POLITECNICO DI MILANO

Scuola di Ingegneria Civile, Ambientale e Territoriale

Facoltà di Ingegneria Civile – Dipartimento di Idraulica

Role of Multiple Sources of Uncertainty on Predicting Gas Migration Through a Caprock

Tesi di Laurea Magistrale di:

Pietro Pisnoli Matr. 858234

Anno accademico 2016/2017

Relatore: Prof. Monica Riva

Correlatore: Anna Russian

Alla mia famiglia per il sostegno morale, spirituale ed
economico offerto in questi anni.
Alla Professoressa Monica Riva per l'opportunità offerta ed
il sostegno indispensabile nel realizzarla.
Ad Anna Russian per gli indispensabili suggerimenti
e la pazienza dimostrata nei miei confronti.

INDEX

Figure Index

Tables Index

ABSTRACT

We propose a one-dimensional stochastic model describing diffusion of gas through a layer of rock under multiple sources of uncertainty including: initial concentration of gas, boundary conditions and diffusion coefficient. Diffusion coefficient comprises not only the classical Fick's molecular diffusion but also the Knudsen diffusion, a process occurring in media characterized by very low permeability, such as caprocks overlying gas/oil reservoirs. All stochastic inputs are assumed independent of each other, constant in space and time and uniformed distributed. We compute statistical moments of gas concentration and of the instantaneous flux solution using three different approaches: we derive the partial differential equations, PDEs, satisfied by mean and variance of the qualities of interest and solve them (1) analytically and (2) numerically; (3) we perform also a set of Monte Carlo simulations.

We investigate the impact of uncertainty settings by consider their single and joint effect by employing a global sensitivity analysis (GSA). The GSA was made with two different techniques: the Sobol's Indices and the multiple moment based metrics (AMA indices).

ABSTRACT (ITA)

Nel seguente elaborato di tesi si propone un modello stocastico monodimensionale che descrive la diffusione di un gas all'interno di uno strato di roccia, considerando molteplici possibili fonti d'incertezza quali: la concentrazione iniziale, una condizione al contorno e il coefficiente di diffusione. Quest'ultimo comprende non solo la classica diffusione molecolare di Fick, ma anche la diffusione di Knudsen, un processo che è stato osservato all'interno di mezzi porosi a bassissime permeabilità, come le caprock che formano i bacini petroliferi. Tutti gli input stocastici sono assunti indipendenti l'uno dall'altro, costanti nello spazio e nel tempo e con una distribuzione di probabilità uniforme.

Sono state derivate le espressioni analitiche delle soluzioni di concentrazione e flusso istantaneo risolvendo la seconda equazione differenziale di Fick. Quindi sono stati calcolati media e varianza delle due soluzioni utilizzando tre differenti approcci: calcolando analiticamente le espressioni dei momenti (1), calcolando numericamente con il metodo di Eulero alle differenze finite le soluzioni (2) ed infine eseguendo un set di simulazioni Montecarlo utilizzando la soluzione analitica (3).

Si è investigato il comportamento e l'impatto delle incertezze delle variabili di input sulle due soluzioni analizzate facendole variare prima una sola variabile alla volta ed infine considerandole tutte e tre stocastiche. Per valutare quanto l'incertezza delle variabili di input influisca sulle soluzioni analizzate è stata condotta un'analisi di sensitività globale (global sensitivity analysis, GSA). A tal fine, sono stati usati gli indici di Sobol ed gli indici di AMA. Questi ultimi, non si limitano a considerare soltanto la varianza della dell'output, ma ne considerano tutti i 4 momenti statistici.

1 INTRODUCTION

Gas transport mechanisms through very low permeability sedimentary rocks have recently acquired and increased importance. The main reasons are the rapid increase in gas production from unconventional reservoirs, the possibility of storing methane gas in exhausted reservoirs and the green options of storing carbon dioxide.

There are many problems in finding a physical and mathematical model that takes into account all the variables that influence the advancement of gas within the rock: the knowledge of initial and boundary conditions, the heterogeneity of the rocks and its physical characteristics, the stress states history, the rock lithology and the fracture net.

Gas migration through low permeability sedimentary rock can be divided into two principal mechanisms: a continuous-phase gas migration and a discontinuous-phase gas migration (Leonid F. Khilyuk, 2000) .

The first is relative a flow when there is a higher water pressure and/or gas saturation respect the normal level therefore it is linked to pressure gradient and concentration gradient. It considers the advective flux and the diffusion process. The diffusion of gas is the main process and it takes place in very low permeability rock through the small diameter of pores. This process is very slow. The advective part is taken into account as a flow in a pipeline. Pressure is the driving force for the gas movement but also the geometry of the path and of the petrophysical features (fracture width, length, water saturation, tortuosity, absolute and relative roughness of the fracture walls, etc.) affect the gas flow.

The second mechanism occurs when gas in form of bubbles migrate through water filled shale rock. It is driven by the capillary forces and by the different stress tension between the phases. Discontinuous buoyancy-type mechanism consists of gas bubbles migrating through the channels of water-filled interconnected pores. The driven force is given by the differential pressures

resulting from differences in specific weight between the gas bubbles and water the pressure drop along path. (Leonid F. Khilyuk, 2000).

In this work, we examine the problem of the diffusion of a gas within a sedimentary rock layer characterized by nanometric porosity.

In order to take into account and quantify in a systematic way the uncertainties that characterize our problem, we chose a stochastic approach. The combination of complexity, uncertainty and ignorance that are present in real problems, not only due to natural phenomena, but also due to human behavior, requires the consideration of randomness in the mathematical models. Random differential equations have been used in the last few decades to deal with errors and uncertainty.

