matrix from pores is named *microscopic scale*. The computation of the problem at this scale is an impossible task, because the modelling geometry of pores and the computational cost would be too hard and too long. Because of this reason the process is studied under the *continuous or macroscopic scale*. The passage from the *microscopic* to the *macroscopic scale* happens by means of a process called *volume averaging* (Bear 2010). Without entering in details of the *volume averaging* the porous medium can be defined as the overlay of the solid matrix upon the continuous fluid. In this passage of scale representative coefficients of geometric pores like hydraulic permeabilities and porosities are created.

In relation to the process of flux in a porous medium the starting point is the balance momentum equations for the fluid inside the micro-pores of the solid matrix written in the vector form:

$$\frac{\partial}{\partial t}\rho \boldsymbol{V} = -\boldsymbol{\nabla} \bullet \rho \boldsymbol{V}\boldsymbol{V} + \boldsymbol{\nabla}\boldsymbol{\sigma} + \rho \boldsymbol{F} \qquad (1.1.1)$$

(a) (b) (c) (d)

Where:

ρ	[M/L^3]	Fluid Density
V	[L/T]	Fluid Velocity
σ	[M/LT]	Strains
F	[ML/T]	Volume Forces

And where terms of Equation (1.1.1) are defined for unit of volume

(a) = momentum quantity rate of accumulation

(b) = advection momentum quantity rate of change

(c) = strain-related momentum quantity rate of change

(d) = volume forces-related momentum quantity variation

Next step is the introduction of the following hypothesis:

- The fluid is Newtonian.
- Inertial forces are negligible if compared to viscous forces.
- The dragging effect of momentum quantity into the fluid, resulting from the velocity gradient at the microscopic level, is forgettable respect to the dragging force developed at the interface between solid and fluid.
- The fluid fully saturates pores.
- The solid matrix is unmoveable.

Finally, a simplified expression is obtained:

$$\boldsymbol{q} = \boldsymbol{V}\Phi = -\frac{\underline{\underline{\boldsymbol{k}}}}{\mu}(\nabla p - \boldsymbol{g}\rho\nabla \boldsymbol{z}) \qquad (1.1.2)$$

Where:

q	[L/T]	mean flow rate specific to the volume or Darcy flux
V	[L/T]	mean seepage velocity of the fluid
Φ	[-]	mean constant value of the porosity in the matrix
$\underline{\underline{\boldsymbol{k}}}$	[L^2]	hydraulic permeability tensor
ρ	[M/L^3]	fluid mean density
p	[M/LT]	fluid mean pressure
μ	[M/LT]	fluid mean viscosity
$\boldsymbol{g}$	[L/T^2]	gravity acceleration vector

The term $-\nabla p$ represents forces provoked by the pressure gradient, while $\boldsymbol{g}\rho\nabla \boldsymbol{z}$ incases the gravity force. It is important to remember that $\underline{\underline{\boldsymbol{k}}}$ describes the attitude of the terrain to be passed through by a generic fluid.

2 The Constitutive Equation of the Fluid:

In this study the Constitutive Equation is a critical point. In fact, it is fully empirical and its values are tabulated. So the following function has not an own identity:

$$\rho = F(p) \tag{1.1.3}$$

3 The Continuity Equation:

Without considering any term of mass source of fluid or solute the Continuity Equation is so written:

$$\Phi \frac{\partial \rho}{\partial t} + \nabla(\rho \boldsymbol{V}) = 0 \tag{1.1.4}$$

4 The Advettive-Diffusive-Dispersive Transport Equation

The last fundamental equation is unveiled in the following way:

$$\Phi \frac{\partial \rho X^{\alpha}}{\partial t} - \nabla \cdot \left(\rho \boldsymbol{V} X^{\alpha} + \rho D_m^{\alpha} \frac{M^{\alpha}}{M} \nabla x^{\alpha}\right) = 0 \qquad \text{with } \alpha = (H_2, H_2O) \tag{1.1.5}$$

Where:

D_m^{α}	$[\frac{m^2}{s}]$	molecular diffusion coefficient
X^{H_2}	$[\frac{Kg_{H_2}}{Kg_{H_2+H_2O}}]$	H_2 mass concentration
M^{H_2}	$[\frac{Kg_{H_2}}{Kmol_{H_2}}]$	H_2 molar mass
M	$[\frac{Kg_{H_2+H_2O}}{Kmol_{H_2+H_2O}}]$	fluid molar mass
x^{α}	$[\frac{Kmol_{H_2}}{Kmol_{H_2+H_2O}}]$	H_2 molar concentration

$X^{\alpha} = \frac{M^{\alpha} x^{\alpha}}{M}$

In this transport equation for solved solutes in a liquid phase two main processes are present:

- Advettive transport of solutes conducted by the fluid according to Darcy Law.

- Diffusive transport generated by molecular diffusion inside the porous medium.

2 METHODOLOGICAL APPROACH

In the following pages two *target quantities* which characterize the intrusion of the gas in the aquifer will be firstly presented and then submitted to the GSA (Global Sensitivity Analysis). The GSA has been executed by means of Sobol indices (obtained through the PCE tool) and by Morris indices (see chapter 3). In this chapter the problem is illustrated and its implementation by means of home-made numerical code is shown.

2.1 PROBLEM SETTING

The configuration of the problem is the following (see figure 2.2):

- Geometry: the domain of interest is a vertical rectangular section of the reservoir (called also source aquifer) and above it there is the aquifer (named also target aquifer) separated by a cap rock but linked to the reservoir by a fault. Above the target aquifer there is the overburden formation.

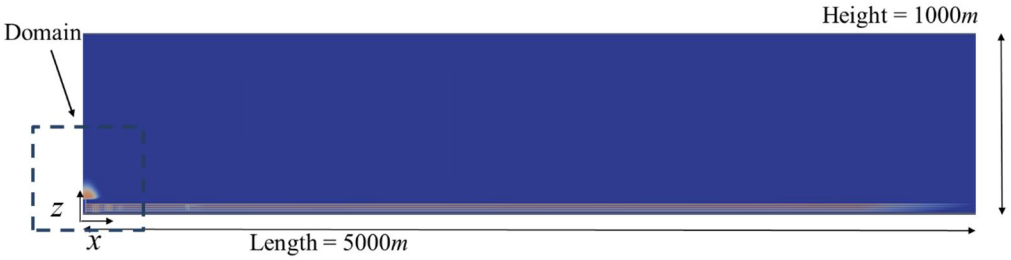


Figure 2.1. Underground Basin Modelling.

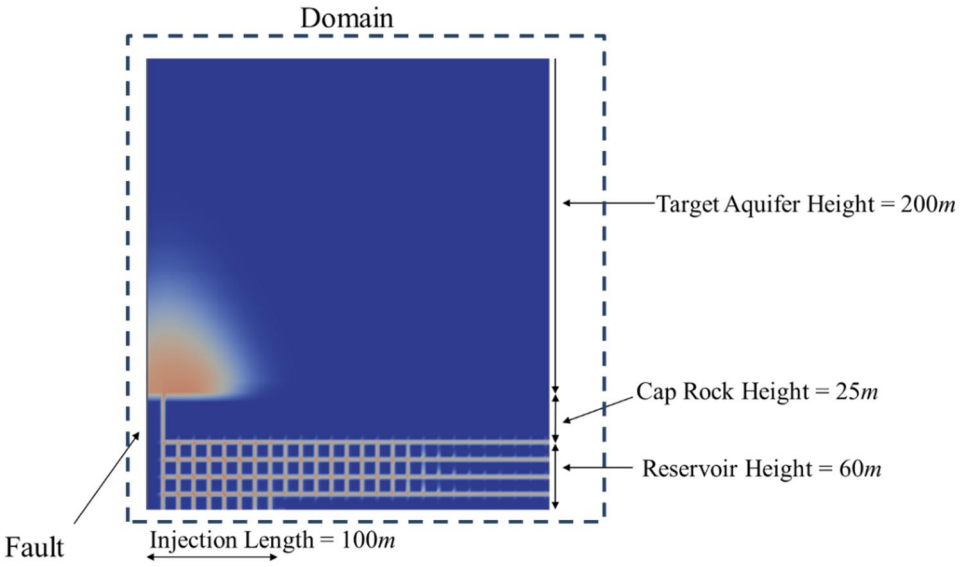


Figure 2.2. Domain of the problem.

- Boundary Conditions (B.C): At the top of the aquifer pressure is equal to atmospheric pressure and its variation with depth follow the classic hydrostatic distribution. There is no flow at boundaries except for the point of injection where during the injection time pressure is equal to the injection pressure P_{INJ}.The mass concentration is equal and constant to a certain value X^{H_2}

during the injection time.

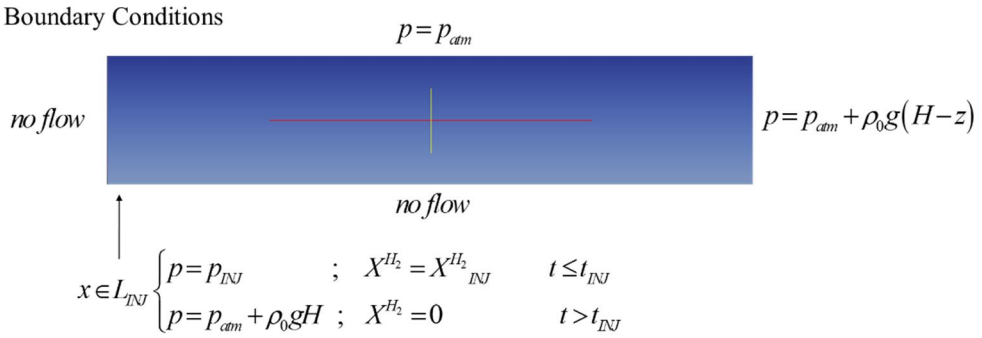


Figure 2.3. Boundary conditions of the problem.

- Initial Conditions (I.C): For these analysis it is necessary to solve flux and transport equations in couple and this fact implies the choice of initial conditions. For the flux a hydrostatic pressure distribution has been chosen and for the transport in all domain the solute concentration is fixed to the zero value.

$$p = p_{atm} + \rho_0 g(H - z) \quad and \quad X_{H_2} = 0 \tag{2.1.1}$$

- Problem Type: The process is isotherm and for what concerns solute transport processes the molecular diffusion and advettive transport are considered!

2.1.1 IMPLEMENTATION WITHIN IN-HOUSE SIMULATOR

The problem has been firstly implemented by means of a home-made numerical code implemented in DUMUX by a working team of Stuttgart University.
This code permits to simulate flux and transport processes of solutes by considering fluid density variations due to effects related to solute concentration.
On the occasion of this thesis the processes are fluxes in a saturated mediums and transport of dissolved solutes in isothermal conditions, without considering any types of reactions. As just mentioned, the densities variations are only associated to the solute concentration.
The calculation domain is a bi-dimensional vertical section 2 meters long and 1 meters high. The computation grill is regular with quadrilateral elements: 128 elements along vertical direction and 256 elements along horizontal direction. The temporal step has been imposed to 60 seconds for all simulations.

2.2 TARGET QUANTITIES AND INPUT PARAMETERS

In this study there was a wide range of possibilities regarding the choice of global target quantities and these quantities have the form of *response surfaces*. A *response surface* of a model refers to a plane of *d* dimensions (*d* is the number of considered factors) that represents the variation of a target response of the model (a state or output variable, or a performance metric) with respect to variations in one or more factors of interest. Factors are features of the model (in this study they are only the parameters of the model) which may vary on continuous domains that define the "factor space". The criterion of decision about *target quantities* is the obtaining of global quantities that describe in an essential way when, how much and how the gas has penetrated in the target aquifer. For this reason, *target quantities* are:

- $M_{H_2,MAX}$ $[kg]$ that is the peak value of accumulated mass of gas in the target aquifer in all the duration of the simulation. It will be evident that accumulated mass grows gradually with time increasing until the end of injection time; then it gradually decreases.

$$M_{H_2}(t) = \int_{\Omega_T} m_{H_2}(t) d\Omega_T \qquad M_{H_2,MAX} = \max\left(M_{H_2}(t)\right) \tag{2.2.1}$$

  Where $M_{H_2}(t)$ is the mass $m_{H_2}(t)$ (specific to the volume and function of time *t*) integrated all over the spatial domain Ω_T.

- $M_{H_2,END}$ $[kg]$ that is the value of accumulated mass of gas in the target aquifer at the end-time of the simulation. This value has been considered interesting for the purpose of the thesis because it reveals how fast the target aquifer disperses the solute in the surrounding environment.

$$M_{H_2,END} = M_{H_2}(t_{END})$$

- $m_{H_2,MAX}$ $[\frac{kg}{m^3}]$ that is the peak value of mass concentration of gas in a cubic meter throughout all the simulation. This mass measure has also the quality to be compatible to real measures sampled in situ because measures of mass are easier to sample and because environmental laws are also based on maximal concentration values defined for each substance and so if the peak of a certain substance is under its related threshold value, the situation is under control.

Nevertheless, this target quantity has been abandoned in this study due to a lack of precision and because the purpose of the thesis isn't the reproduction of sampling methods.

The second part of this paragraph is dedicated to the schedule of input parameters and the choice of its intervals; some parameters will be assumed as known and others will be considered uncertain and so they will have a specific range of uncertainty. The choice of uncertain parameters will vary according to the group of simulations considered.

Remembering the Darcy Law (1.1.2), the Fluid Constitutive Equation (1.1.3), the Continuity Equation (1.1.4) and the Advettive-Diffusive Transport Equation (1.1.5) every *global quantity* ξ can be written as:

$$\xi = \mathrm{F}(P_{INJ}, Z_{BOR}, Z_{RES}, Z_{CR}, Z_{target}, t_{END}, t_{INJ}, M_{fraction}, L_{INJ}, b_{SAf}, b_F, \Phi_{SAm} + \Phi_{CR}, \Phi_T, \Phi_{OV}, \Phi_{SAf}, \Phi_F, k_{SAm}, k_F, k_{SAf}, k_T, k_{CR}, k_{OV}, B.C., I.C) \quad (2.2.2).$$

P_{INJ} is the pressure at which vapour is injected in the reservoir in order to improve productivity, Z parameters are geometric parameters defining the geometry of the problem, $M_{fraction}$ is the fraction of mass of H_2, b parameters define width of fractures and faults, t parameters define the timing of problem, Φ parameters are porosities of the several formations of the space problem and k parameters are its correspondent permeabilities. Moreover, fluid properties of water and H_2 (like densities and viscosities) are well known and uniform for all the duration of simulations and constant throughout the experiment.

Follow a table describing these parameters (their names, symbols adopted and units of measurement):

Table 2.1. Factors composing the model.

Spatial Parameters		
Reservoir Permeability	k_{SA_m}	$[m^2]$
Fault Permeability	k_F	$[m^2]$
Fracture Permeability	k_{SA_f}	$[m^2]$
Target Permeability	k_T	$[m^2]$

Overburden Permeability	k_{OV}	$[m^2]$
Cap Rock Permeability	k_{CR}	$[m^2]$
Reservoir Porosity	Φ_{SA_m}	$[-]$
Fault Porosity	Φ_F	$[-]$
Fracture Porosity	Φ_{SA_f}	$[-]$
Target Porosity	Φ_T	$[-]$
Overburden Porosity	Φ_{OV}	$[-]$
Cap Rock Porosity	Φ_{CR}	$[-]$
Operational and Geometric Parameters		
BOR Depth	Z_{BOR}	$[m]$
Reservoir Height	Z_{RES}	$[m]$
Cap Rock Thickness	Z_{CR}	$[m]$
Fracture Width	b_{SA_f}	$[m]$
Fault Width	b_F	$[m]$
Injection Pressure	P_{INJ}	$[bar]$
Simulation Time	t_{END}	$[s]$
Injection Time	t_{INJ}	$[s]$
H_2 Fraction Mass Injected	$M_{H_{2fraction}}$	$\left[\frac{kg_{H_2}}{kg_{H_2O}}\right]$
Injection Length	L_{INJ}	$[m]$
Target Thickness	Z_{TARGET}	$[m]$
Fluid properties and others		
Water Viscosity	μ_{H_2O}	$\left[\frac{kg}{ms}\right]$
Diffusion Coefficient	$D_m^{H_2,H_2O}$	$\left[\frac{m^2}{s}\right]$
Water Density	ρ_{H_2O}	$\left[\frac{kg}{m^3}\right]$
Gravity Force	g	$\left[\frac{m}{s^2}\right]$

As said before, uncertain parameters will vary in the following simulations. The reason is that to execute a sensitivity analysis some input parameters must vary. So it isn't only a matter of uncertainty but rather a need to explore how the response modifies if the set of uncertain parameters and their values change. So for each group of simulation, the set of uncertain parameters will be specified. It must be said that when a parameter is chosen as uncertain, its standard value given in the origin will be more or less in the centre of its variation range. The only exception of this concept is applied to the injection pressure because of its implications that will be explained in the next chapters.

2.3 MODEL REDUCTION TECHNIQUE BY GENERALIZED POLYNOMIAL CHAOS EXPANSION (g-PCE)

The starting concept to understand why a surrogate model is necessary can be the observation that the execution of a single simulation of the full model requires a big computational cost or in other words it requires a lot of time, more than five minutes with a standard processor. This time could seem small, but it must be considered that thousands simulations are necessary to execute the GSA and hundreds of thousands simulations are needed to obtain PDFs.