Stochastic processes are ways of quantifying the dynamic relationships of sequences of random events. Stochastic models play an important role in elucidating many areas of the natural and engineering sciences. They can be used to analyze the variability inherent in biological and medical processes, to deal with uncertainties affecting managerial decisions and with the complexities of psychological and social interactions, and to provide new perspectives, methodology, models, and intuition to aid in other mathematical and statistical studies. (Howard & Karlin, 1998) One of the most common methods for solving stochastic problems is the Montecarlo method.

The Monte Carlo method is a robust way to study the propagation of uncertainties. Monte Carlo simulation performs risk analysis by building models of possible results by substituting a range of values—a probability distribution—for any factor that has inherent uncertainty. It then calculates results over and over, each time using a different set of random values from the probability functions. Depending upon the number of uncertainties and the ranges specified for them, a Monte Carlo simulation could involve thousands or tens of thousands of recalculations before it is complete. Monte Carlo simulation produces distributions of possible outcome values. Depending upon the number of uncertainties and the ranges specified for them, a Monte Carlo analysis could involve thousands or millions of realizations before it is

statistically relevant. In practice if we have a model $C = f(x, Y, Z, w)$ where Y and Z are random variables characterized by a certain pdf, we compute the value of C for n different combinations of Y and Z, taken randomly according to their pdf. At the end, we compute statistical analysis on C e.g. mean and variance of C.

We consider the gas diffusion problem in homogenous medium. This problem is designed to simulate conditions that could be present in a gas basin developed beneath a rock shale layer, which acts as caprock, without fractures. Indeed, these last, may act as a preferential way and thus nullify the diffusion effect.

The work is structured in the following way. The last part of the Introduction is devoted to a brief explanation of the structure of the reservoir system, to an analysis of the various types of mechanisms that exist in the transport of gases. In the first chapter there is a theoretical introduction on the global sensitivity analysis and on the diffusion problem.

Then we expose the analytical solution of second Fick's law, computed using the separation of variables technique. (Appendix A). After that we considered three of the possible variables of the problem as stochastic variable, considering the medium as homogeneous. We took the initial condition, that is the initial concentration in the cap-rock, the boundary condition, that is the reservoir gas concentration considered constant in time, and the diffusion coefficient in the caprock. At the beginning we consider one parameter random at a time keeping deterministic the others and after we studied the problem with all three variables random.

We obtained the solutions (concentration and instantaneous flux) for the homogeneous case with three different method: with a large number of Monte Carlo realizations, using the analytical solution of the second Fick's law and finally numerically, using the Euler finite difference method.

When We analyzed the problem with all three inputs random, we also made a global sensitivity analysis to better understand the influence of each parameter in the solution.
After that, we analyzed the heterogenous problem by taking into account a correlated random field that we obtained by using the Karhunen-Loève expansion.

Gas reservoir, in geology and natural gas production, is a naturally storage area, characteristically a folded rock formation such an anticline, that traps and holds natural gas. The reservoir rock must be permeable and porous to contain the gas, and it has to be capped by impervious rock in order to form an effective seal that prevents the gas from escaping upward or laterally, the caprock (Encyclopædia Britannica, 2012). The natural gas migrates into reservoir from the less-permeable source rock because of the pressure difference between the source rocks, which are compressed by the weight of overlying rocks, and the reservoir rocks, which are at lower pressure. Even oil basins are often characterized by the presence of a higher layer, in contact with the caprock, where lighter hydrocarbons, mainly methane and ethane, are concentrated in gaseous state. This is called gas cap and it has the same characteristics of a gas reservoir.
Natural gas is a combustible, gaseous mixture of simple hydrocarbon compounds, found in deep underground reservoirs formed by porous rock. It is composed almost entirely of methane, but it may contain small amounts of other higher alkanes (ethane, propane butane and pentane) and sometimes a small percentage of carbon dioxide, nitrogen, hydrogen sulphide or helium.

Methane	CH_4	70-90 %
Ethane	C_2H_6	0-20 %
Propane	C_3H_8	
Butane	C_4H_{10}	
Carbon Dioxide	CO_2	0-8 %
Oxygen	O_2	0-0.2 %
Nitrogen	N_2	0-5 %
Hydrogen Sulphide	H_2S	0-5 %
Rare Gases	A, He, Ne	trace

Tab 1 Typical composition of natural gas [naturalgas.org].

Natural gas is odorless and colorless; the slightly sour smell that we associate with the gas coming from a stovetop is due to an odorization process (for safety and leak detection) which adds mercaptan compounds to the end-use gas. More precisely, odorization compounds are mixtures of t-butyl mercaptan, isopropyl mercaptan, tetrahydrothiophene, dimethyl sulfide and other sulfur compounds.

As a fossil fuel, natural gas has formed from the decaying remains of pre-historic plant and animal life. As with petroleum, most natural gas formation is due to the breakdown of prehistoric marine zooplankton.

Typically, it is found at the top of petroleum reservoirs, where it has been formed by the combined action of methanogenic bacteria (they produce methane while they decompose organic material) and through catagenesis (the thermal decomposition of kerogen). As marine sediments have buried deep within the earth, high temperatures and pressures lead to varying degrees of the completion of catagenesis, which is the process that produces both petroleum and natural gas. Higher temperatures and pressures favor the formation of lighter hydrocarbons (natural gas), and so oil/gas formations that

are deeper in the earth tend to have a greater ratio of gas to petroleum. (Fuel Chemistry Division Public Education & Outreach Committee, n.d.).

Methane, as already mentioned, is the major component of natural gas. It is the simplest alkane with chemical formula CH_4 and it has a molar mass 16.04 [g/mol].

Methane solubility in aqueous solutions depends on the balance between the chemical potential of CH_4 in the liquid phase, μlCH_4, and in the vapor phase, μvCH_4. Based on a simple fugacity coefficient equation Zhenhao Duan (Zhenhao Duan *, 2006) have presented CH_4 solubility in pure water and aqueous NaCl solutions, gas phase compositions and liquid phase density with the best experimental accuracy from 273 to 523 K and from 1 to 2000 bar. In the typical reservoir range (373-450 K and 200-600 bar) the methane solubility can be 0.4 mol/kg.