Firstly, we define a target quantity $f(\boldsymbol{x})$, representing a given *global quantity* and this function depends on N independent random variables kept together in the $\boldsymbol{x}$ vector ($\boldsymbol{x} \equiv x_1, \ldots, x_N$); $\boldsymbol{x}$ vector contains uncertain parameters chosen for the simulation. Since no detailed information on geochemical compaction model parameters is typically available, each x_n is described by a uniform distribution within the interval $\Psi_n = [x_{n,min}, x_{n,max}]$. Moreover, $f(\boldsymbol{x})$ can be decomposed as

$$f(\boldsymbol{x}) = f_0 + \sum_{i=1}^{N} f_i(x_i) + \sum_{i,j=1}^{N} f_{i,j}(x_i, x_j) + \cdots f_{1,2,\ldots,N}(x_1, x_2, x_N) \qquad (2.3.1)$$

Where:

$$f_0 = \int_{\Psi} f(\boldsymbol{x}) p_{\Psi}(\boldsymbol{x}) d\boldsymbol{x}; \qquad f_i(x_i) = \int_{\Psi_{-i}} f(\boldsymbol{x}) p_{\Psi\_i}(\boldsymbol{x}) d\boldsymbol{x} - f_0$$

$$f_{i,j}(x_i, x_j) = \int_{\Psi_{-i,j}} f(\boldsymbol{x}) p_{\Psi_{-i,j}}(\boldsymbol{x}) d\boldsymbol{x} - f_0 - f_i(x_i) - f_j(x_j) \qquad (2.3.2)$$

Here $\Psi = \Psi_1 \times \ldots \times \Psi_N$ is the hypercube of the space of variability of $\boldsymbol{x}$ and $p_\Psi(\boldsymbol{x})$ is the joint probability density of $\boldsymbol{x}$ over Ψ. It must be said that integration over Ψ_i is performed over the space of $\boldsymbol{x}$ excluding Ψ_i, where p_{Ψ_i} being the corresponding density function. Another interesting characteristic is that f_0 is the mean of $f(\boldsymbol{x})$.

Going back to the construction of the surrogate model the tool that was chosen to accomplish the task was the so called g-PCE (generalized Polynomial Chaos Expansion). First approaches to the PCE were tried back in the 30s by Wiener. Several different methods based on Wiener 's theory have been developed during past years but the fundamental concept of a g-PCE is unique: the spectral expansion of $f(\boldsymbol{x})$ in terms of a set of orthonormal polynomials representing a basis of the probabilistic space Ψ within which an approximation of the model response surface is built. The choice of the specific family of polynomials is linked to the type of probability distribution of the uncertain model parameters. In this work each x_n is assumed to be uniformly distributed and as a consequence the family of multivariate Legendre polynomials are chosen. Without entering in details of how building multidimensional Legendre polynomials starting from univariate Legendre polynomials it is written straightaway the expression of the g-PCE:

$$f(\boldsymbol{x}) = \theta_0 + \sum_{i=1}^{N} \sum_{\boldsymbol{s} \in \Upsilon_i} \theta_{\boldsymbol{s}} L_{\boldsymbol{s}}(\boldsymbol{x}) + \sum_{i,j=1}^{N} \sum_{\boldsymbol{s} \in \Upsilon_{i,j}} \theta_{\boldsymbol{s}} L_{\boldsymbol{s}}(\boldsymbol{x}) + \cdots; \qquad (2.3.3)$$

With:

$$L_{\boldsymbol{s}}(\boldsymbol{x}) = \prod_{i=1}^{N} L_{i,s_i}(x_i); \qquad \theta_{\boldsymbol{s}} = \int_{\Psi} f(\boldsymbol{x}) L_{\boldsymbol{s}}(\boldsymbol{x})\, p_{\Psi}(\boldsymbol{x}) d\boldsymbol{x}$$

Here $\boldsymbol{s} = \{s_1, s_2, \ldots, s_N\} \in \boldsymbol{N}^N$ is a multi-index expressing the degree of each univariate Legendre polynomial, $L_{n,s_n}(x_n)$, employed to construct the multivariate orthogonal Legendre polynomial $L_{\boldsymbol{s}}(\boldsymbol{x})$. $\theta_{\boldsymbol{s}}$ is the associated polynomial coefficient and Υ_i contains all indices such that only the ith component does not vanish, i.e., $\Upsilon_i = \{s_i \neq 0, s_k = 0 \; for \; k \neq i\}$.

Some words must also be spent about the way the coefficients $\theta_{\boldsymbol{s}}$ are computed because without some expedient an elevate number of high-dimensional integrals should be calculated. The applied method is the Sparse Grid polynomial approximation of *f*. The goal of this method is to reduce computational cost of the tensor grid interpolants while maintaining a high accuracy. This can be obtained by building the approximation as a particular linear combination of tensor grids, where

each one of these includes only a limited number of points. The expression of sparse grid approximation of *f* is written as

$$f_{SG}(\boldsymbol{x}) = \sum_{\boldsymbol{I}\in\Lambda} c_{\boldsymbol{I}}\, f_{TG,\boldsymbol{I}}(\boldsymbol{x})$$

where $\Lambda \in N^N$ is a set of multi-indices that selects the set of tensor grids that form the sparse grid, $c_{\boldsymbol{I}}$is the suitable computable coefficient and $f_{TG,\boldsymbol{I}}(\boldsymbol{x})$ is the interpolant tensor associated to the tensor grid $\mathcal{H}_{\boldsymbol{I}} = \mathcal{H}_1^{l_1} \times \mathcal{H}_2^{l_2} \times \ldots \times \mathcal{H}_N^{l_n}$, constituted by N $\mathcal{H}_n^{l_n} = \{x_{n,1}, x_{n,2}, \ldots, x_{n,l_n}\} \subset \Psi_n$, which are sets of interpolation points that provide a precise interpolation with respect to the uniform probability measure $p_{\Psi_i}(x_i)$.

The conversion from sparse grid to g-PCE expansion is established on the characteristic that a sparse grid is constituted by a linear combination of tensor grid interpolants and is grounded on two steps: (a) each tensor Lagrangian interpolant is reformulated in terms of Legendre polynomials and (b) the Legendre polynomials associated with different tensor grids are summed. In essence, without entering in details (for details refers to Barthelmann ‘s works), converting the sparse grid into a g-PCE requires solving as many systems as the number of tensor grids in the sparse grid.

3 SENSITIVITY ANALYSIS

The complexity of reservoir modelling is growing fast due to the new challenges the Oil & Gas world is facing. New fields are more expensive to be explored because of their remote locations. So there is a need to characterize reservoirs with few information and at the same time to compute models that represent as much as possible the real underground conformations, geometries and types of fluids. As a consequence, such models are more and more complex and computationally intensive. Sensitivity Analysis has become an undeniable instrument in the manipulation and analysis of such models; but as models become more complex, sensitivity analyses increment their computational cost. So it is evident the importance of developing strategies for SA that are both effective and efficient.

SA is multitasking, in fact it can be oriented on several aspects (Razavi & Gupta 2015):

- Test the level of similarity between the functioning of the model and the underlying system.
- Quantify the relative importance of each factor in relation to the model/system response.
- Identify if there are some regions in the factor space where variations in the factor produces great variability in response.
- Investigate on the eventual type and level of correlation between factors.
- Find noninfluential factors in order to lift them from the model and so simplify the model itself.

3.1 A REVIEW OF SENSITIVITY ANALYSIS METHODS

A variety of approaches, theories and philosophies regarding sensitivity analysis are present in the literature, based on different philosophies and theories and sometimes the definition of sensitivity is not clear or at least unique.
Essentially sensitivity analysis methods are classified into two categories (Razavi & Gupta 2015, Sudret,2008; Saltelli,2006):

- LSA (*Local Sensitivity Analysis*): This analysis gives back the variation of the response around a prefixed value and respect to input parameters.
  In other words, the sensitivity of a model output function Y (also called response Y) to a factor x_i is defined as the derivative $\frac{\partial Y}{\partial x_i}$ of the model response Y in the direction of increasing values of parameter x_i. So consider the function

$$Y = f(x_1, \dots, x_d) \tag{3.1.1}$$

  Where $x_1, \dots, x_d$ are parameters of interest characterized by some distributions and intervals. The rate of change s_i of response Y with respect to factor x_i $(1 \leq i \leq d)$ can be evaluated at a *specific base point* $(x_1^*, \dots, x_d^*)$ in the factor space as follows:

$$s_i = \left(\frac{\partial Y}{\partial x_i}\right)_{(x_1^*,\dots,x_d^*)} \tag{3.1.2}$$

  Under this definition s_i characterizes the independent effect of parameter x_i when all other parameters remain constant.
  The sensitivity of the model representing the interaction between two parameters x_i and x_j (with $i \neq j$ and $1 \leq i, j \leq d$) on the model response is obtained by using a second-order partial derivative; at the same way third-order partial derivative defines sensitivity due to three-factor x_i, x_j, x_k interactions (with $i \neq j \neq k$ and $1 \leq i, j, k \leq d$) and so on:

$$s_{ij} = \left(\frac{\partial^2 Y}{\partial x_i x_j}\right)_{(x_1^*,\ldots,x_d^*)} \tag{3.1.3}$$

$$s_{ijk} = \left(\frac{\partial^3 Y}{\partial x_i x_j x_k}\right)_{(x_1^*,\ldots,x_d^*)} \tag{3.1.4}$$

As said before this method provides results which are valid only in the close vicinity of the *base point* in the parameter space.

- GSA (*Global Sensitivity Analysis*): The task of this analysis is to determine the entity and ways of how the response surface is influenced by single or multiple combinations of parameters, considering them individually or in their mutual relationships.

  The general procedure for a GSA method requires a certain number of points belonging to the response surface computed at different points of the domain of interest that are representative of the entire domain.

In literature many different methods applied to the GSA are present but between these five can be pinpointed as the most important:

- *Factorial design methods:* all the factor space is discretized into a certain number of levels (i.e. grid points) and the number of the levels for each factor will determine the quality of GSA. In fact, the model response is computed for each combination of factor levels and points obtained are used to estimate the "main effects" of each factor and also the "interaction effects" between various combination of factors.

- *Regression-based methods:* The standardized regression coefficients (SRC) are based on a regression of the output on the input vector. The coefficients of linear and second-order interaction terms represent respectively the main and interaction effects. The input/output Pearson correlation coefficients measure the effect of each input variable by the correlation it has with the model output. These coefficients are useful to measure the effect of the input variables if the

model is linear, i.e. if the coefficient of determination R^2 of the regression is close to one.

- *Regional sensitivity analysis (RSA)*: First step is to partition the marginal distribution of samples obtained for each factor into two or more distributions based on empirically selected threshold values for model response. The idea behind is that if the factor does not have a considerable influence on model output throughout the factor space, the two distributions should not be statistically distinguishable.
- *Globally aggregated measures of local sensitivities*: Such methods compute the local sensitivity coefficients for each factor at multiple points across the factor space and analyse the distributional properties of these values to assess the global sensitivities of model response to individual factors. The pioneer of this technique was Max Morris (1991) who proposed a way to aggregate local sensitivities measures in order to obtain a representation of global sensitivities. From his first approach many techniques have been developed upon it and in this work Campolongo 's improvement is adopted (Campolongo et al., 2007).
- *Variance-based methods:* These methods aim at decomposing the variance of the output as a sum of contributions of each input variable, or combinations thereof. These methods are also called ANOVA techniques for 'Analysis Of VAriance'. Sobol indices belong to this last group and their aim is to determine the sensitivity for non-linear models (Sobol 2001, Sobol 2005). Usually the calculation of these indices need the utilization of Monte Carlo techniques, which have a high computational cost. ANOVA techniques include another method whose name is FAST (Fourier Amplitude Sensitivity Test) (Saltelli 1999).

  In more general terms we can imagine a model like $Y = f(X)$, where $X = \{x_1, x_2, \dots, x_d\}$ are input parameters (aleatory and independent variables), f is a deterministic function (e.g system of differential equations) and Y the result of the scalar model. Global Sensitivity Analysis doesn't regard any distinction about parameter input values assumed but instead it provides an analysis of variance of the response Y, based on the whole variation range of the vector of parameters X.

Sobol indices are an effective instrument for GSA, in fact they decompose the variance of the response into the several contributes related to input parameters. The calculation of Sobol indices can be obtained by employing a Polynomial Chaos Expansion (PCE) (Sudret, 2008) of the system response, based upon input parameters. The utilization of a generalized PCE allows:

- The creation of a surrogate model constructed following certain rules. The surrogate model allows to analyse scalar quantities related to the system under consideration, in function of a set of parameters defined with affordable computational costs.
- Sobol indices can be obtained analytically by means of a suited manipulation of PCE coefficients and the advantage is that computation of Sobol indices is very time-short.

The hardest part for what concerns computational costs regard the calculation of PCE coefficients, because they are obtained from a multi-dimensional integer (which is built from input parameters). For the evaluation of this integer it is necessary to pass through the complete numerical-mathematical model chosen in the relative ambit (in our case we have used a home-made model) for a well-defined number of times. The determination of PCE coefficients in this thesis ambit has been made by using the polynomial interpolation based upon *sparse grids*. In this way the computational cost has been reduced significantly respect to Monte Carlo standard method.
Just below a brief introduction to mathematical theory regarding global analysis of variance is given. Related tools which are used are Sobol indices, polynomial chaos expansion and the calculation of coefficients by means of *sparse grids*.
The methodology here introduced has been implemented into numerical code created at MOX centre, Politecnico di Milano. Thanks to this code *sparse grids* are worked out and by *sparse grids* the number of necessary points to obtain PCE coefficients is optimized. The code permits also to build the polynomial chaos expansion (PCE) and the requested Sobol indices.

3.2 SOBOL INDICES

Consider a mathematical model with d input parameters, contained in a vector $\boldsymbol{x}$ and a scalar output y:

$$y = f(\boldsymbol{x}), \boldsymbol{x} \in \Psi^d \tag{3.2.1}$$

where Ψ^d is the parameter space $\Psi^d = \{x: 0 \leq x_i \leq 1, i = 1, \ldots, d\}$. If it is expressed by summations, The Sobol decomposition (2001) of $f(x)$ results:

$$f(x_1, \ldots, x_d) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \leqslant i \leqslant j \leqslant d}^{d} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\ldots,d}(x_1, \ldots, x_d) \tag{3.2.2}$$

where f_0 is a constant and the integral of each summand $f_{i_1,\ldots,i_s}(x_{i_1,\ldots,i_s})$ over any of its independent variables is zero, i.e.:

$$\int_0^1 f_{i_1,\ldots,i_s}(x_{i_1}, \ldots, x_{i_s}) \bullet dx_{i_k} = 0 \quad for\ 1 \leqslant k \leqslant s \tag{3.2.3}$$

The classical properties of this decomposition are the following:

i. The sum in Eq. (3.2.2) contains a number of summands equal to

$$\sum_{j=1}^{d} \binom{d}{j} = 2^d - 1 \tag{3.2.4}$$

ii. The constant f_0 is the mean value of the function:

$$f_0 = \int_{\Psi^d} f(\boldsymbol{x}) d\boldsymbol{x} \tag{3.2.5}$$

iii. Due to Eq. (3.2.3) the summands are orthogonal to each other in the following sense:

$$\int_{\Psi^d} f_{i_1,\ldots,i_s}(x_{i_1}, \ldots, x_{i_s})\, f_{j_1,\ldots,j_s}(x_{j_1}, \ldots, x_{j_s}) d\boldsymbol{x} = 0 \tag{3.2.6}$$
$$for\ \{i_1, \ldots, i_s\} \neq \{j_1, \ldots, j_t\}$$

With the above assumptions, the decomposition in Eq. (3.1.2) is *unique* whenever f($\boldsymbol{x}$) is integrable over Ψ^d. In addition, the terms in the decomposition can be obtained analytically. So the univariate terms read:

$$f_i(x_i) = \int_{\Psi^{d-1}} f(\boldsymbol{x}) d\boldsymbol{x}_{\sim i} - f_0 \tag{3.2.7}$$

In this last expression $\int_{\Psi^{d-1}} d\boldsymbol{x}_{\sim i}$ means that integration is over all variables except x_i. In the same way the bivariate terms can be written as

$$f_{ij}(x_i, x_j) = \int_{\Psi^{d-2}} f(\boldsymbol{x}) d\boldsymbol{x}_{\sim\{ij\}} - f_i(x_i) - f_j(x_j) - f_0 \tag{3.2.8}$$

Also here $\int_{\Psi^{d-2}} d\boldsymbol{x}_{\sim\{ij\}}$ represents the integration over all parameters except x_i and x_j. As a result, any summand $f_{i_1,\dots,i_s}(x_{i_1}, \dots, x_{i_s})$ can be written as the difference of a multidimensional integral and summands of lower order.

Now it is assumed that input parameters are independent random variables uniformly distributed in the interval [0, 1]:

$$\boldsymbol{X} = \{X_1, \dots, X_d\} \qquad X_i \sim \Im(0,1), \quad i=1,\dots,d. \tag{3.2.9}$$

Because of this the model response Y=*f*(**X**) is a random variable, whose total variance D has this form:

$$D = Var[f(\boldsymbol{X})] = \int_{\Psi^d} f^2(\boldsymbol{x}) p_\Psi(\boldsymbol{x}) d\boldsymbol{x} - f_0^2 \tag{3.2.10}$$

Where $p_\Psi(\boldsymbol{x})$ is the probability distribution defined by the product of single variables independent probabilities $\boldsymbol{x}$, this term disappears from the moment a uniform distribution is present. It is possible to divide the total variance like follows:

$$D = \sum_{i=1}^{n} D_i + \sum_{1 \le i < j \le d}^{n} D_{ij} + \dots + D_{1,2,\dots,d} \tag{3.2.11}$$

where partial contributes are defined in the following expansion:

$$D_{i_1,\dots,i_s} = \int f^2_{i_1,\dots,i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_1}, \dots, dx_{i_s} \tag{3.2.12}$$
$$with\ 1 \le i_1 < \dots < i_s \le d, s = 1, \dots, d$$

Sobol indices are defined as

$$S_{i_1,\dots,i_s} = \frac{D_{i_1,\dots,i_s}}{D}\ or\ in\ the\ equivalent\ form$$

$$S_{i_1,\dots,i_s} = \frac{1}{D}\int_{\Psi_{i_1,\dots,i_s}} f^2_{i_1,\dots,i_s}\left(x_{i_1},\dots,x_{i_s}\right)dx_{i_1},\dots,dx_{i_s} \qquad (3.2.13)$$

According to (3.2.11) they respect the following rule:

$$\sum_{i=1}^{n} S_i + \sum_{1\le i<j\le d}^{n} S_{ij} + \cdots + S_{1,2,\dots,d} = 1 \qquad (3.2.14)$$

Every index $S_{i_1,\dots,i_s}$ is a sensitivity measure which describes the amount of the total variance due to the uncertainties in the set of input parameters $\{i_1,\dots,i_s\}$. The first order S_i is the representation of the contribute to the total variance of a single input parameter, without the interaction with other parameters. The total sensitivity indices S_{T_i} identify the contribution to the total variance of a single input parameter considering also interactions with other input parameters.

$$S_i^T = \sum_{S} S_{i_1,i_2,\dots,i_s} \qquad (3.2.15)$$

A clear consequence is that:

$$S_i^T = 1 - S_{\sim i},$$

where $S_{\sim i}$ is the sum of all $S_{i_1,\dots,i_s}$ that do not include index *i*.

Sobol indices are known to be good descriptors of the sensitivity of the model respect to its input parameters, because there is no linearity or monotony in the functional bond which links the response to parameters (Saltelli 2006). If no analytical expressions of the response are available, the full description of Sobol indices requires the evaluation of 2^d Monte Carlo integrals. This operation isn't usually affordable it is too expensive in a computational sense (Sudret, 2008). So the first way to solve the matter is computing only indices of first and second order and the second way is to derivate Sobol indices analytically from the g-PCE built for the model and this is the method applied for these studies. Recalling the discussion about the decomposition of the response (2.3.1) and the following g-PCE decomposition (2.3.3) makes possible deriving the equivalence between the Sobol indices and the coefficients $\theta_{\boldsymbol{s}}$ of the gPCE representation of $f(\boldsymbol{x})$, i.e.

$$S_{i_1,\dots,i_s} = \frac{1}{D} \sum_{\boldsymbol{s} \in Y_{i_1,\dots,i_S}} \theta_{\boldsymbol{s}}^2 \; ; \; f_0 = \theta_0 \; ; \quad D = \sum_{\boldsymbol{s} \in \boldsymbol{N}^N} \theta_{\boldsymbol{s}}^2 \; ; \tag{3.2.16}$$

Eq. (3.2.16) can be made workable with some type of truncation, e.g. upon truncation of the summation to a set of polynomials with total degree *w*, i.e., $\sum_i s_i \leq w$. The precision of the resulting gPCE approximation increases with the regularity of $f(\boldsymbol{x})$ and as $w \to \infty$. To recap, in Sobol works (Sobol,2001), it is shown that if the representation of a scalar response by means of PCE is known, it is possible to determine the complete list of Sobol indices without any relevant additive cost, because the calculation of indices requires only elementary operations of maths. So in the end representing the model response through a PCE expansion brings two advantages (Sudret,2008):

- Sobol sensitivity indices are obtained in a simple manner because they derivate from the post-processing of PCE coefficients
- It was shown that it gives a surrogate model, useful to reduce computational costs and to execute Monte Carlo simulations based on the response initially chosen.

3.3 MORRIS INDICES

The other tool applied in the sensitivity analysis of this thesis is based on Morris theory. In 1991 Morris proposed a method which it proved to be well-suited for models with high computational costs and for cases with many uncertain input factors. The method is founded on computing for each input parameter a number of incremental ratios, called Elementary Effects (EE), from which basic statistical operations are executed in order to obtain sensitivity information. Although it was being proven that the EE method is a very good tool for sensitivity analysis of large models (it has an optimal balance between accuracy and efficiency), it is still not much used. In this work the Morris updated method implemented by F. Campolongo, J. Carboni and A. Saltelli is adopted.
The original EE method (Morris, 1991) has the purpose to determine which input factors may be considered to have effects which are (a) negligible, (b) linear and additive, or (c) non-linear or involved in interactions with other factors. For every input parameter two sensitivity measures are calculated: μ, which embodies the overall influence of the parameter on the output, and σ, which gives back an estimation of the ensemble of the factor's higher order effects (like non-linear and/or other factors linked effects). The experimental plan is built by individually randomised OAT experiments. OAT (Once Factor At Time) method computes a finite difference approximation of the local slope of the response surface around a base point in the factor space. This approach is computationally efficient (it requires only k+1 models run for a k-dimensional input parameter space) but it suffers by a lack of precision. In fact, because the size Δx of the input parameter change in OAT is typically some fraction of (e.g., 1-10%) of the factor range, the method actually produces larger-scale trends (lower frequency variations) in the response surface. Moreover, OAT does not capture parameters interactions. But if OAT is used in the Morris contest, it becomes an optimal tool for sensitivity analysis.

Each model input $X_i, i = 1, \dots, k$ varies across z selected levels in the domain of input parameters. The space of experimentation Ω is thus a k-dimensional z-level grid. Like in Sobol analysis parameters have a uniform distribution in their uncertainty range and

more over their intervals are normalized to the [0,1] interval and later transformed from the unit hypercube to their actual distributions. It is the moment to define the Elementary Effect of the *i*th input parameter:

$$d_i(\boldsymbol{X}) = \left(\frac{y(X_1, \dots, X_{i-1}, X_i + \Delta, X_{i+1}, \dots, X_k) - y(\boldsymbol{X})}{C} \right) \qquad (3.3.1)$$

- Δ is a value in $\left\{ \frac{1}{(z-1)}, \dots, 1 - \frac{1}{z-1} \right\}$. An intelligent choice for z and Δ is z even and Δ equal to $\frac{z}{[2(z-1)]}$. This choice gives a certain symmetric treatment of inputs, which is surely desirable (for details see Morris, 1991).
- z is the arbitrary number determining Δ and as just said it will be an even number. In these studies $z = 4$ and $z = 8$ will be chosen.
- $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ is any selected value in Ω in a way that the transformed point $(\boldsymbol{X} + \boldsymbol{e}_i \Delta)$ remains in the domain Ω for every index $i = 1, \dots, k$.
- $\boldsymbol{e}_i$is a vector of zeros except for a unit as its *i*th component.

The finite distribution of Elementary Effects related to the *i*th factor comes from the random sampling of different $\boldsymbol{X}$ from Ω and its notation is F_i, i.e. $d_i(\boldsymbol{X}) \sim F_i$.The number of components of each F_i will be $z^{k-1}[z - \Delta(z-1)]$.

The sensitivity measures μ and σ introduced before represent respectively the mean and the standard deviation of the distribution F_i. Originally Morris first approach to estimate these quantities was to sample *r* Elementary Effects from each F_i by means of an efficient design that constructs *r* trajectories of (*k*+1) points in the input space, each supplying *k* Elementary Effects, one per input parameter. The total computational cost is thus $r(k+1)$ simulations. So more *r* is high and more the analysis will result precise.

But this first method sometimes was prone to II Type errors, so it can fail in the identification of an influential parameter for the model. In other words, if the distribution F_i contains both positive and negative elements, i.e. if the model is not monotonic, some effects may delete each other out and as a consequence there will be low μ values even for an important factor. The standard deviation σ helps to identify this type of error (if there are low μ values and high σ values probably I have a non-

monotonic model with Elementary Effects of different sign). At the same time, it is remarkable the resilience against I Type errors.

Morris method is established on the construction of *r* trajectories in the input space, normally they are between 10 and 50; trajectories are built by firstly generating a random starting point for every trajectory and secondly by moving one factor at time (OAT technique) in a random order. This design could bring to a non-optimal coverage of the input space, particularly for models with a large number of input factors.

The idea of the improvement brought by F.Campolongo et al. is to choose the *r* trajectories in such a way as to maximise their dispersion in the input space. The first move is the generation of a high number of *M* Morris trajectories ($M \sim 500 \div 1000$), the second move is the arbitrary decision about *r* value ($r \sim 5 \div 50$), where *r* value is the number defining the *r* trajectories with the highest '*spread*'. The meaning of *spread* is in closed in the following definition of distance d_{ml} between a couple of trajectories *m* and *l*:

$$d_{ml} = \begin{cases} \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \sqrt{\sum_{z=1}^{k} \left[X_i^m(z) - X_j^l(z)\right]^2} & for\ m \neq l \\ 0 & otherwise \end{cases} \qquad (3.3.2)$$

where *k* is the number of input factors and $X_i^m(z)$ represents the *z*th coordinate of the *i*th point of the *m*th Morris trajectory. In other simpler words d_{ml} is the sum of the geometric distances between all the couples of points of the two fixed trajectories.

So the best *r* trajectories are those that maximise the distance d_{ml} among them.

Another refinement which is present in this study is the introduction of the μ^* measure, which lifts problems with II type error by solving problems of the effects of opposite signs affecting non-monotonic models. μ^* is the estimate of the mean of the distribution of the absolute values of the Elementary Effects and its denotation is $|d_i(\boldsymbol{X})| \sim G_i$. The only drawback coming from the utilization of μ^* is the loss of information on the sign of the effect and this is the reason because estimates of μ will be presented coupled to μ^*estimates (moreover there is no extra computational cost).

3.4 METHOD OF ANALYSIS

In the following pages, by the utilization of simple examples, it will be demonstrated that two Morris and Sobol methods have to be used in couple to avoid any type of shortcomings. While Sobol ‘s method attempts to provide more general results by characterizing the nature of response sensitivity over the entire factor space, Morris ‘s method evaluates the local sensitivity coefficients (which are the elementary effects for each factor at multiple points across the factor space) and analyses the distributional properties of these values to assess the global sensitivities of the model response in relation to the individual factors. Both methods are necessary to have a clear idea of the surface response.

1. ***First Example***

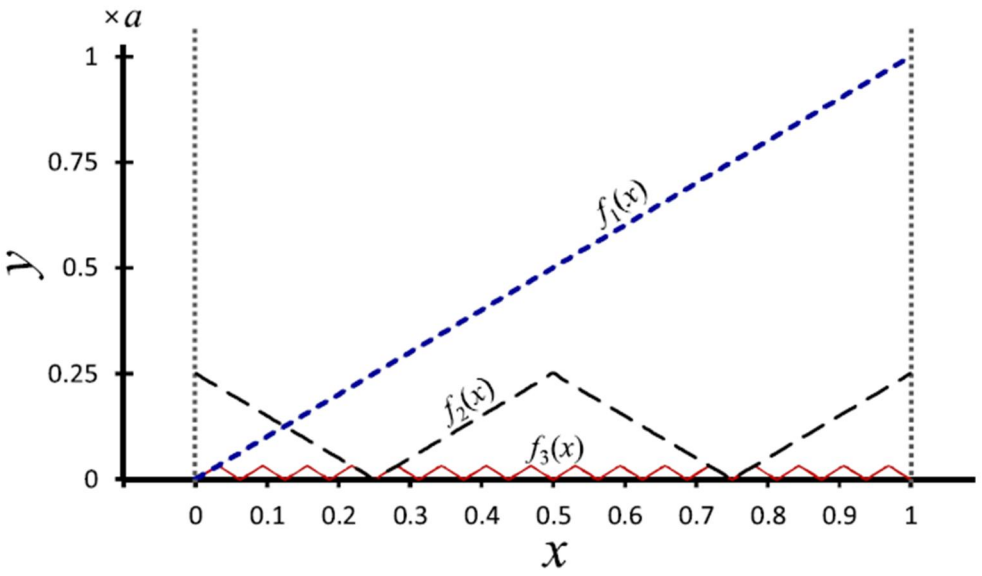


Figure 3.1. Response surfaces of the first example.

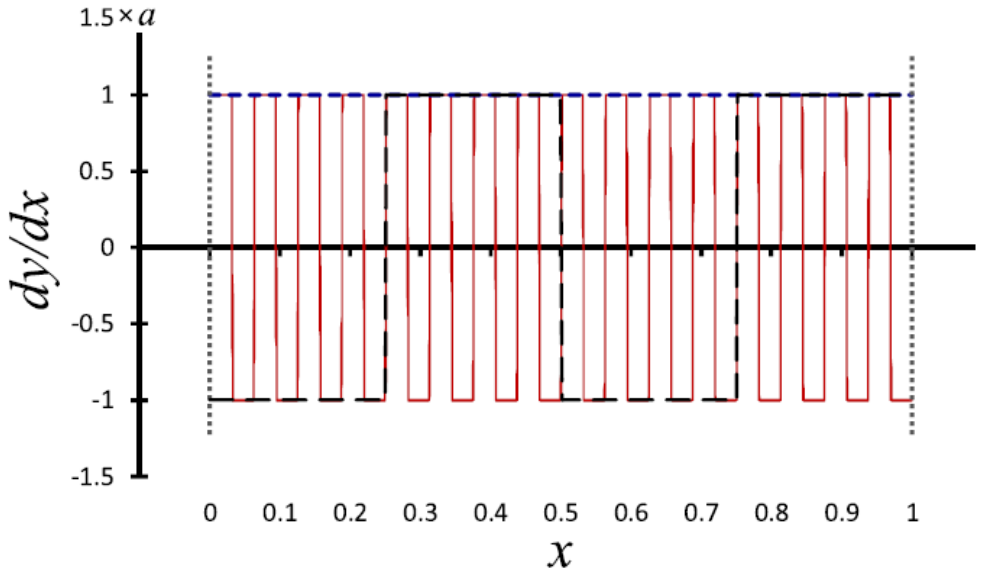


Figure 3.2. Derivative functions of the response surfaces of the first example.

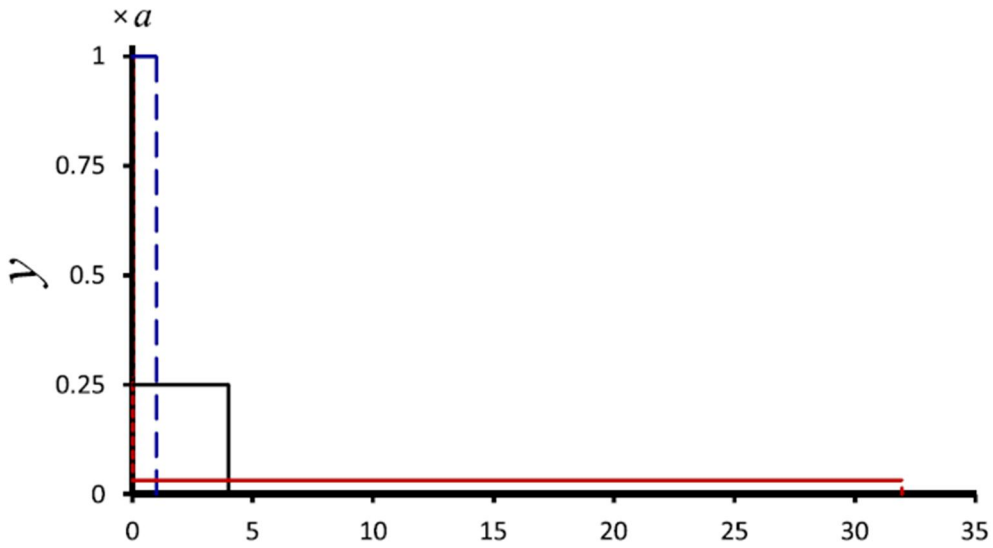


Figure 3.3. Probability density functions of the response values of the first example.

Table 3.1. Performance of conventional sensitivity measures, first example.

function	$D(y)$	$E\left(\left\|\frac{\partial y}{\partial x}\right\|\right)$	$E\left(\left(\frac{\partial y}{\partial x}\right)^2\right)$	$D\left(\frac{\partial y}{\partial x}\right)$
$f_1(x)$	$\frac{a^2}{12}$	a	a^2	0
$f_2(x)$	$\frac{a^2}{192}$	a	a^2	a^2
$f_3(x)$	$\frac{a^2}{12288}$	a	a^2	a^2

Figure (3.1) presents three functions that cover radically different output ranges. It can be seen how they have constant and identical values of the local sensitivities across the factor range (except at singular points where the derivatives change sign) and at the same time completely different response surface PDFs (see Fig. 3.2 and Fig. 3.3). This means that f_1 has the highest sensitivity of the response *y* respect to variations in factor x and it is confirmed by the fact that changes in *x* control a larger range of the output.

In this example the measure of D(y) states $f_1(x)$ is respectively 16 and 1026 times more sensitive to factor *x* than functions $f_2(x)$ and $f_3(x)$; however, measures $E\left(\left|\frac{\partial y}{\partial x}\right|\right)$ and $E\left(\left(\frac{\partial y}{\partial x}\right)^2\right)$ state that all functions are sensitive at the same way, while measure $D\left(\frac{\partial y}{\partial x}\right)$ does not recognize differences between $f_2(x)$ and $f_3(x)$. In the light of these considerations, in this example the results supplied by D(y) would be more consistent for a sensitivity analysis.