Based on reservoir conditions and depending on its solubility, methane can be found as:

- Free-phase form under caprock and in fracture network
- Dissolved in water
- Adsorbed form in micro fracture network or shale matrix

The process leading to the formation of the methane, in addition to the thermogenic ones, common to other hydrocarbons, is the biogenic. It is done by methanogens bacteria that decompose organic matter under anoxic conditions, referred to as biogenic methane. These microorganisms are active in the intestine of most animals and are responsible for methane release from decomposing landfill waste. In the process of petroleum formation, methane may be formed in this manner during the early stages burial.

Geochemical indicators, such as stable carbon and hydrogen isotopes of methane, stable carbon isotopes of ethane, and hydrocarbon ratios, have been used to evaluate methane sources however their utility is complicated by influences from multiple physical (e.g., mixing) and geochemical (e.g., redox) processes. Many authors give different models to understand from the isotopic

ratio and other ratios, the origin of the gas and if there was mixing inside the reservoir (Alain Prinzhofer, 1997). He suggests testing a mixing hypothesis with bacterial methane to use a diagram displaying ethane/methane ratios versus δ13C.

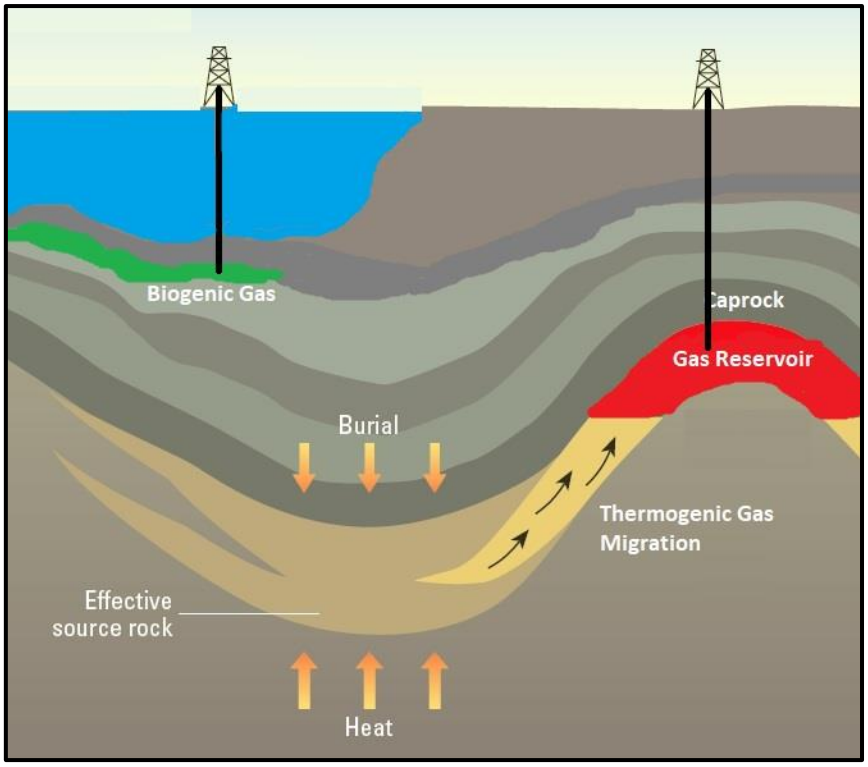


Fig 1-1 Gas Sources

The study of gas diffusion in the caprock is also very important for geologic sequestration.

Geologic sequestration begins with capturing CO2 from the exhaust of fossil-fuel power plants and other major sources. The captured CO2 is piped 1 to 4 kilometers below the land surface and injected into porous rock formations. Compared to the rates of terrestrial carbon uptake-logic sequestration is currently used to store only small amounts of carbon per year. Much larger rates of sequestration are envisioned to take advantage of the potential permanence and capacity of geologic storage. The permanence of geologic sequestration depends on the effectiveness of several CO2 trapping mechanisms. After CO2 is injected underground, it will rise buoyantly until it

is trapped beneath an impermeable barrier, or seal. In principle, this physical trapping mechanism, which is identical to the natural geologic trapping of oil and gas, can retain CO2 for thousands to millions of years. Some of the injected CO2 will eventually dissolve in ground water, and some may be trapped in the form of carbonate minerals formed by chemical reactions with the surrounding rock. All these processes are susceptible to change over time following CO2 injection. Scientists are studying the permanence of these trapping mechanisms and developing methods to determine the potential for geologically sequestered CO2 to leak back to the atmosphere. The capacity for geologic carbon sequestration is constrained by the volume and distribution of potential storage sites. Unmineable coal beds have also been proposed for potential CO2 storage, but more information is needed about the storage characteristics and the impacts of CO2 injection in these formations. Scientists are developing methods to refine estimates of the national capacity for geologic carbon sequestration. To fully assess the potential for geologic carbon sequestration, economic costs and environmental risks must be considered. Infrastructure costs will depend on the locations of suitable storage sites. Environmental risks may include seismic disturbances, deformation of the land surface, contamination of potable water supplies, and adverse effects on ecosystems and human health. Scientists are pioneering the use of new geophysical and geochemical methods that can be used to anticipate the potential costs and environmental effects of geologic carbon sequestration (Sundquist, Burruss, Faulkner, & and all.)

2 THEORETICAL BACKGROUND

In the following chapter the main theoretical framework used in the "methodology chapter" will be presented. Initially we briefly describe the main mechanisms of gas advancement inside the caprock; then the diffusion equation and mathematical framework will be analyzed. Successively we detail the process studied and the approach used in the thesis work. Finally, there will be the theoretical part concerning the global sensitivity analysis with the two methodologies used.