2. ***Second Example***

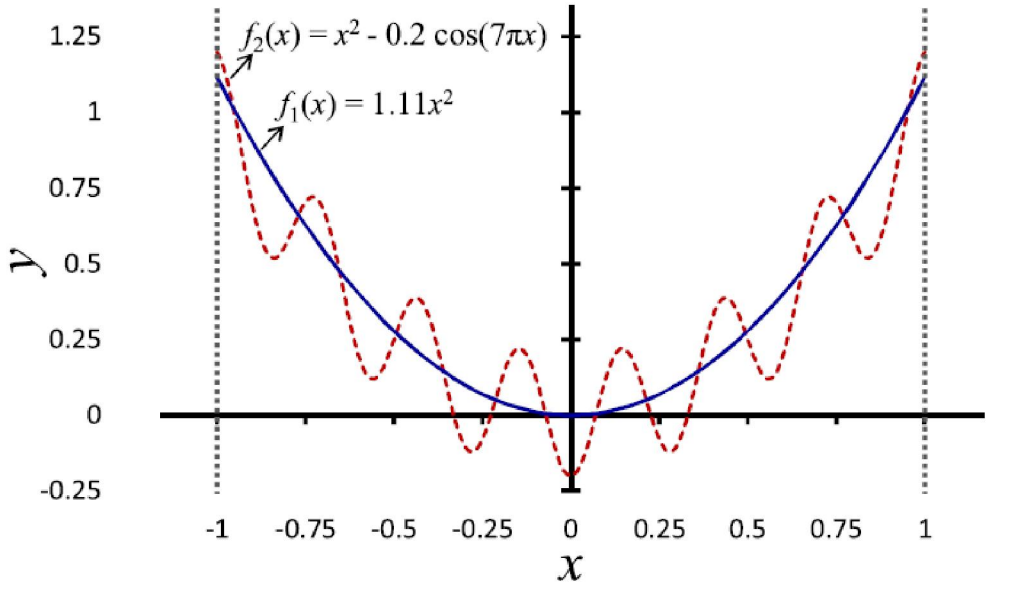


Figure 3.4. Response surfaces of the second example.

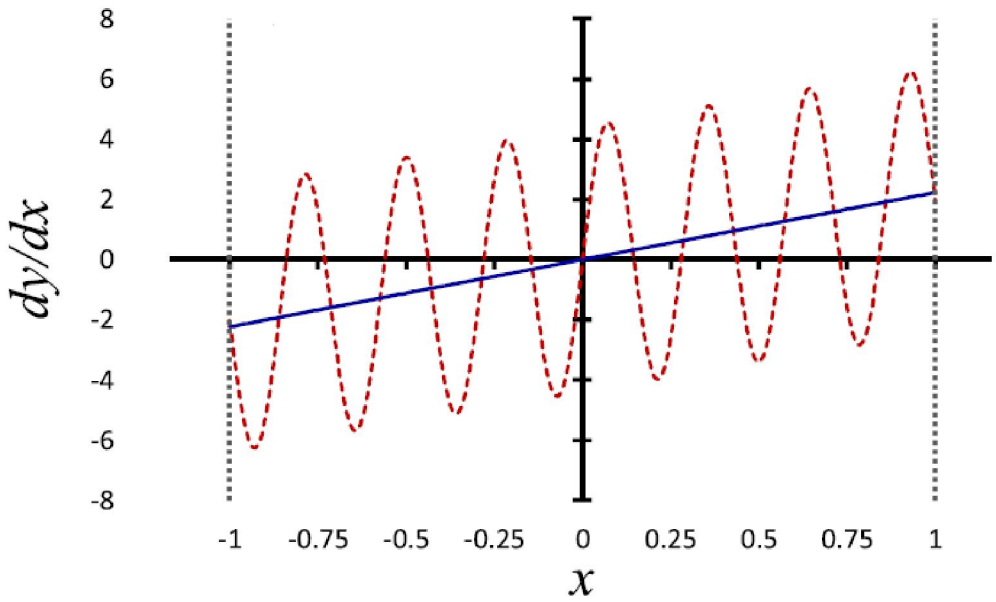


Figure 3.5. Derivative functions of the response surfaces of the second example.

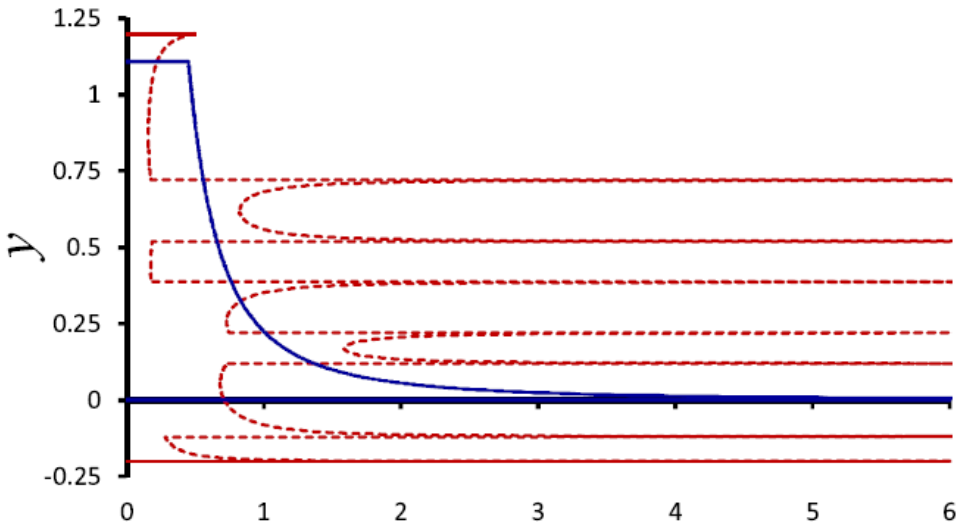


Figure 3.6. Probability density functions of the response values of the second example.

Table 3.2. Performance of conventional sensitivity measures, second example.

function	$D(y)$	$E\left(\left\|\frac{\partial y}{\partial x}\right\|\right)$	$E\left(\left(\frac{\partial y}{\partial x}\right)^2\right)$	$D\left(\frac{\partial y}{\partial x}\right)$
$f_1(x)$	0.11	1.11	1.64	1.64
$f_2(x)$	0.11	3.01	11.80	11.80

This example contains two functions having the same average absolute local sensitivity equal to 0 (figure 3.5). Although two very different response surfaces PDFs result, there is an identical *variance* of the response *y* for the two functions. This happens because the variance approach measures overall variability of the response but it is not sensitive to the structure of the surface- i.e., how the values of the response surface are organized in the factor space. Consequently, variance-based methods are not able to take into account important structural information such as multimodality. In contrast Morris type methods indicate that $f_2(x)$ is significantly more sensitive to factor *x* than $f_1(x)$ but there is a danger of exceeding Δx values and as a consequence the surface response obtained would not be truthful. This complication has been called "*scale issue*".

In conclusion the variance-based Sobol approach is based entirely on characterizing the global variance of model response, and its decomposition into components associated with individual contributing factors. As such, it is unable to distinguish between response surface structures that have identical global variance of model response but different distributions and spatial organizations (response surface structures) of the model response and its derivatives. The Morris approach and its extensions attempt to globally aggregate local sensitivity information (first-order partial derivatives) across the factor space. However, all implementations of this approach are prone to the *scale issue*, and the step size of the analysis could have a significant impact on the conclusions about underlying sensitivities.
So the combined utilization of Sobol and Morris techniques permit to overcome shortcomings of both about surfaces responses representations.

4 RESULTS

This chapter is many-sided and multifaceted. The starting point of this work was a blank page, in the sense there were no previous studies related to the case taken in consideration. Anyway the full model was already implemented and also the global quantities and parameters describing them were already suggested (global quantities of interest were in function of 27 parameters without considering grid and fracture network parameters). So the first approach was to identify most influent parameters for some global quantities. Initially the analysis was carried out only by means of Sobol indices. At a certain point it became clear that Sobol indices were producing the influence regarding only the variance of the mean of global quantities, while the influence of input parameters on the absolute value of the mean of global quantities was not taken into consideration. From then also Morris analysis was implemented in couple with Sobol analysis.

So initial experiments are here reported as “Preliminary Results” because they are an optimal example to understand the right way to execute a sensitivity analysis without shortcomings and because they were propaedeutic for the final definitive case.

Then there is a part called “Results of Sensitivity Analysis” which show final results regarding to initial questions that they were supposed to be answered.

Finally, the last section is named “Probability Distributions of Target Variables” and it is concerned with the analyses of surface responses through PDFs tools.

4.1 PRELIMINARY RESULTS

As first thing it is reported a table with all input parameters in standard configuration, i.e. fixed values of parameters when they are considered certain (see Table 2.1 for symbols meanings).

Table 4.1. Tabulation of standard-configuration values of input parameters.

Spatial Parameters		
k_{SA_m}	1×10^{-15}	$[m^2]$
k_F	1×10^{-8}	$[m^2]$
k_{SA_f}	5×10^{-8}	$[m^2]$
k_T	1×10^{-14}	$[m^2]$
k_{OV}	1×10^{-13}	$[m^2]$
k_{CR}	1×10^{-18}	$[m^2]$
Φ_{SA_m}	0.1	$[-]$
Φ_F	0.5	$[-]$
Φ_{SA_f}	0.5	$[-]$
Φ_T	0.25	$[-]$
Φ_{OV}	0.1	$[-]$
Φ_{CR}	0.1	$[-]$
Operational and Geometric Parameters		
Z_{BOR}	1000	$[m]$
Z_{RES}	60	$[m]$
Z_{CR}	25	$[m]$
b_{SA_f}	0.005	$[m]$
b_F	0.01	$[m]$
P_{INJ}	500	$[bar]$
t_{END}	792000	$[s]$
t_{INJ}	72000	$[s]$

$M_{H_2 fraction}$	0.001	$\left[\frac{kg_{H_2}}{kg_{H_2O}}\right]$
L_{INJ}	100	$[m]$
Z_{TARGET}	200	$[m]$
Fluid properties and others		
μ_{H_2O}	1×10^{-3}	$\left[\frac{kg}{ms}\right]$
$D_m^{H_2,H_2O}$	1×10^{-9}	$\left[\frac{m^2}{s}\right]$
ρ_{H_2O}	1000	$\left[\frac{kg}{m^3}\right]$
g	9.81	$\left[\frac{m}{s^2}\right]$

First Case:

The first focus was on those input parameters that seemed more determinant for global quantities of interest. So four permeabilities and the injection pressure were chosen. The range of variation was chosen arbitrarily according to the standard range those parameters could assume in real cases.

Table 4.2. First case choice of uncertain input factors and their range of variation.

Input Parameter	Lower Bound	Upper Bound
k_{SA_m}	1×10^{-16}	1×10^{-14}
k_F	1×10^{-8}	1×10^{-6}
k_{SA_f}	1×10^{-8}	1×10^{-6}
k_T	1×10^{-14}	1×10^{-12}
P_{INJ}	600	800

The standard values of all porosities were formerly set to 0.1 (case 1_A). This was unrealistic especially for Φ_{SA_f} and Φ_F, because they are typically higher than other

porosities. So the first thing done was to set to 0.5 Φ_{SA_f} and Φ_F porosities (case 1_B). Then full model was run in order to see if and how results about global quantities Θ were changing. For what concerns the choice of total degree *w*, the criterion was initially to adopt *w*=2 and *w*=3 degrees and if the respective results were not in agreement go for *w*=4 degree, otherwise stop to *w*=3. In fact, the full model simulations needed for the creation of the surrogate model of degree *w*=2 and *w*=3 is respectively 61 and 241, while for order *w*=4 is 781, a considerable number.

Table 4.3. Simulation results for case 1_A and 1_B and for total degree w=2.

$w = 2$	$M_{H_2,MAX}$		$M_{H_2,END}$	
	1_A	1_B	1_A	1_B
Θ	554	554	352	352
$\sqrt{D}$	2.7×10^2	2.7×10^2	1.6×10^2	1.6×10^2
D	7.5×10^4	7.3×10^4	2.5×10^4	2.6×10^4

Table 4.4. Simulation results for case 1_A and 1_B and for total degree w=3.

$w = 3$	$M_{H_2,MAX}$		$M_{H_2,END}$	
	1_A	1_B	1_A	1_B
Θ	558	555	347	351
$\sqrt{D}$	2.7×10^2	2.8×10^2	1.6×10^2	1.6×10^2
D	7.3×10^4	7.7×10^4	2.6×10^4	2.4×10^4

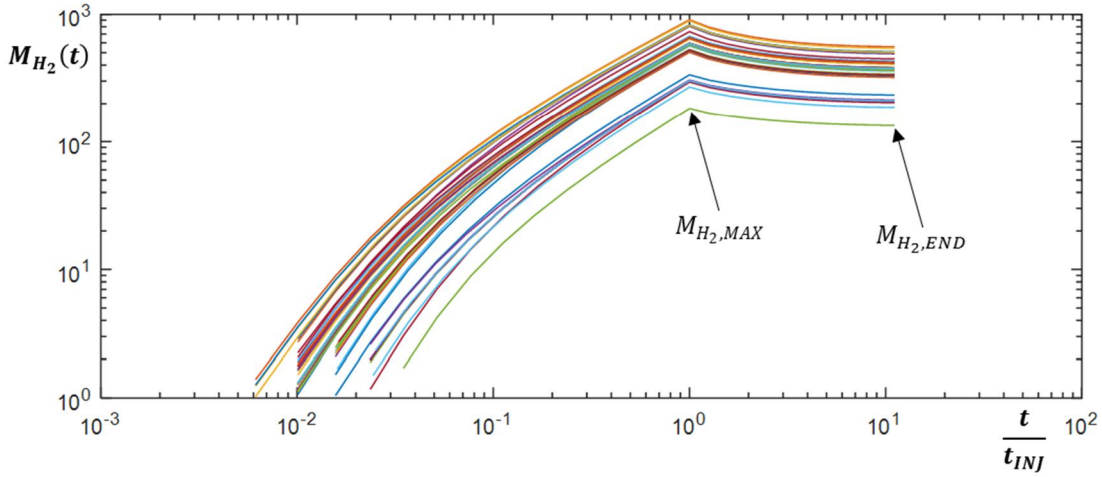


Figure 4.1 First Case representation of full model responses for w =2 degree.

All responses seem to assume the same trend, with accumulated mass growing until the end of injection time and then slowly decreasing. The values above are nearly the same, so for the sake of authenticity the next results will be only based on 1_B subcase. Here below principal and total Sobol indices are reported. The analysis of both Sobol indices and absolute values of the variance is not a superfluous choice, because Sobol indices reveal the magnitude of contribute to the total variance but they don't show its absolute value, which must also be kept in consideration.

Table 4.5. Principal Sobol Indices of case 1_B

Principal Indices	$M_{H_2,MAX}$		$M_{H_2,END}$	
	$w = 2$	$w = 3$	$w = 2$	$w = 3$
$S_{k_{SA_m}}$	0.000	0.000	0.000	0.002
$S_{k_{SA_f}}$	0.000	0.000	0.000	0.001
S_{k_F}	0.002	0.003	0.003	0.004
S_{k_T}	0.939	0.937	0.927	0.927
$S_{P_{INJ}}$	0.049	0.049	0.060	0.058

Table 4.6 Total Sobol Indices of case 1_B

Total Indices	$M_{H_2,MAX}$		$M_{H_2,END}$	
	$w = 2$	$w = 3$	$w = 2$	$w = 3$
$S^T_{k_{SA_m}}$	0.000	0.000	0.000	0.000
$S^T_{k_{SA_f}}$	0.000	0.000	0.000	0.002
$S^T_{k_F}$	0.002	0.003	0.003	0.004
$S^T_{k_T}$	0.949	0.948	0.937	0.937
$S^T_{P_{INJ}}$	0.059	0.059	0.070	0.068

First results obtained state clearly that the most influent parameter for the two global quantities considered is the permeability of the target aquifer above the reservoir; P_{INJ} seems to play some kind of role, but its value is nearly twenty times lower than k_T value, so more cases will be implemented to have a better knowledge of its importance. Another evident observation is the similarity between Principal and Total indices that outlines the lack of interaction between different parameters. Furthermore, the almost equality between *w* =2 and *w* = 3 results are an index of high linearity of the surrogate model obtained through the g-PCE, because the nearly equality means that it is sufficient to stop to *w*=2 degree in order to obtain a surrogate model that approximates well the full model response. This consideration is supported directly by linear Sobol indices (here not reported) which are very similar to principal Sobol indices, indicating precisely the high linearity of the surrogate model and so of the response.

Second Case:

Next step was the investigation of some porosities, so five porosities and target permeability were implemented in this second case. For the moment P_{INJ} was not present in this simulation because from previous case its influence seemed secondary or near negligible and because every parameter increases significantly computational costs.

Table 4.7 Second case choice of uncertain input factors and their range of variation.

Input Parameter	Lower Bound	Upper Bound
Φ_{SA_m}	0.05	0.3
Φ_F	0.5	0.7
Φ_{SA_f}	0.5	0.7
Φ_T	0.2	0.5
Φ_{CR}	0.05	0.3
k_T	1×10^{-14}	1×10^{-12}

Even if computational costs would have been increased, in order to increase the veracity level of these results it has been decided to extend the analysis to $w = 4$ degree:

- $w = 2$ ⟶ Number of full model runs = 85
- $w = 3$ ⟶ Number of full model runs = 389
- $w = 4$ ⟶ Number of full model runs = 1433

Here follow results.

Table 4.8. Second case simulation results for the peak of mass response.

	$M_{H_2,MAX}$		
$degree$	$w = 2$	$w = 3$	$w = 4$
Θ	284	279	288
$\sqrt{D}$	1.5×10^2	1.5×10^2	1.5×10^2
D	2.4×10^4	2.4×10^4	2.3×10^4

Table 4.9. Second case simulation results for the mass at end response.