2.1 CONTINUUM ONE PHASE FLOW

Gas transport mechanisms in shale rocks matrix can be divided into two main processes: (1) the advective process, where Darcy Flow and Slip Flow are dominant and (2) the diffusion process, where Fick's, Knudsen and Surface Diffusion are dominant. All these processes are influenced by various physical phenomena and conditions like: adsorbing/desorbing effect, type of gas, the history of the rock, stress and temperature, lithography, Total Organic Carbon (TOC), highly anisotropic permeability as well as the fracture geometry.

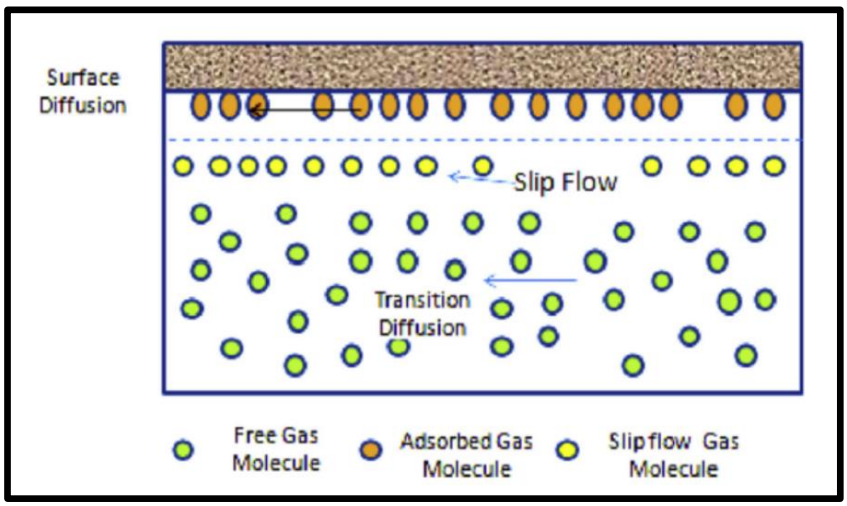


Fig 2-1 Gas transport mechanisms (Yanfeng He, 2017)

The surface diffusion plays a minority role in the gas transport through a shale caprock, and therefore we neglect it. This is due to the fact that it is a type of diffusion that is present for the gas molecules adsorbed by the solid matrix surface. This adsorbed gas part is directly related to the present proportion of kerogen in the rock. As seen above, most caprock shale has very low percent of kerogen, it is therefore justified to exclude this.

We can therefore identify three main processes involved in the transport of the gas absorbed by the water inside the solid shale matrix:

- **Slip Flow.** Slip flow (Klinkenberg phenomenon) is a non-Darcy effect that occurs when the mean free path length of gas free molecules is close to the average size of pores in a porous medium. This condition results in the acceleration of individual gas molecules along the flow path. Slip flow can also be interpreted as a viscosity reduction because of interactions between gas molecules and pore walls start to overcome molecule–molecule interactions which is the definition of internal fluid friction and the basis of viscosity. Slip flow is particularly dominant in narrow fractures as well as the matrix systems of most coals, which are typically characterized by micro- and meso-scale pore throats.

- **Fick's Diffusion.** Fick's diffusion is the classical diffusion process based on the concentration gradient and prevails on all others when the frequency of collision molecules-molecules is higher than the frequency of collision molecules-pore wall.

- **Knudsen Diffusion.** Knudsen diffusion is a diffusion process that occurs when the scale length of a system is comparable to the mean free path of the gas particles. Consider the diffusion of gas molecules through very small pores, if pore diameter is smaller than the mean free path of the gas molecules and the density of the gas is low, the gas molecules collide with pore walls more frequently than with each other's. It's then a complementary process of Fick's diffusion.

The main driving forces that are found are a concentration gradient and a possible abnormal pressure. The first is given by the accumulation of gas below the impermeable layer and hence a greater absorption into the lower layers of the caprock that will exhibit higher concentrations. This is true if there is no biogenic or thermogenic gas produced within the same shale.

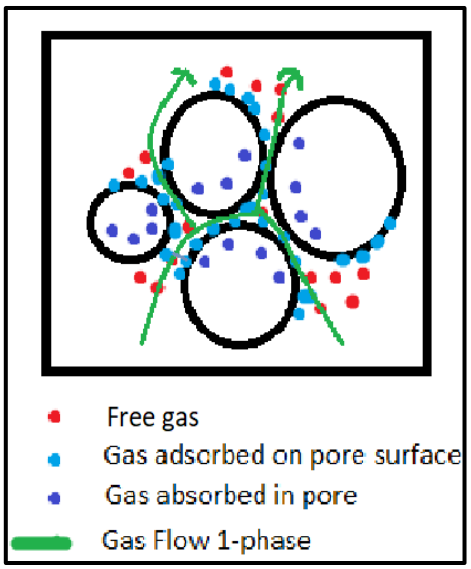


Fig 2-2 Gas types in caprock matrix

The second driving force is present when pressure is above or below hydrostatic. Fluid pressure below hydrostatic is termed under pressure. Fluid pressures in excess of hydrostatic are termed overpressures. Overpressures in sedimentary basins tent to be more common than under pressures. Hypothetical mechanisms for the creation of abnormal fluid pressures in the Earth's crust in literature are: arrested compaction of shale, aqua thermal effects, tectonic phenomena, thermal cracking of organic matter (may cause an increase in the volume of fluids, which would in turn cause an increase in pressure). The first is the most commonly accepted cause. Compaction requires the expulsion of pore water. When clays first start to compact, they are quite permeable and most of the water moves upward. As compaction continues, however, the clay flakes become parallel, reducing vertical permeability. Sands and silts compact less than clays and shales and can maintain permeability to greater depths. If there is a silty or sandy bed within a few feet of the shales, the shales continue a normal compaction trend. However, if no sandy beds are present, the water remains in the shale pores. As additional overburden is deposited, the shale then has to sustain all or part of the additional weight. This results in high pressure in the shale pore water. If there is a small, isolated sand body enclosed by the shale, whatever fluid it

contains (water, oil, or gas), will share the same pressure. The fact that overpressures have been maintained for hundreds of millions of years over small vertical intervals indicates that the permeability of the enclosing shales can be virtually zero. (Barker & Horsfield, 1982)