	$M_{H_2,END}$		
$degree$	$w = 2$	$w = 3$	$w = 4$
Θ	162	177	146
$\sqrt{D}$	8.7×10^1	8.3×10^1	9.5×10^1
D	7.6×10^3	6.9×10^3	9.1×10^3

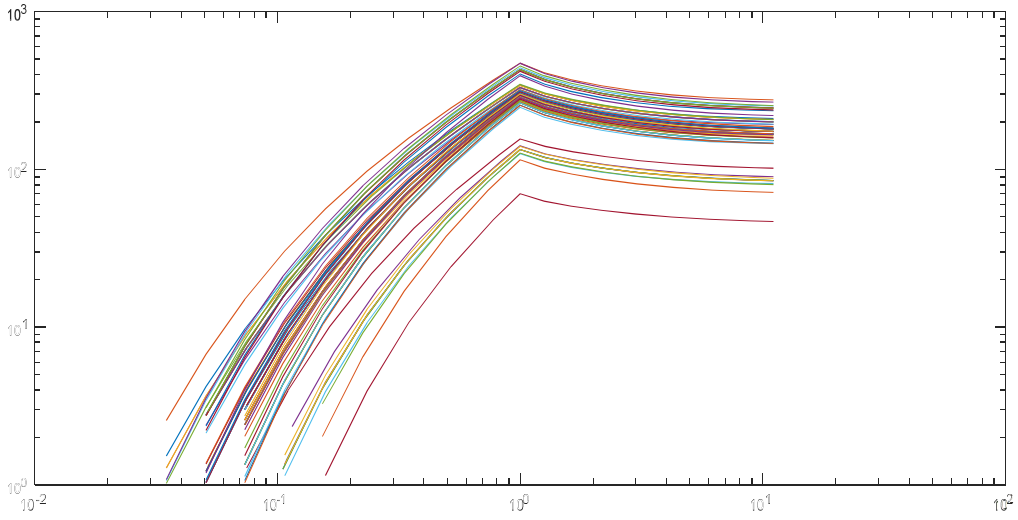


Figure 4.2. Second case representation of full model responses for w =2 degree.

Results are counter-intuitive. The change of the mean and of the variance of two global quantities is big: more or less new values of the mean are the half and new values of the variance are one third of the previous case. At the same time trends of global quantities are equal to the first case. Before starting to find an explanation it is wise to look at Sobol indices.

Table 4.10. Principal Sobol Indices of second case.

Principal Indices	$M_{H_2,MAX}$			$M_{H_2,END}$		
	$w = 2$	$w = 3$	$w = 4$	$w = 2$	$w = 3$	$w = 4$
$S_{\Phi_{SA_m}}$	0.039	0.034	0.048	0.061	0.079	0.037
$S_{\Phi_{CR}}$	0.003	0.003	0.004	0.005	0.006	0.002
S_{Φ_T}	0.010	0.011	0.012	0.005	0.007	0.002
$S_{\Phi_{SA_f}}$	0.000	0.000	0.000	0.000	0.000	0.006
S_{Φ_F}	0.000	0.000	0.000	0.000	0.000	0.000
S_{k_T}	0.940	0.944	0.926	0.913	0.896	0.942

Table 4.11. Total Sobol Indices of second case.

Total indices	$M_{H_2,MAX}$			$M_{H_2,END}$		
	$w = 2$	$w = 3$	$w = 4$	$w = 2$	$w = 3$	$w = 4$
$S^T_{\Phi_{SAm}}$	0.043	0.038	0.053	0.066	0.087	0.045
$S^T_{\Phi_{CR}}$	0.004	0.003	0.005	0.005	0.007	0.004
$S^T_{\Phi_T}$	0.014	0.014	0.016	0.006	0.008	0.003
$S_{\Phi_{SA_f}}$	0.000	0.000	0.000	0.000	0.002	0.008
$S^T_{\Phi_F}$	0.000	0.000	0.000	0.000	0.000	0.002
$S^T_{k_T}$	0.948	0.952	0.935	0.929	0.906	0.949

First observation is about the dominance of permeability k_T over other parameters: with values of both principal and total indices k_T remains the most influent factor for both global quantities taken on examination. Another parameter with notable values is the porosity of the reservoir matrix Φ_{SA_m}. Considerations about responses linearity and low-interaction between parameters of first case are also valid in this case. From these results a fundamental statement comes as a consequence: Injection Pressure doesn't affect too much the variance of the responses but instead it changes dramatically the absolute value of them. It is a natural consequence if one looks at standard and fixed

value $P_{INJ} = 500\ [bar]$ that injection pressure assumes in this case. In fact, when P_{INJ} is an uncertain factor varying between 600 and 800 [bar] global responses exhibit high values, while when it is fixed to 500 [bar] like in this case they show low values. In order to have the final proof of this observation a third case is implemented.

Third Case:

Parameters of third case are the same of those of second case except for injection pressure which is added to the analysis to verify the previous statements. What is expected is that with the addition of Injection pressure the mean of global quantities and their variance will increase and they will be around first case values.

Table 4.12. Third case choice of uncertain input factors and their range of variation.

Input Parameter	Lower Bound	Upper Bound
Φ_{SA_m}	0.05	0.3
Φ_F	0.5	0.7
Φ_{SA_f}	0.5	0.7
Φ_T	0.2	0.5
Φ_{CR}	0.05	0.3
k_T	1×10^{-14}	1×10^{-12}
P_{INJ}	600	800

Also here $w = 4$ degree has been implemented in order to maintain the best compromise with veracity level and computational costs (here below reported).

- $w = 2$ ⟶ Number of full model runs = 113
- $w = 3$ ⟶ Number of full model runs = 589
- $w = 4$ ⟶ Number of full model runs = 2437

And now results are presented.

Table 4.13. Third case simulation results for the peak of mass response.

	$M_{H_2,MAX}$		
$degree$	$w = 2$	$w = 3$	$w = 4$
Θ	485	496	487
$\sqrt{D}$	2.6×10^2	2.6×10^2	2.6×10^2
D	6.6×10^4	6.8×10^4	6.8×10^4

Table 4.14. Third case simulation results for the mass at end response.

	$M_{H_2,END}$		
$degree$	$w = 2$	$w = 3$	$w = 4$
Θ	297	277	281
$\sqrt{D}$	1.5×10^2	1.4×10^2	1.4×10^2
D	2.1×10^4	2.1×10^4	2.1×10^4

As it was predicted values of mean and variance returned similar to the first case and so the previous considerations are confirmed. Now look at Sobol indices.

Table 4.15. Principal Sobol Indices of third case.

Principal Indices	$M_{H_2,MAX}$			$M_{H_2,END}$		
	$w = 2$	$w = 3$	$w = 4$	$w = 2$	$w = 3$	$w = 4$
$S_{\Phi_{SAm}}$	0.014	0.014	0.015	0.028	0.027	0.028
$S_{\Phi_{CR}}$	0.003	0.000	0.002	0.002	0.012	0.009
S_{Φ_T}	0.010	0.008	0.009	0.003	0.011	0.002
$S_{\Phi_{SA_f}}$	0.000	0.000	0.000	0.000	0.000	0.000
S_{Φ_F}	0.000	0.000	0.000	0.000	0.000	0.000
S_{k_T}	0.940	0.895	0.899	0.884	0.888	0.885
$S_{P_{INJ}}$	0.055	0.062	0.060	0.070	0.046	0.059

Table 4.16. Total Sobol Indices of third case.

Total indices	$M_{H_2,MAX}$			$M_{H_2,END}$		
	$w = 2$	$w = 3$	$w = 4$	$w = 2$	$w = 3$	$w = 4$
$S^T_{\Phi_{SA_m}}$	0.016	0.017	0.018	0.030	0.031	0.031
$S^T_{\Phi_{CR}}$	0.002	0.001	0.001	0.002	0.012	0.010
$S^T_{\Phi_T}$	0.013	0.012	0.012	0.004	0.012	0.011
$S_{\Phi_{SA_f}}$	0.000	0.000	0.000	0.000	0.004	0.002
$S^T_{\Phi_F}$	0.000	0.000	0.000	0.000	0.004	0.003
$S^T_{k_T}$	0.919	0.914	0.918	0.898	0.898	0.897
$S^T_{P_{INJ}}$	0.066	0.076	0.072	0.081	0.054	0.056

Permeability of the target aquifer remains the most dominant parameter. A first important remark regards Φ_{SA_m}. In the third case, the addition of P_{INJ} to the list of uncertain parameters lowers Sobol indices values of the reservoir matrix. It is clear that all input factors should be taken as uncertain in order to have a truthful analysis of the influence of input parameters on the variance of global quantities. And it is also evident that Sobol analysis is not a good tool for identifying the most influent parameters for the mean of global quantities. Moreover, in addition to injection pressure, other mean-influent parameters could be present and in the contest of these studies the mean of global quantities is more important than their variance. So for these last reasons Morris analysis will be implemented.

Before introducing the next section, a last question has to be solved: “Do surrogate models of these three cases approximate respective full models?”. To answer it is necessary to develop scatter plots of some (number is arbitrary) random points. The procedure is the following: random sets of parameters are chosen and they are run using both full model and its surrogate. Then the two responses resulting are plotted: black segment represents full model solutions and red points surrogate model solutions. More red points are near full model segments and more surrogate models

are good approximations. And since surrogate models of degree $w = 2$ suite well full models there is no need of executing scatter plots for higher degrees surrogate models, because they surely suite better full models. Here are scatter plots.

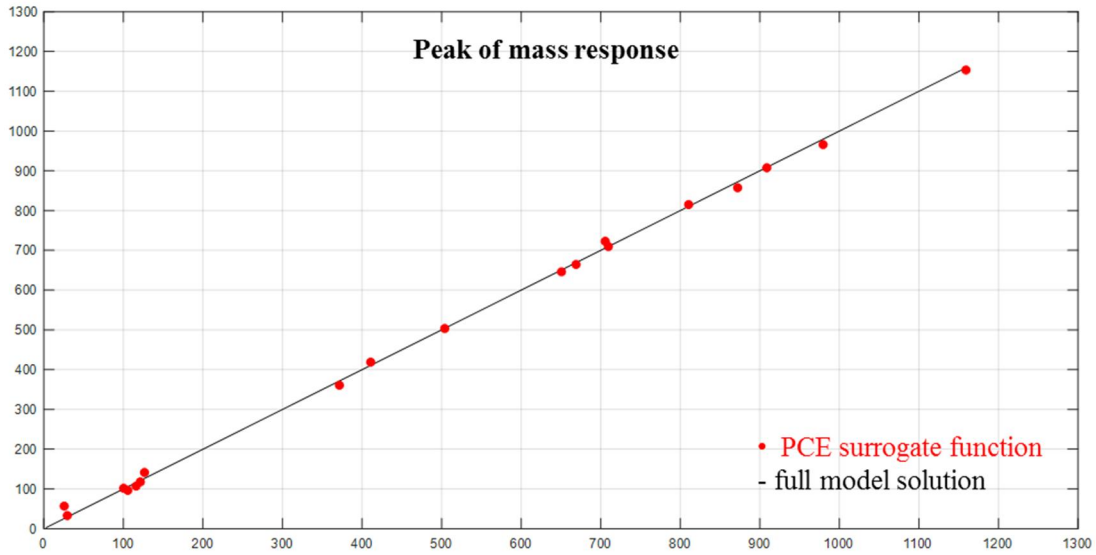


Figure 4.3. First case: scatter plot of "Peak of mass" response.

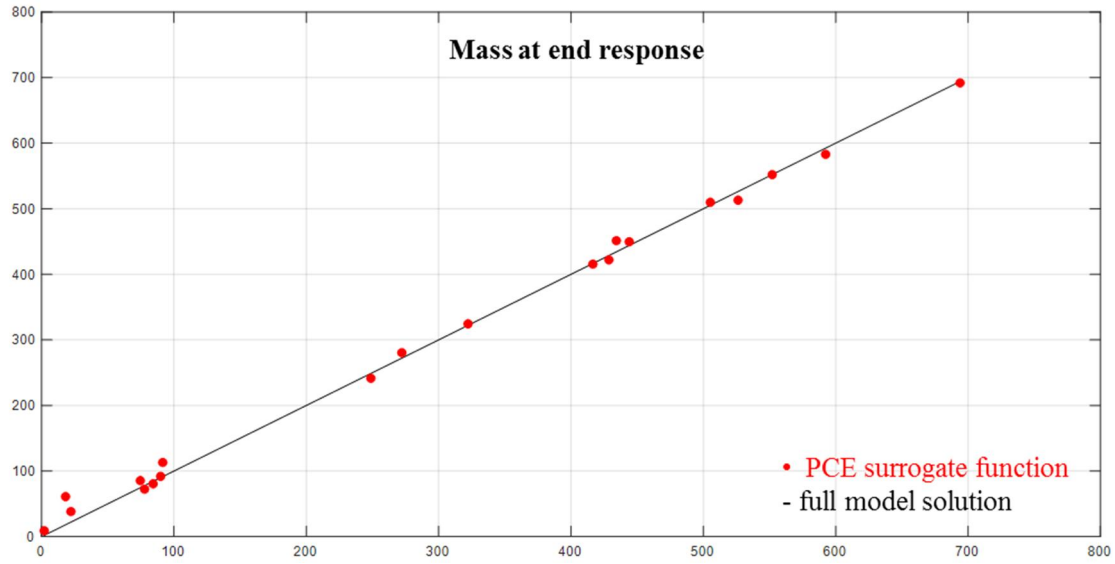


Figure 4.4. First case: scatter plot of "Mass at end" response.

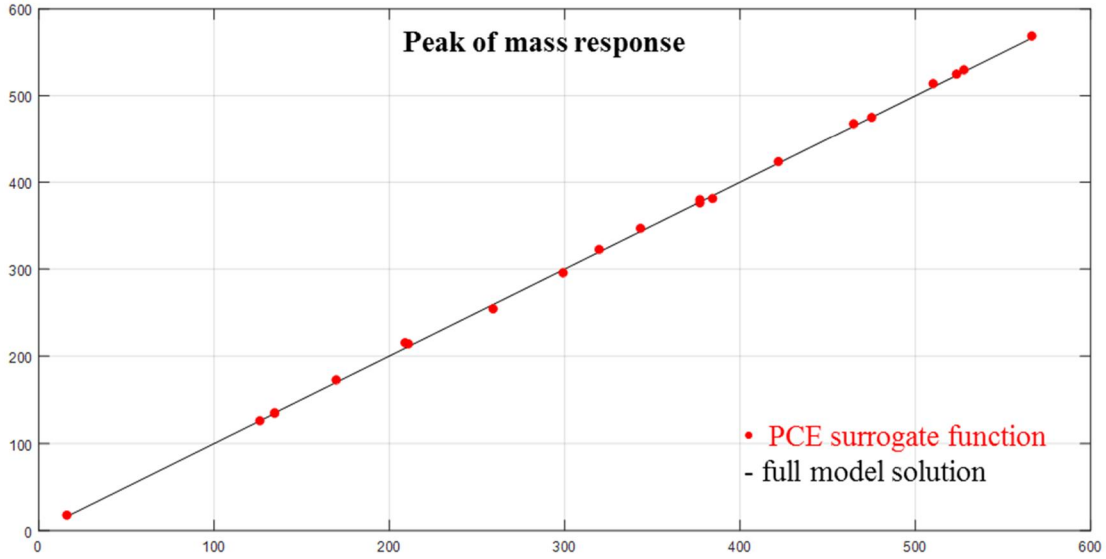


Figure 4.5. Second case: scatter plot of "Peak of mass" response.

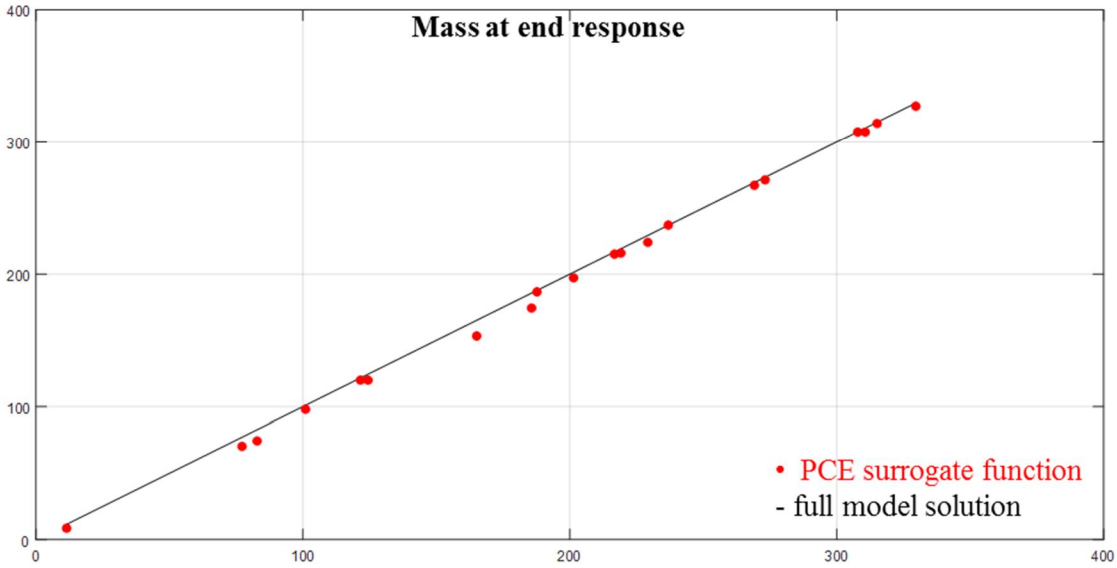


Figure 4.6. Second case: scatter plot of "Mass at end" response.

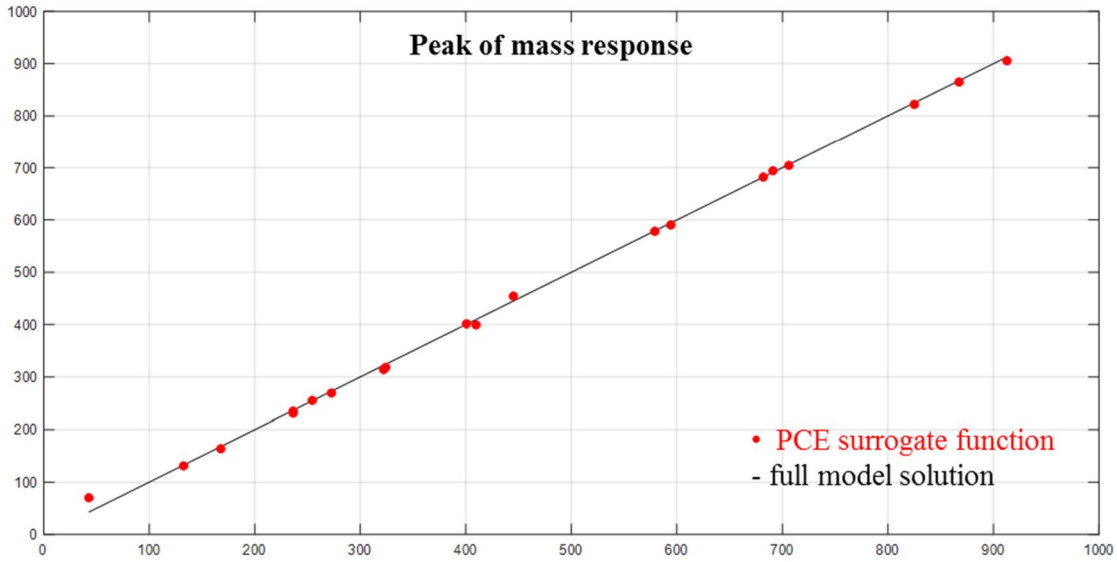


Figure 4.7. Third case: scatter plot of "Peak of mass" response.

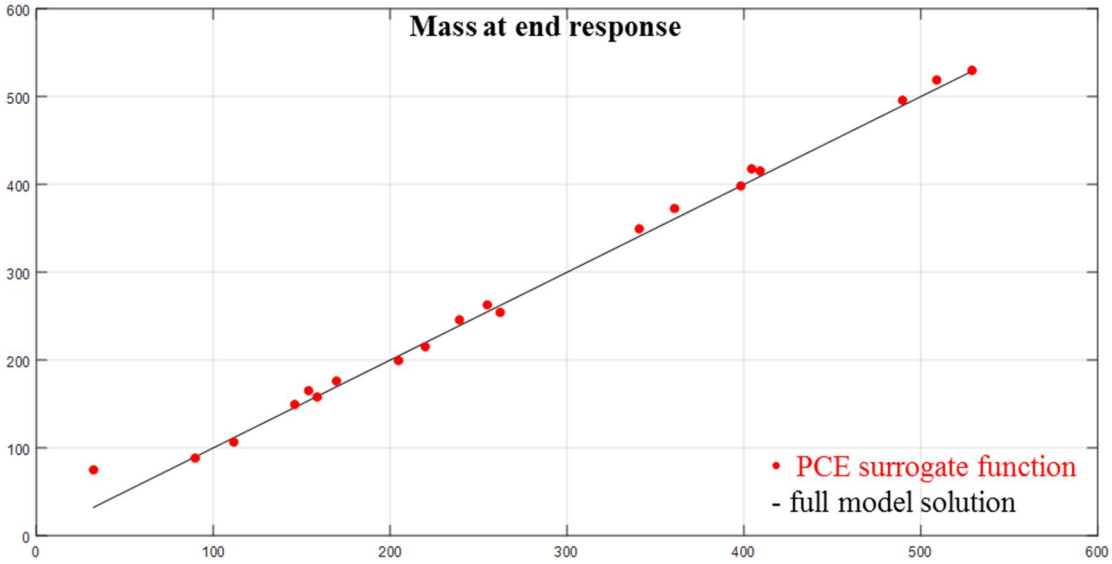


Figure 4.8. Third case: scatter plot of "Mass at end" response.

Scatter plots indicate a good approximation made by surrogate models, enough to confirm the validity of Sobol analyses. Next section is dedicated to the implementation of final analyses.

4.2 RESULTS OF SENSITIVITY ANALYSIS

In this will be structured as follows. Firstly, Morris analysis is applied to the problem, then two "final" cases are analysed through Sobol indices and results are matched with Morris results. As last PDFs of global responses are built.

Morris Analysis:

The theory, the implementation and the way of proceeding have been discussed in section 3.2 and the reader is sent here for relative information; here only the specific case and relative results are taken in examination.
In the first attempt the choice of uncertain parameters was the widest: nearly all parameters were chosen.

Table 4.17. First choice of parameters and bounds in Morris analysis.

Input Parameter	Lower Bound	Upper Bound
Φ_{SA_m}	0.05	0.3
Φ_F	0.5	0.7
Φ_{SA_f}	0.5	0.7
Φ_T	0.2	0.5
Φ_{CR}	0.05	0.3
Φ_{OV}	0.05	0.3
k_{SA_m}	1×10^{-16}	1×10^{-14}
k_F	1×10^{-8}	1×10^{-6}
k_{SA_f}	1×10^{-8}	1×10^{-6}
k_T	1×10^{-14}	1×10^{-12}
k_{OV}	1×10^{-14}	1×10^{-12}
k_{CR}	1×10^{-17}	1×10^{-20}
Z_{BOR}	900	1200
Z_{RES}	50	80
Z_{CR}	20	40
b_{SA_f}	0.001	0.01

b_F	0.005	0.01
P_{INJ}	600	800
t_{END}	780000	800000
t_{INJ}	70000	100000
$M_{H_2 fraction}$	0.0005	0.01
L_{INJ}	90	120
Z_{TARGET}	180	220

Results coming from this choice were inconsistent and deceptive because they were affected by grid problems (if some geometric and temporal parameters become uncertain the grid becomes no more reliable). So instead of showing useless and untruthful results the smartest choice is illustrated. In order to derivate useful but at the same time consistent results, parameter choice will lie on permeabilities, porosities and injection pressure.

Table 4.18. Definitive choice of parameters and bounds in Morris analysis.

Input Parameter	Lower Bound	Upper Bound
Φ_{SA_m}	0.05	0.3
Φ_F	0.5	0.7
Φ_{SA_f}	0.5	0.7
Φ_T	0.2	0.5
Φ_{CR}	0.05	0.3
Φ_{OV}	0.05	0.3
k_{SA_m}	1×10^{-16}	1×10^{-14}
k_F	1×10^{-8}	1×10^{-6}
k_{SA_f}	1×10^{-8}	1×10^{-6}
k_T	1×10^{-14}	1×10^{-12}
k_{OV}	1×10^{-14}	1×10^{-12}
k_{CR}	1×10^{-17}	1×10^{-20}
P_{INJ}	600	800

Global quantities analysed are the same of those chosen for Sobol analyses ($M_{H_2,MAX}$ and $M_{H_2,END}$). Other specifications are here below reported.

Table 4.19. Morris analysis specifications.

Number of factors	k	13
Discretization number	z	4
Generation number	N	2000
Number of trajectories	r	20
Full model runs	$r(k+1)$	280

Also this choice comes from a trading between computational costs and precision is present: more *r* and *N* are high more the analysis will be accurate but at the same time computational costs will increase. About *z* the point is more controversial. *z* defines discretization pass Δ because $\Delta = \frac{z}{[2(z-1)]}$. So results are dependent from *z* choice. $z = 4$ assures a good discretization and next results will be based on but in order to see how much results change modifying *z*, a results comparison will be act between $z = 4$ and $z = 8$ case.

The first aspect under investigation is the monotony of the response function respect to the several uncertain input factors. So the mean and the absolute mean of finite distribution of elementary effects are plotted for each parameter. If two points of a certain factor coincide then the response function is monotonic respect to that factor.

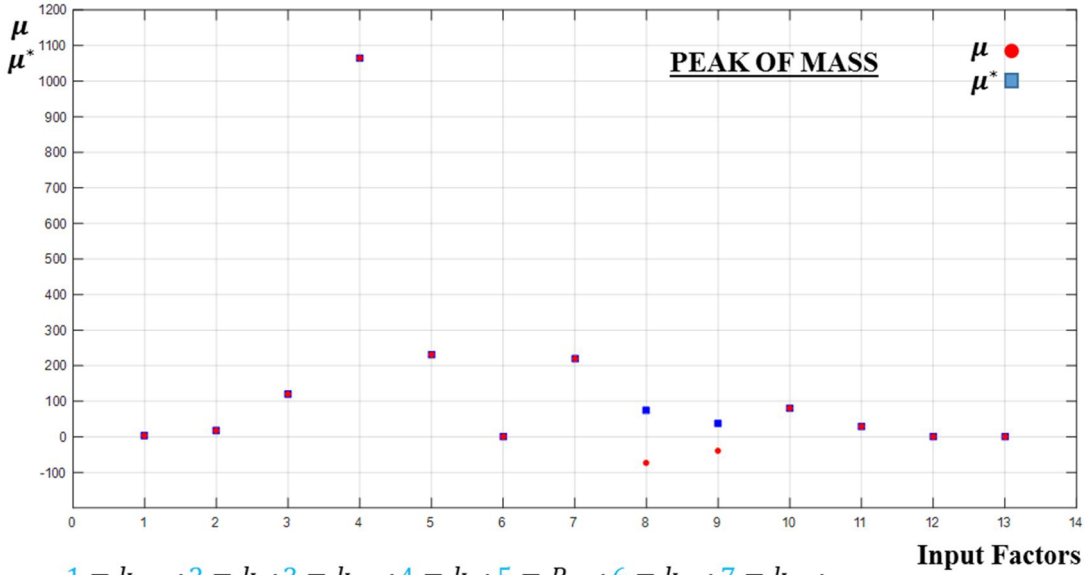


Figure 4.9. Morris Analysis: mean and absolute mean of F_i and G_i for "Peak of mass".

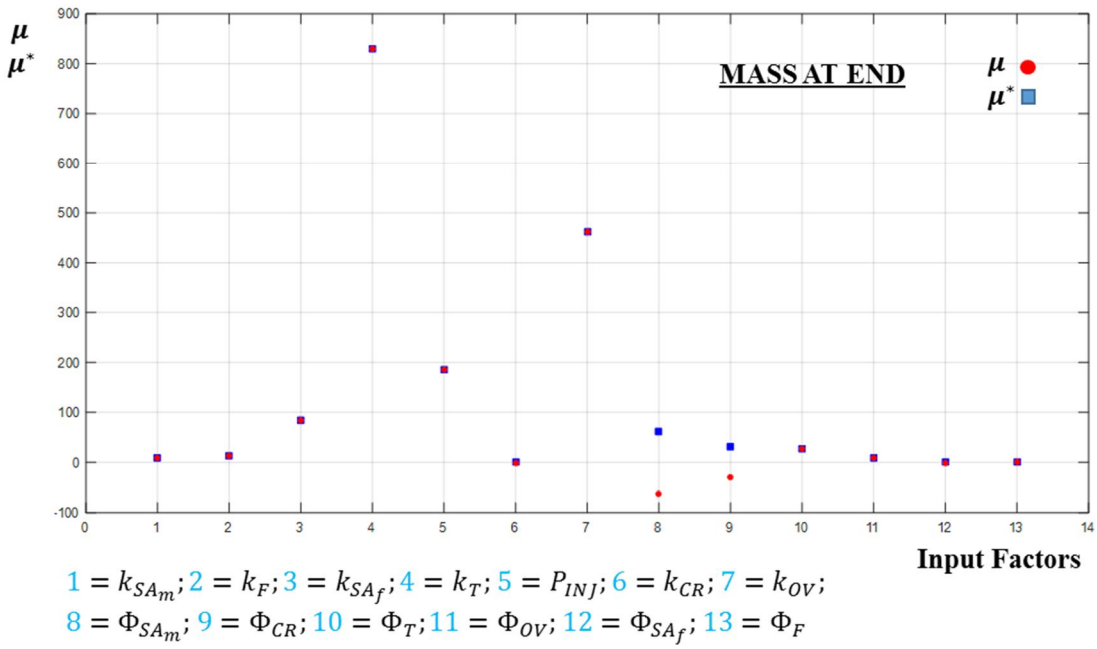
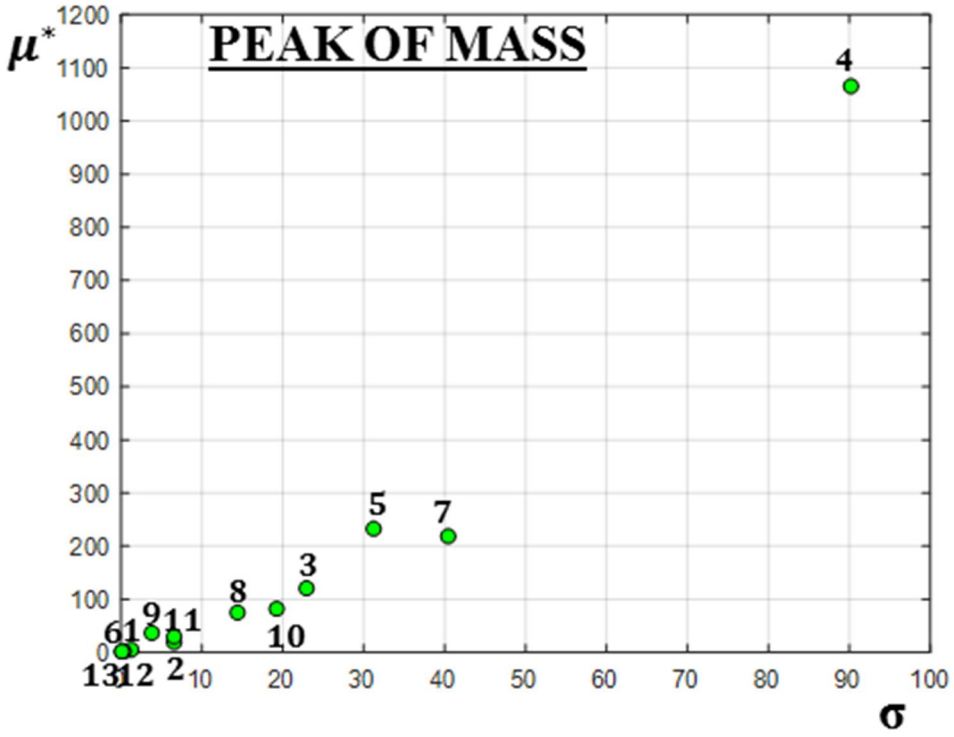


Figure 4.10. Morris Analysis: mean and absolute mean of F_i and G_i for "Mass at end".

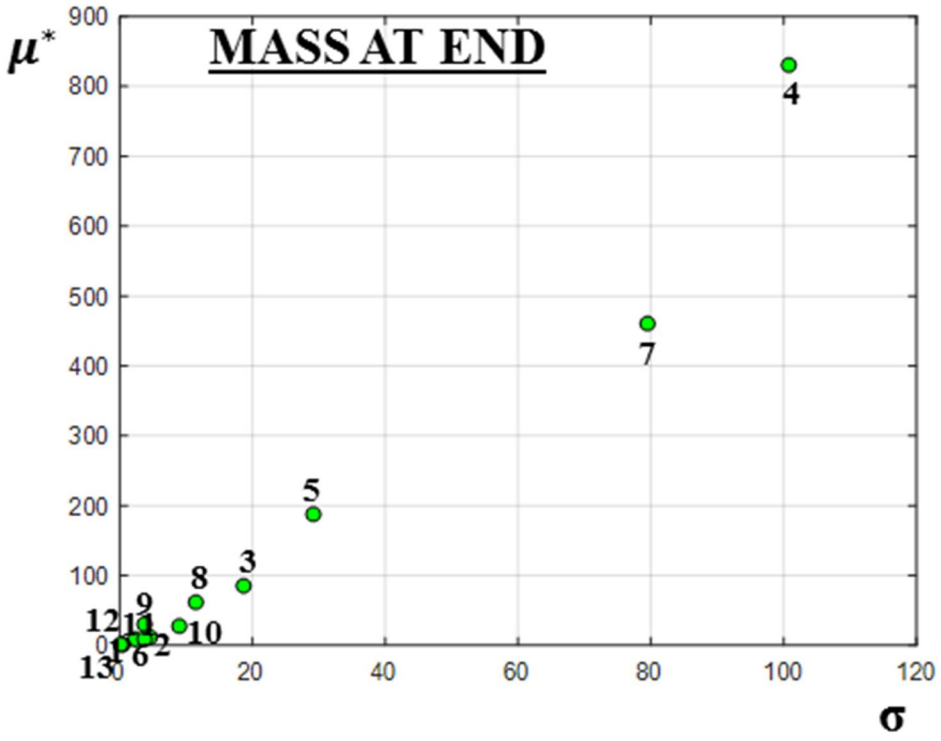
Plots state that with the exception of the porosity of the matrix reservoir and of the cap rock, global output functions of the global quantities considered are monotonic respect to uncertain parameters.

About variances of finite distributions of elementary effects, a big variance of distribution F_i will outline a notable interaction of the *i* parameter with other parameters and/or nonlinear order effects. Obviously higher means of finite distributions will generally have higher variances; the trend higher mean-higher variance is evident in the following plots.



$1 = k_{SA_m}; 2 = k_F; 3 = k_{SA_f}; 4 = k_T; 5 = P_{INJ}; 6 = k_{CR}; 7 = k_{OV};$
$8 = \Phi_{SA_m}; 9 = \Phi_{CR}; 10 = \Phi_T; 11 = \Phi_{OV}; 12 = \Phi_{SA_f}; 13 = \Phi_F$

Figure 4.11. Morris Analysis: absolute mean and variance of F_i and G_i for "Peak of mass".



$1 = k_{SA_m}; 2 = k_F; 3 = k_{SA_f}; 4 = k_T; 5 = P_{INJ}; 6 = k_{CR}; 7 = k_{OV};$
$8 = \Phi_{SA_m}; 9 = \Phi_{CR}; 10 = \Phi_T; 11 = \Phi_{OV}; 12 = \Phi_{SA_f}; 13 = \Phi_F$

Figure 4.12. Morris Analysis: absolute mean and variance of F_i and G_i for "Mass at end".