2.1.1 Knudsen Number and Flow Regime

The Knudsen Number (*Kn*) is used to characterize types of gas flow that occurs within the shale matrix. This is a dimensionless number defined as the ratio of the molecular mean free path length, λ, to representative physical length scale, d, (often pore diameter).

$$Kn = \frac{\lambda}{d} \qquad \lambda = \frac{k_b T}{\sqrt{2}\pi d_m^2 P} \qquad \textit{Eq. 2-1}$$

where k_b is Boltzmann constant (1.38064x1023 J/K), d_m is a characteristic dimension of the gas molecule, T is temperature and P is the pressure.

The Knudsen number is used to determinate different flow regimes. When *Kn* is near or greater than one, the mean free path of a molecule is comparable to a length scale of the problem, and the continuum assumption of fluid mechanics is no longer a good approximation.

We can so define four regime types:

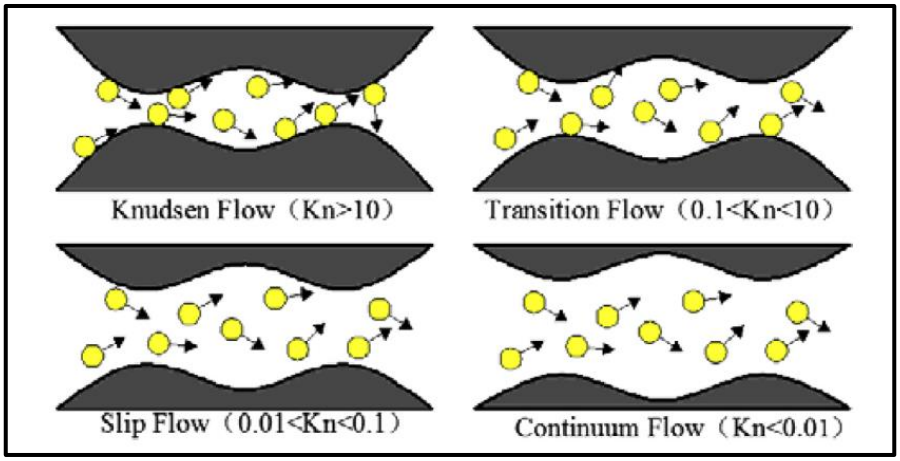


Fig 2-3 Different Molecules Behaviors for Kn (Keliu Wua, 2016)

In the following we divide the diverse flow regimes:

- Kn > 10: Free Molecular Flow

  Under certain circumstances, the continuum hypothesis is inappropriate. Such circumstances occur when the distance between the molecules or, more correctly, the mean free path between collisions with other molecules, λ, is comparable with physical dimension of the flow channel. In this region, the Knudsen diffusion and the Surface diffusion are predominant.

- 0.1 < Kn < 10: Transition Flow

  In the transition regime, an intermediate behavior between the continuum and free-molecule behavior develops. It is generally described by the Knudsen diffusion and it is less affected by slip flow.

- 0.001 < Kn < 0.1: Slip Flow

  In the Slip Flow regime, the behavior naturally is still intermediate between the continuum and free-molecule behavior but the slip flow is the most important contribution to transport compared to diffusive terms.

- Kn < 0.001: Continuum Flow

  The continuum assumption, an idealization of continuum mechanics under which fluids can be treated as continuous, is valid and so the flow is dominated by the Darcy's law.

  The flow region in low permeability rocks is characterized by a Knudsen number between 0.002 and 6 and thus between transition region and the slip flow region.

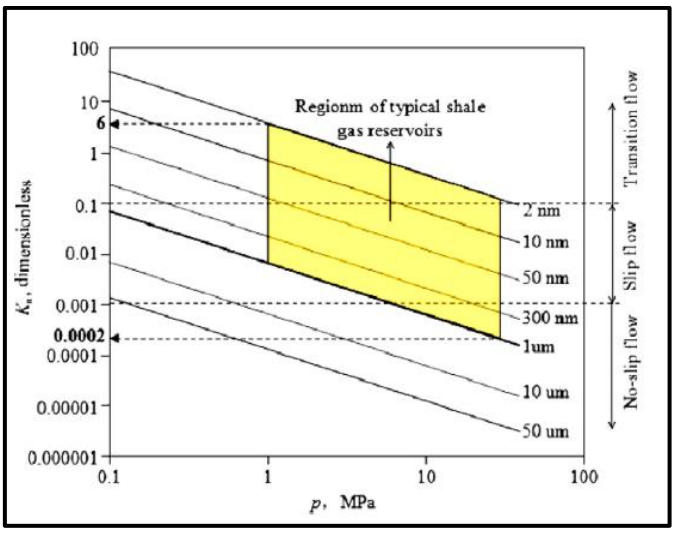


Fig 2-4 Shale Knudsen Range (Keliu Wua, 2016)

2.1.2 Slip Flow

We assume the equation derived on basis of the Hagen-Poiseuille equation. The classical reformulation of it with slip correction is given by:

$$Q = \frac{\pi {r_0}^4}{16\mu L}(P_1^2 - P_2^2)\left[1 + 4\left(\frac{2}{f} - 1\right)\frac{\lambda_m}{r_0}\right] \qquad \textit{Eq. 2-2}$$

Here Q is the volumetric flow rate; r_0 is the pore radius; P_1 and P_2 are the downstream and upstream pressures, respectively; L is the length between downstream and upstream, λ_m is the mean free path length and μ is the dynamic gas viscosity; f is the fraction of gas molecules diffusely reflected in pore walls (Maxwell, 1890); (1 - f) Represents the fraction which reflects the complementary. For instance, f = 0.5 means the pore surface acts as if it is half perfectly reflecting and half perfectly absorbent.

2.1.3 Knudsen Diffusion

When the mean free path of gas molecules is on the same order as the tube dimensions, free-molecular diffusion (i.e. Knudsen Diffusion) becomes

important. Due to the influence of walls, Knudsen diffusion includes the effect of the porous medium. The molecular flux of gas due to Knudsen diffusion is given by the general diffusion equation:

$$J_D = -D_K \frac{\partial C}{\partial x} \qquad \textit{Eq. 2-3}$$

where D_K is the Knudsen diffusion coefficient. It can be esteemed as:

$$D_K = \frac{d}{3}\sqrt{\frac{8RT}{\pi M}} \qquad \textit{Eq. 2-4}$$

where M represents the molecular weight of gas, and *d* is the mean pore size of the porous media. R is gas universal constant the and T is the temperature.

Generally, the Knudsen process is significant only at low pressures and small pore diameters. However, instances exist where both Knudsen diffusion and molecular diffusion are important. If we consider that Knudsen diffusion and molecular diffusion compete with one another by a "resistances in series" approach, then the total diffusivity of gas D_{tot} is determined as

$$\frac{1}{D_{tot}} \cong \frac{1}{D_{fick}} + \frac{1}{D_K}. \qquad \textit{Eq. 2-5}$$

The above relationships for the effective diffusion coefficient are based on diffusion within straight and cylindrical pores aligned in a parallel array. However, in most porous media, pores of various diameters are twisted and interconnected with one another, and the path for diffusion of the gas molecules within the pores is "tortuous". If an average pore diameter is assumed, reasonable approximation for the effective diffusion coefficient in random pores is given by:

$$D_{tot}^{eff} = \frac{\varphi}{\tau} D_{tot} \qquad \textit{Eq. 2-6}$$

where D_{tot}^{eff} is the effective total diffusion coefficient, φ and τ are the porosity and tortuosity of the porous media, respectively.

The four possible types of pore diffusion are illustrated in Fig 2-5 , with each featured with their respective diffusivity correlation. The first three, pure molecular diffusion, pure Knudsen diffusion, and Knudsen and molecular

combined diffusion, are based on diffusion within straight and cylindrical pores that are aligned in parallel array. The fourth involves diffusion via "tortuous paths" that exist within the compacted solid.

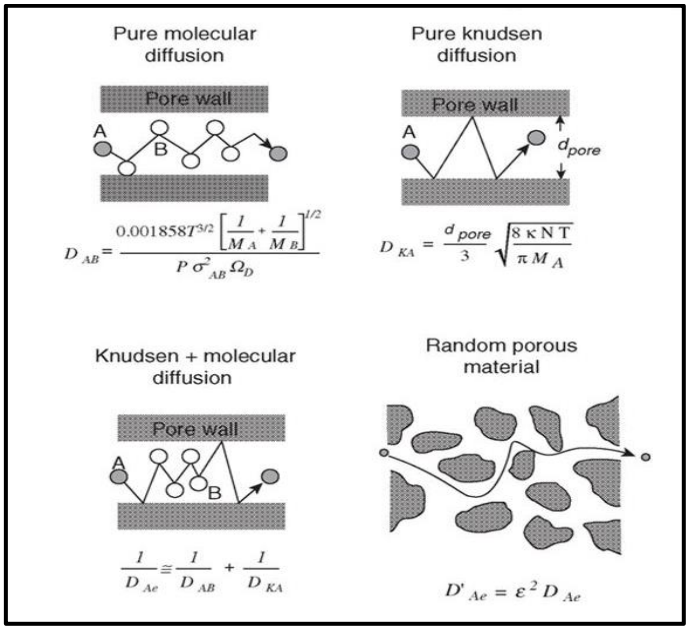


Fig 2-5 Different diffusion types (Welty, Wicks, Rorrer, & Wilson, 2007)

2.2 DISCONTINUOUS TWO PHASE FLOW

Although diffusion and advection of dissolved gas in pore water exist at any level of gas pressure, in comparison to the two-phase flow process, the amount of transported gas is minimal (Song & Zhang, 2013). When the gas pressure exceeds the summation of reservoir pressure and entry capillary pressure, volumetric or Darcy flow is observed.

Natural gas begins to displace the pore water and a two-phase flow process is observed within the caprock. Water-saturated shale caprocks are distinct from the gas shale. The gas in caprocks is not generated within the shale. The pore water in the caprocks has to be driven out first. This displacement requires overcoming the entry capillary pressure. The entry capillary pressure in "micropores" of the shale matrix is higher than any fracture, so water would be expected to remain in the micropores even after gas from reservoir invasion

into any fractures (Wang & Peng, 2014). Further, most of caprocks are observed to be a fractured porous-medium-composed of fracture network and matrix. Two-phase flow is only observed in the fracture network.

The principal driving forces that we have in this flow type is the possible overpressure and the buoyancy. These forces are contrasted by the capillary entry pressure and by the drag forces generated by the flow inside the fractures.