As last most dominant parameters are shown in terms of absolute means of finite distributions of elementary effects.

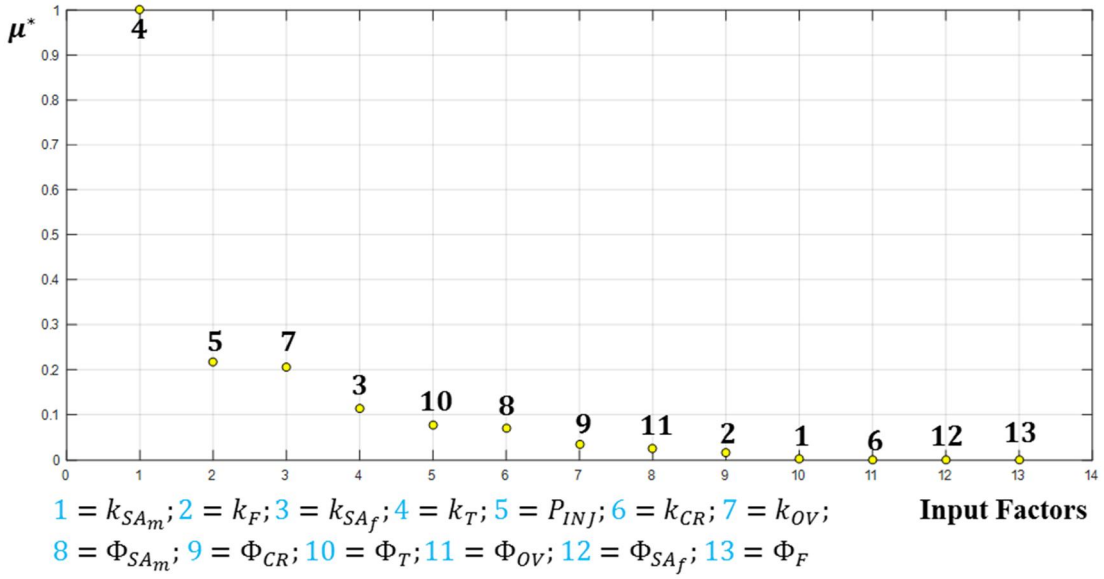


Figure 4.13. Morris Analysis: most dominant factors in terms of absolute mean for the "Peak of mass".

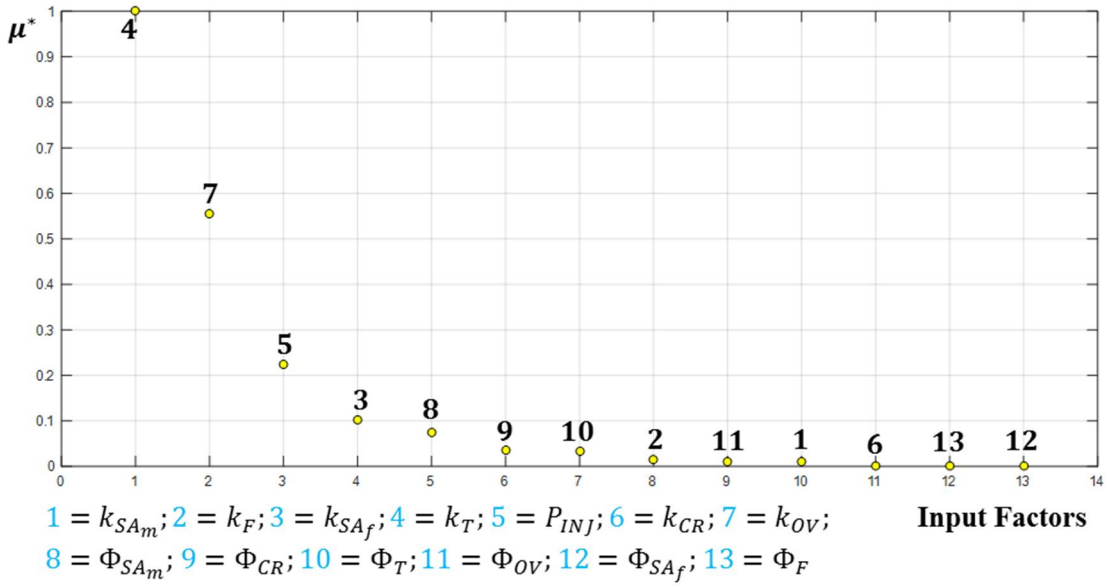


Figure 4.14. Morris Analysis: most dominant factors in terms of absolute mean for "Mass at end".

Table 4.20. Morris Analysis: most dominant factors in terms of absolute mean.

Peak of Accumulated Mass			**Accumulated Mass at the End of the Experiment**		
1°	k_T	1	1°	k_T	1
2°	P_{INJ}	0.2171	2°	k_{OV}	0.5560
3°	k_{OV}	0.2053	3°	P_{INJ}	0.2247
4°	k_{SA_f}	0.1142	4°	k_{SA_f}	0.1018
5°	Φ_T	0.0763	5°	Φ_{SA_m}	0.0740
6°	Φ_{SA_m}	0.0694	6°	Φ_{CR}	0.0364
7°	Φ_{CR}	0.0355	7°	Φ_T	0.0333
8°	Φ_{OV}	0.0275	8°	k_F	0.0157
9°	k_F	0.0172	9°	Φ_{OV}	0.0105
10°	k_{SA_m}	0.0038	10°	k_{SA_m}	0.0104
11°	k_{CR}	0.0006	11°	k_{CR}	0.0010
12°	Φ_{SA_f}	0.0003	12°	Φ_{SA_f}	0.0009
13°	Φ_F	0.0002	13°	Φ_F	0.0003

There is a semblance of agreement between Morris and previous Sobol analysis in the evaluation of k_T as the most influent parameter affecting output responses. Moreover, k_{OV} (that was not taken in consideration in the previous cases) plays an important role as well as P_{INJ}.But a clearer overview is achieved through the implementation of a Sobol analysis containing all and only thirteen input factors of Morris analysis. Before that it is necessary to verify how much results change if discretization number z is modified, for example it becomes $z = 8$.

Table 4.21. Morris Analysis: Comparison between different discretization choices ("Peak of mass" results).

PEAK OF ACCUMULATED MASS					
$z = 4$			**$z = 8$**		
1°	k_T	1	1°	k_T	1
2°	P_{INJ}	0.2171	2°	P_{INJ}	0.2396
3°	k_{OV}	0.2053	3°	k_{OV}	0.1699
4°	k_{SA_f}	0.1142	4°	Φ_T	0.0832
5°	Φ_T	0.0763	5°	k_{SA_f}	0.0820
6°	Φ_{SA_m}	0.0694	6°	Φ_{SA_m}	0.0592
7°	Φ_{CR}	0.0355	7°	Φ_{CR}	0.0406
8°	Φ_{OV}	0.0275	8°	Φ_{OV}	0.0280
9°	k_F	0.0172	9°	k_F	0.0149
10°	k_{SA_m}	0.0038	10°	k_{SA_m}	0.0033
11°	k_{CR}	0.0006	11°	k_{CR}	0.0006
12°	Φ_{SA_f}	0.0003	12°	Φ_F	0.0004
13°	Φ_F	0.0002	13°	Φ_{SA_f}	0.0002

Table 4.22. Morris Analysis: Comparison between different discretization choices ("Mass at the end" results).

ACCUMULATED MASS AT THE END OF THE EXPERIMENT					
$z = 4$			**$z = 8$**		
1°	k_T	1	1°	k_T	1
2°	k_{OV}	0.5560	2°	k_{OV}	0.5795
3°	P_{INJ}	0.2247	3°	P_{INJ}	0.2666
4°	k_{SA_f}	0.1018	4°	k_{SA_f}	0.0967
5°	Φ_{SA_m}	0.0740	5°	Φ_{SA_m}	0.0725

6°	Φ_{CR}	0.0364	6°	Φ_{CR}	0.0520
7°	Φ_{T}	0.0333	7°	Φ_{T}	0.0387
8°	k_F	0.0157	8°	k_F	0.0140
9°	Φ_{OV}	0.0105	9°	k_{SA_m}	0.0110
10°	k_{SA_m}	0.0104	10°	Φ_{OV}	0.0078
11°	k_{CR}	0.0010	11°	k_{CR}	0.0014
12°	Φ_{SA_f}	0.0009	12°	Φ_F	0.0010
13°	Φ_F	0.0003	13°	Φ_{SA_f}	0.0003

From tables 4.21. and 4.22 is under eyes how values noticeably change if discretization number *z* change. There is not a right and a wrong discretization choice but it is preferable a domain with a finer discretization so results will be referred to *z* = 4 choice.

Final Sobol Case:

The starting idea of the procedure applied to this work was to execute several analyses with a limited number ($6 \div 7$) of uncertain parameters (every parameter was taken as uncertain at least one time) in order to find those parameters which can be considered surely uninfluential for global quantities considered (parameters whose Sobol indices assumed 0 value) and at the same maintain acceptable computational costs. But it was discovered that for the sake of model reliability many parameters cannot be considered uncertain. So the smartest decision is to focus the study on certain categories of parameters (permeabilities, porosities and injection pressure) and execute a unique Sobol analysis although its relative high computational cost. The construction of the surrogate model required following computational costs:

- *w* = 2 ⟶ Number of full model runs = 365
- *w* = 3 ⟶ Number of full model runs = 3329
- *w* = 4 ⟶ Number of full model runs = 23245

Analyses of orders *w* = 2 and *w* = 3 were affordable, while that of order *w* = 4 was an impossible task for the computing power at disposal. But the *Sparsity-of-effects*

principle indicates that most systems or processes are usually sensitive to only a subset of factors and their low-order (e.g. second-order) interactions, and that high-order interactions are typically insignificant or negligible (Razavi & Gupta 2015). So if results obtained with $w = 2$ and $w = 3$ degrees are consistent ones with the others and if scatter plots don't find any inconsistency, analysis with $w = 4$ is no more necessary.

Final results are shown here below.

Table 4.23. Final case simulation results for the peak of mass response.

$M_{H_2,MAX}$		
$degree$	$w = 2$	$w = 3$
Θ	644	639
$\sqrt{D}$	3.5×10^2	3.5×10^2
D	1.2×10^5	1.2×10^5

Table 4.24. Final case simulation results for the mass at end response.

$M_{H_2,END}$		
$degree$	$w = 2$	$w = 3$
Θ	517	509
$\sqrt{D}$	2.9×10^2	2.9×10^2
D	8.6×10^4	8.6×10^4

Table 4.25. Principal Sobol Indices of final case.

Principal Indices	$M_{H_2,MAX}$		$M_{H_2,END}$	
	$w = 2$	$w = 3$	$w = 2$	$w = 3$
$S_{k_{SA_m}}$	0.000	0.000	0.000	0.000
$S_{k_{SA_f}}$	0.003	0.008	0.004	0.001
S_{k_F}	0.000	0.000	0.000	0.001
S_{k_T}	0.914	0.896	0.777	0.789

$S_{k_{CR}}$	0.000	0.002	0.000	0.003
$S_{k_{OV}}$	0.014	0.022	0.123	0.103
$S_{\Phi_{SA_m}}$	0.002	0.000	0.002	0.006
$S_{\Phi_{CR}}$	0.001	0.004	0.002	0.000
S_{Φ_T}	0.003	0.004	0.000	0.000
$S_{\Phi_{SA_f}}$	0.000	0.000	0.000	0.000
S_{Φ_F}	0.000	0.000	0.000	0.000
$S_{\Phi_{OV}}$	0.001	0.000	0.000	0.000
$S_{P_{INJ}}$	0.043	0.044	0.041	0.035

Table 4.26. Total Sobol Indices of final case.

Total Indices	$M_{H_2,MAX}$		$M_{H_2,END}$	
	$w = 2$	$w = 3$	$w = 2$	$w = 3$
$S^T_{k_{SA_m}}$	0.000	0.000	0.000	0.000
$S^T_{k_{SA_f}}$	0.003	0.008	0.005	0.002
$S^T_{k_F}$	0.000	0.000	0.000	0.001
$S^T_{k_T}$	0.933	0.915	0.827	0.848
$S^T_{k_{CR}}$	0.000	0.002	0.000	0.003
$S^T_{k_{OV}}$	0.021	0.029	0.164	0.151
$S^T_{\Phi_{SA_m}}$	0.003	0.001	0.002	0.007
$S^T_{\Phi_{CR}}$	0.001	0.005	0.002	0.000
$S^T_{\Phi_T}$	0.003	0.005	0.000	0.000
$S_{\Phi_{SA_f}}$	0.000	0.000	0.000	0.000
$S^T_{\Phi_F}$	0.000	0.000	0.000	0.000
$S^T_{\Phi_{OV}}$	0.001	0.001	0.000	0.000
$S^T_{P_{INJ}}$	0.054	0.053	0.052	0.048

From tables above a new important parameter comes to the light: the overburden permeability controls the variance of mass global quantities and its influence on accumulated mass at the final time is the highest with the exception of the target permeability. So between these 13 parameters 3 seem to be relevant in the ambit of this work: Target permeability, Overburden permeability and Injection pressure. Other possible considerations about linearity and parameters interactions are the same of those dealt with other three cases. Before going into further discussion there is the duty to check scatter plots of this new case, because the increment of parameters could deteriorate the quality of the surrogate model. Scatter plots illustrated here below confirm the good quality of surrogate models.

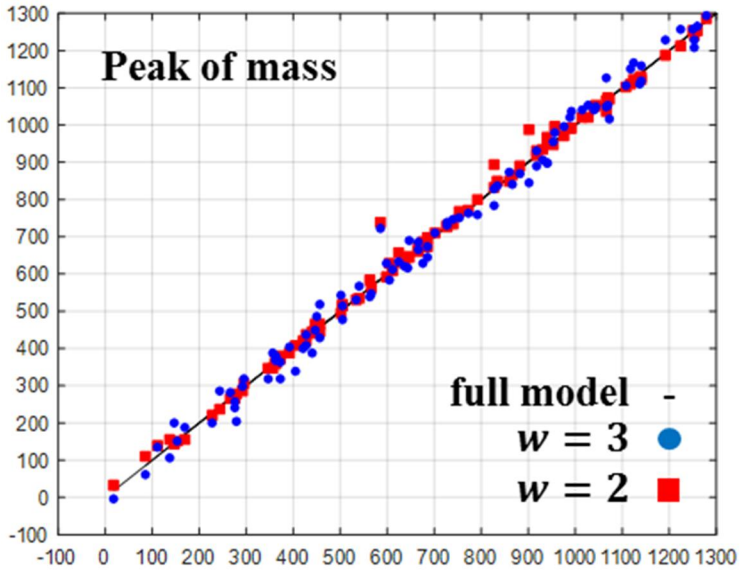


Figure 4.15. Final case: scatter plot of "Peak of mass" response.

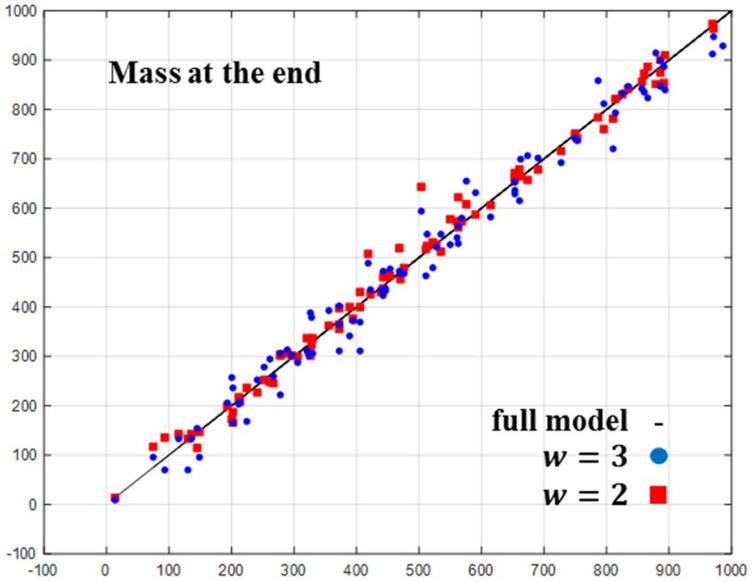


Figure 4.16. Final case: scatter plot of "Mass at end" response.

Results Comparison and Discussion:

It has been already discussed about different objects of the Morris and Sobol analyses (one is concerned with local sensitivities and the other has a more global view) but they both try to answer to the question: which parameters control the intrusion of the mass from the gas reservoir into the target aquifer? Both Sobol final case and Morris analyses agree in the identification of three parameters as the most influent for global quantities considered: Target permeability, Overburden permeability, and Injection pressure.

From the table below it is evident how the importance of the overburden permeability is higher for the global quantity "Mass at the end" respect to the "Peak of Mass". This can be explained by the fact that once the injection is interrupted, the dispersion of the gas from the target aquifer to the surrounding environment is strictly related to the permeability of the environment; while during injection the rate of accumulation into the aquifer is less influenced by the surrounding formations.

Table 4.27. Comparison between Total Sobol indices and Morris results.

	Peak of mass			**Mass at the end**		
	Morris	$w = 2$	$w = 3$	Morris	$w = 2$	$w = 3$
k_T	1	0.9331	0.9146	1	0.8267	0.8475
k_{OV}	0.1699	0.0212	0.0295	0.5795	0.1639	0.1514
P_{INJ}	0.2396	0.0542	0.0531	0.2666	0.0517	0.0478

Another remark is how the importance of overburden permeability varies by changing target quantity of interest: Mass at End global quantity is more influenced by k_{OV} than Peak of Mass global quantity. A possible explanation can be that the gas dispersion from the target aquifer is mainly controlled by the permeability formations surrounding it and the quantity of gas present in the target aquifer at the end of the experiment depends strongly from how fast the gas disperses from it.