The entry capillary pressure or the threshold pressure constrains the capacity of gas containment of a satured porous medium. This pressure is determined by the pore geometry and gas/rock/water wettability. According to Kelvin's law, the entry capillary pressure, P_c , is given by:

$$P_c = \frac{2\sigma cos\Theta}{r} \qquad \textit{Eq. 2-7}$$

where r is the representative radius of the pores in the caprock, θ is the conctact angle between the two phases and σ is the interfacial tension. For a fracture with the aperture b_i and the aperture variation with deformation, Δr, is (Wang, Liu, & Kabir, 2013):

$$\Delta r = b_i \left[1 + n \frac{1-R_f}{\varphi_{f0}}\right] \Delta\varepsilon_{ej} \qquad \textit{Eq. 2-8}$$

where n is the spatial dimension, R_f is a strain ratio, φ_{f0} is the porosity of the fracture and $\Delta\varepsilon_{ej}$ is the change of effective strain in the jth direction.

If the interfacial tension and wettability do not change with deformation, the current gas entry pressure is evolved as:

$$P_c = \frac{\frac{2\sigma cos\Theta}{b_i}}{1+\left[1+n\frac{1-R_f}{\varphi_{f0}}\right]\Delta\varepsilon_{ej}} \qquad \textit{Eq. 2-9}$$

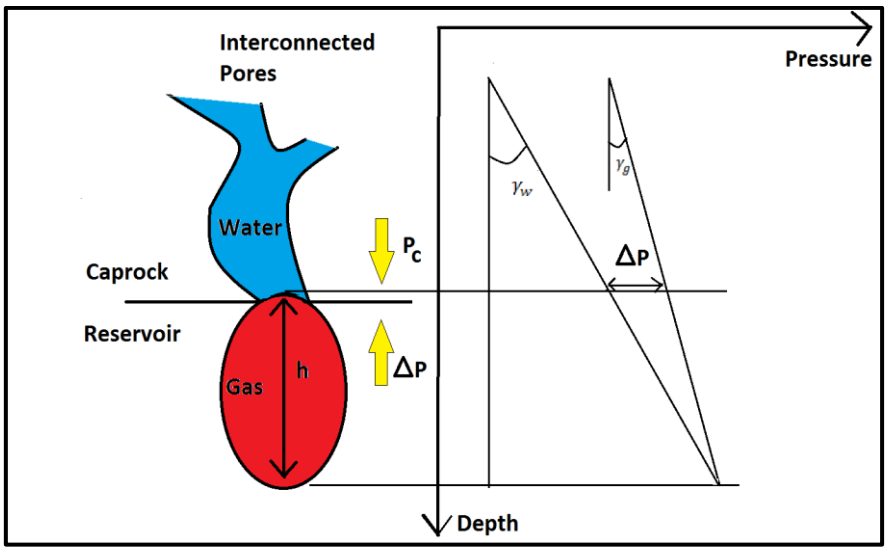


Fig 2-6 Capillary effect

The principle consideration to understand the process is that the pressure within the gas bubble has an inferior gradient respect hydraulic gradient and we have the same pressure between gas and water at the bottom of the bubble; so, we have an excess pressure at the top of it that compute as (Al-Bazali an J.Zhang)

$$\Delta p = (\gamma_w - \gamma_g)h \qquad \textit{Eq. 2-10}$$

γ_w is the water specific weight while γ_g is the gas specific weight.

When the excess pressure exceeds the capillary entry pressure we have the gas intrusion in the fracture.

To determinate the minimum vertical height of the bubble that it can exist before have an intrusion in the capillary, we can compare the equation above and we obtain:

$$h > \frac{2\sigma cos\theta}{r(\gamma_w - \gamma_g)}. \qquad \textit{Eq. 2-11}$$

2.3 DIFFUSION

Diffusion is the process by which matter is transported from one part of a system to another as a result of random molecular motions. Random molecular motions is when no molecule has a preferred direction of motion, this has to be reconciled with the fact that a transfer of gas molecules from the region of higher to that of lower concentration is nevertheless observed. Consider any horizontal section in the solution and two thin, equal, elements of volume one just below and one just above the section. Though it is not possible to say which way any particular gas molecule will move in a given interval of time, it can be said that on the average a definite fraction of the molecules in the lower element of volume will cross the section from below, and the same fraction of molecules in the upper element will cross the section from above, in a given time. Thus, simply because there are more gas molecules in the lower element than in the upper one, there is a net transfer from the lower to the upper side of the section as a result of random molecular motions.

This mechanism governs the total diffusion process, as the mean free path of gas molecules is at least one order larger than the pore diameter of the porous media. Fick's law is the most popular approach to evaluate gas diffusion in clear fluids and gases due to its simplicity. It has two forms. Fick's first law describes the correlation between the diffusive flux of a gas component and the concentration gradient under steady-state conditions. Fick's second law relates the unsteady diffusive flux to concentration gradient.

2.3.1 Fick's Laws

Transfer of heat by conduction is also due to random molecular motions, and there is an obvious analogy between the two processes. This was recognized by Fick (1855), who first use the mathematical equation of heat conduction derived some years earlier by Fourier. The theory of diffusion in isotropic substances (J.C.Cortes, 2005) is therefore based on the hypothesis that the rate

of transfer of diffusing substance through unit area of a section is proportional to the concentration gradient measured normal to the section:

$$J = -D\frac{\partial C}{\partial x} \qquad \textit{Eq. 2-12}$$

where J is the rate of transfer per unit area of section, c the concentration of diffusing substance, x the space coordinate measured normal to the section, and D is the diffusion coefficient (Crank, 1975).

The above forms of Fick's law are appropriate for clear fluids or gases. For application in porous media, diffusion coefficient is written as:

$$D^{eff} = \frac{\varphi}{\tau}D \qquad \textit{Eq. 2-13}$$

Where D^{eff} is the effective diffusion coefficient, φ and τ is the porosity and tortuosity of the porous media, respectively. The same of eq.2-9. I would erase the 2.9 paragraph and keep this….