4.3 PROBABILITY DISTRIBUTIONS OF TARGET VARIABLES

The aim of this section is to enrich the analysis about two global quantities considered by showing different PDFs associated to different cases and *w* degrees. The Probability Distribution Function is a remarkable tool that describes surface responses or in other words the distribution in the output space of all simulations belonging to a defined case. Moreover, it permits to match visually the responses obtained by different surrogate models having different g-PCE degrees or/and different sets of uncertain parameters. In the contest of this analysis the addition of a case having as uncertain parameters only the 3 most influent parameters of the "Final Sobol case" will enrich the discussion. So before unclosing PDFs plots the new case called "3 Parameters case" will be briefly introduced.

3 Parameters case:

This case contains only the three most significant parameters of the "Final Sobol case" which are target permeability, overburden permeability and injection pressure. There is no sense in computing Sobol analysis for this case because a more complete analysis has been executed in the "Final Sobol case". So only results regarding means and variances of global target quantities are reported.

Table 4.28. 3 Parameters case simulation results for the peak of mass response.

$M_{H_2,MAX}$			
$degree$	$w = 2$	$w = 3$	$w = 4$
Θ	577	577	576
$\sqrt{D}$	3×10^2	3×10^2	3×10^2
D	9.3×10^4	9.2×10^4	9.3×10^4

Table 4.29. 3 Parameters case simulation results for the mass at end response.

$M_{H_2,END}$			
$degree$	$w = 2$	$w = 3$	$w = 4$
Θ	483	483	483
$\sqrt{D}$	$2{,}7 \times 10^2$	$2{,}7 \times 10^2$	$2{,}7 \times 10^2$
D	7.4×10^4	7.5×10^4	7.5×10^4

Here computational costs are reduced thanks to the small amount of uncertain parameters:

- $w = 2$ ⟶ Number of full model runs = 25
- $w = 3$ ⟶ Number of full model runs = 69
- $w = 4$ ⟶ Number of full model runs = 165

Looking at the mean and variation values of target quantities it is evident a reduction respect to the previous "Final Sobol case". Before entering in further discussions scatter plots are made (and as shown here below expected they give good responses) and PDFs are presented.

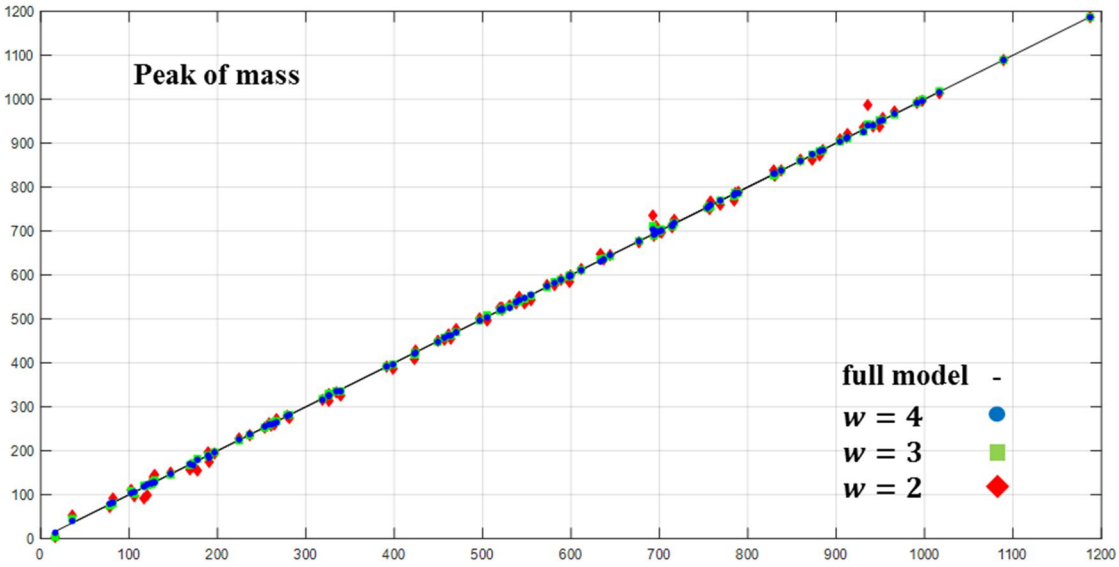


Figure 4.17. 3 Parameter case: scatter plot of "Peak of mass" response.

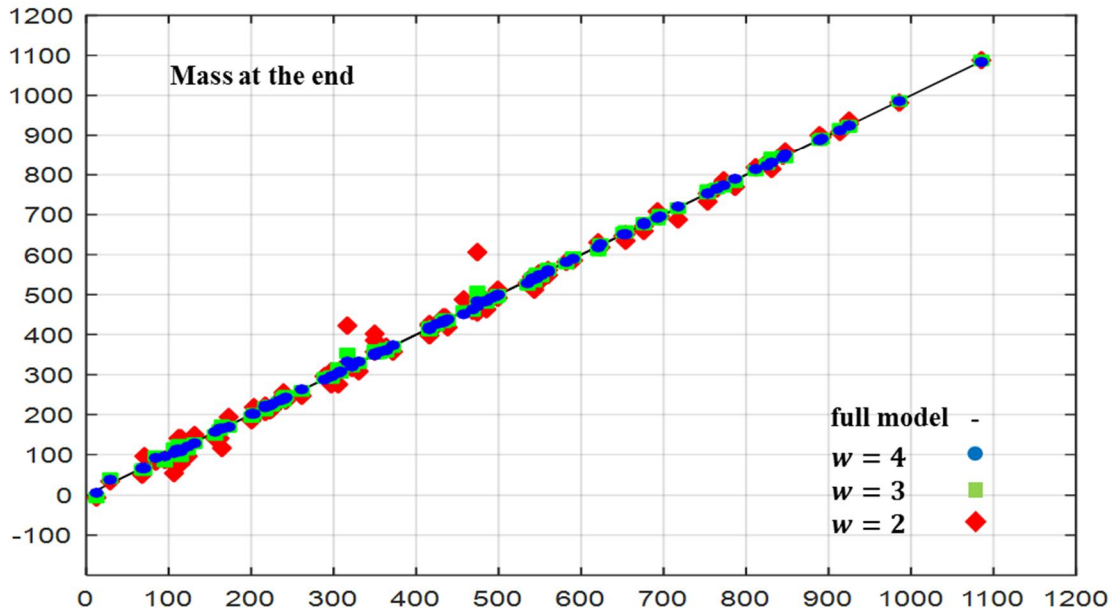


Figure 4.18. 3 Parameter case: scatter plot of "Mass at the end" response.

PDFs construction:

The construction of PDFs for *global quantities* analysis relies on g-PCE mainly for computational reason. If the surrogate model obtained through the g-PCE suites optimally the full model, it is possible to build PDF by running hundreds of thousands of times the surrogate model with very small computational costs if compared to execution of the same PDFs by the full model. Scatter Plots are used to understand the quality of the surrogate model. More simulations are performed and more PDFs will be accurate; obviously computational cost related to a PDF construction grows proportionally to the number of simulations. In this study 100.000 simulations for each case were identified as the best compromise between PDF's accuracy and computational cost.

PDFs comparison:

Now PDFs of peak of accumulated mass and accumulated mass at the end of experiment are presented. Each PDFs is composed by 100000 random simulations generated by surrogate models obtained by means of g-PCE. Dashed lines represent the Final Sobol case and continuous lines 3 Parameters case. Moreover, different colours (red, green and blue) stand for different PCE degree ($w = 2, w = 3, w = 4$).

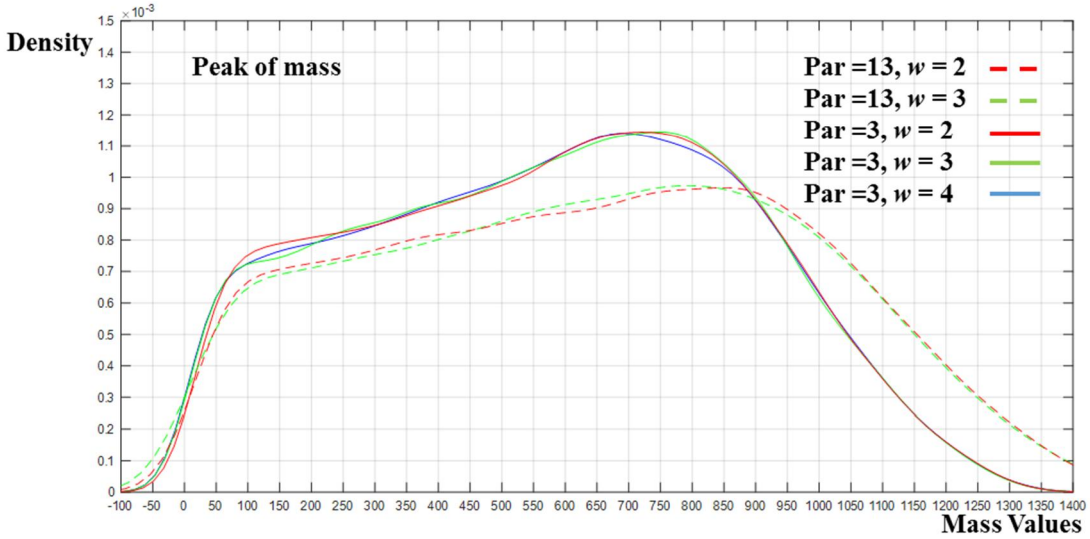


Figure 4.19. PDFs of Peak of mass global quantity.

Looking at Final Sobol case lines, the reader can see how they remain lower than 3 Parameters case until the peak density but after it they decrease slower and so this behaviour determines a slightly higher variance for the Final Sobol case and at the same time a lower peak density. Another consideration is about the precision level of PCE degrees: There is much similarity between output functions obtained with $w = 2$ and $w = 3$ (and $w = 4$) degrees in the ambit of the same case.

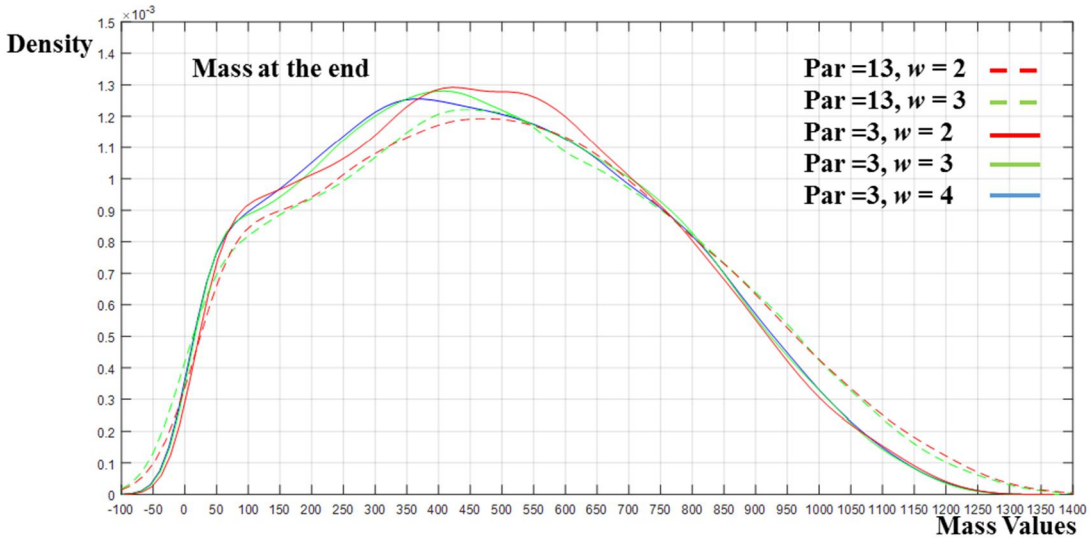


Figure 4.20. PDFs of Mass at the end global quantity.

About accumulated "Mass at the end" of the experiment, density peaks are obviously placed at lower mass values but they have higher number than "Peak of mass" density peaks.

Unlike the other global quantity PDFs of Final Sobol case are more similar to PDFs of 3 Parameters case and unlike the previous group of PDFs, lines of the same case don't agree perfectly, especially in the 3 Parameters case.
Both *global quantities* have high variances: they can assume a wide range of values, from nearly 0 to over 1000. This indicates that even if the process is controlled by a few parameters and even if their values are well known, it remains difficult to predict how much mass will enter in the aquifer.

5 CONCLUSIONS

The purpose of the thesis is the analysis of the gas intrusion from a reservoir into an aquifer, in particular the analysis is concerned about rules of some key parameters in flux and transport processes. Analysis is carried on by means of *global sensitivity analysis* (GSA), which is based on the determination of Sobol indices (for their calculation a *Polynomial Chaos Expansion* has been used), and by Morris analysis.
The study of the problem has been executed through *global quantities* which characterize the gas intrusion, because the characterization of solutes in every point of the domain was an impossible task.
Processes taken in consideration are flux processes and transport processes.
Global quantities considered are the peak of accumulated mass inside the aquifer throughout the time experiment and the accumulated mass inside the aquifer at the final moment of the experiment.
The development of the GSA analysis starts with the identification of those parameters which govern phenomena under study and the task is achieved by several attempts and by using both Sobol and Morris indices. In order to compute Sobol indices PCE is built by means of *sparse grids*, a method that allows to reduce the number full model simulations performed by a numerical model in comparison to the standard Monte Carlo standard method.
Sobol indices are a good way to describe the variance of *global quantities*, while Morris indices are excellent indicators of the mean of *global quantities*.
Results are reported in the fourth chapter. From the combination of Sobol and Morris analyses is clear how the dominant parameters in the process of aquifer contamination

are the target aquifer, the injection pressure and the overburden permeability. About the peak of accumulated mass target permeability is the most influent and other two factors have more or less the same importance, while for what concerns the accumulated mass at the final time of the experiment overburden permeability is more dominant than the injection pressure but less than target permeability.
Then 100’000 simulations of PCE model are performed for each case considered in order to build PDFs of responses. The analysis of responses surfaces gives back many considerations.
Between these two most important are: The peak of mass the surrogate model of degree equal to 2 is already a good approximation of the full model and the other is that especially for accumulated mass at the end “3 Parameters case” is very similar to “Final Sobol case” which has 13 uncertain parameters.
For future developments and studies 2 principal suggestions are presented.
First suggestion is the adoption of a unique sensitivity tool for the sensitivity analysis which can cover several aspects of the response surfaces. Razavi and Gupta have proposed a method called VARS (Variogram Analysis of Response Surface). VARS approach obtains robust results describing locally and globally response surfaces and it is not influenced by noise/roughness in the underlying response surface.
Second suggestion is to implement a dimensional analysis. This could allow a generalization of the problem and a collapse of many response curves into a few dimensionless curves. Moreover, the dimensional analysis could characterize the problem with some dimensionless groups which could simplify analyses.
Anyway future possibilities are infinite but at the same time this study could always be a good starting point or a good comparison.

REFERENCES

Archer G.E.B., Saltelli A. and Sobol I.M. Sensitivity measures, anova-like Techniques and the use of bootstrap, 2007.

Barthelmann V., E. Novak and K. Ritter.High dimensional polynomial interpolation on sparse grids. Adv. Comput. Math., 12: 273_288, 2000.

Bear J, Cheng AHD, Modeling groundwater flow and contaminant transport, Dordrecht Heidelberg London New York: Springer 2010.

Blunt M.. Reservoir engineering lecture notes.

Bp Energy Outlook. Outlook to 2035 – energy use to rise by a third;2016.

Bp Statistical Review of World Energy June 2016. Natural gas – 2015 in review; 2016.

Campolongo F., Carboni J., Saltelli A. An effective screening design for sensitivity analysis of large models, 2006.

Dell' Oca A. Studio dei processi di flusso e trasporto a densità variabile nel mezzo poroso: intrusione salina in acquiferi costieri; (2013).

Formaggia L., Guadagnini A., Imperiali I., Lever V., Porta G., Riva M., Scotti A., Tamellini L. Global sensitivity analysis through polynomial chaos expansion of a basin-scale geochemical compaction model; 2013.

Ioos B. and Leimatre P. A review on global sensitivity analysis methods, 2014

Morris M.D. Factorial Sampling Plans for Preliminary Computational Experiments; 1991.

Organization of the Petroleum Exporting Countries, Public Relations and Information Department. Annual Report 2015; 2015.

Porta G., Tamellini L., Lever V. and Riva M. Inverse modeling of geochemical and mechanical compaction in sedimentary basins through Polynomial Chaos Expansion;2014.

Riva M., Guadagnini A., Dell'Oca. Probabilistic assessment of seawater intrusion under multiple sources of uncertainty; 2014

Saltelli A., Ratto M., Tarantola S., Campolongo F., 2006, "Sensitivity analysis practices: Strategies for model-based inference", *Reliability Engineering and System Safety,* 91, 1109–1125.

Saman Razavi, Hoshin V.Gupta. A New Framework for Comprehensive, Robust, and Efficient Global Sensitivity Analysis: Part I – Theory;2015.

Saman Razavi, Hoshin V.Gupta. A New Framework for Comprehensive, Robust, and Efficient Global Sensitivity Analysis: Part II – Application;2015.

Saman Razavi, Hoshin V.Gupta. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models; 2015.

Sobol I. , Sensitivity estimates for nonlinear mathematical models. Math Modeling Comput Exp 1993;1:407-14.

Sobol I., 2001, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates", *Mathematics and Computers in Simulation,* 55, 217-280.

Sudret B., 2008, "Global sensitivity analysis using polynomial chaos expansions", *Reliability Engineering & System Safety*, 93, 964–979.