Diffusion coefficient can be estimated from the Champman-Enkog theory (Hirschfelder, Bird, & Curtiss, 1954) as:

$$D = \frac{0.00186T^{3/2}}{P\sigma_i^2\delta}\left(\frac{2}{M_i}\right)^{0.5} \qquad \textit{Eq. 2-14}$$

where D is in cm^2/s T is the absolute temperature in Kelvin, P is the pressure in atmospheres and M_i is the molecular weights of gas. The quantities σ and δ are molecular property characteristics of the detailed theory. The first is the collision diameter given in angstroms, which is equal to 3.758 for the methane, while the second, dimensionless, is typically on the order of 1 for gas species in porous media.

The fundamental differential equation of diffusion in an isotropic medium is derived from the first Fick's equation above. Using the first Fick's equation and the continuity equation and if the diffusion coefficient is constant we obtain the second Fick's Law:

$$\frac{\partial C}{\partial t} = D\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2}\right). \qquad \textit{Eq. 2-15}$$

2.3.1.1 Problem Settings

Here we define the problem of interest with its geometry. We want to investigate the diffusion behavior within a caprock modelling this latter with a 1D model and with a Dirichlet boundary conditions.

The whole general deterministic problem is the following parabolic equation:

$$\frac{\partial C(z,t)}{\partial t} = \frac{\partial D(z)\partial C(z,t)}{\partial z^2}, \quad 0 < z < L, t > 0 \qquad \textit{Eq. 2-16}$$

with boundary and initial condition given by:

$$C(0,t) = c_t, \qquad t > 0$$

$$C(L,t) = c_b, \qquad t > 0$$

$$C(z,0) = c_0(z), \qquad 0 < z < L$$

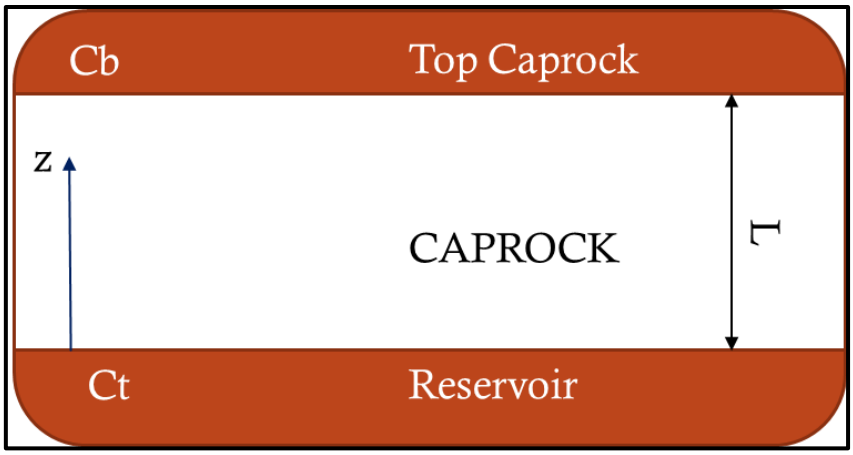


Fig 2-7 Cap rock: sketch of the Model

where c_t and c_b are respectively the boundary conditions that coincide with the concentration inside the reservoir and the concentration at the top of the caprock, considered constants in time. c_0 is the initial condition.

2.3.1.2 Analytical Solution of the Deterministic Problem

Firstly we consider an homogeneous diffusion coefficient D and constant boundary condition. An analytical solution of problem (see Appendix A) can be written in the form of a trigonometrical series (Crank, 1975):

$$C(z,t) = c_t + \frac{c_b - c_t}{L} z + \sum_{n=1}^{\infty} \frac{2}{L} \int_0^L \left(c_o(z) - c_b \frac{z'}{L} - c_t \frac{(L-z')}{L} \right) sin(z'\lambda) dz' \, sin(z\lambda) \, e^{-\lambda^2 D t} \qquad \textit{Eq. 2-17}$$

with

$$\lambda = \frac{n\pi}{L}$$

We can write a generic solution for this problem in a form like this:

$$C(z,t) = \omega(z,t) + \sum_{n=1}^{\infty} k_n u_n(z,t) \qquad \textit{Eq. 2-18}$$

where:

$$\omega(z,t) = c_t + (c_b - c_t) \frac{z}{L} \qquad \textit{Eq. 2-19}$$

and

$$k_n = \frac{2}{L} \int_0^L (c_0(z) - \omega(z,t)) \, sin \frac{n\pi z'}{L} dz' \qquad \textit{Eq. 2-20}$$

$$u_n = sin(\frac{n\pi z}{L}) \, e^{-n^2 D t \pi^2 / L^2} \qquad \textit{Eq. 2-21}$$

This form is more convenient (WHY? If you said it you have to explain… otherwise change term ;)) and it can be used for other boundary condition types just changing the $\omega(z,t)$ term.

The series is convergent in mean square sense for each (z,t) ϵ [0,L] x [0,inf] and thus, the process u(z,t) is well defined (J.C.Cortes, 2005), see Appendix B for details.

Note that in the case of $C(z,0) = c_0$ Eq 2-17 reduce to the analytical solution in (Hantschel & Kauerauf, 2009) :

$$C(z,t) = c_t + (c_b - c_t)\frac{z}{L} + \sum_{n=1}^{\infty} 2[c_o - c_t + (-1)^n(c_b - c_o)]\frac{sin\, z\lambda}{\lambda L} e^{-\lambda^2 Dt} \qquad \textit{Eq. 2-22}$$

with $\lambda = \frac{n\pi}{L}$.

Note that the time dependecy scales quadratical with height L due tot the form of exponent $\lambda^2 Dt = Dt\mathrm{n}^2\pi^2/\mathrm{h}^2$. Hence a caprock with doubled thickness of 400 m needs four times as long and with 10 times thickness of 2 km about 100 times as long for the convergence to the stationary solution.

2.3.1.3 Gas Flux

The cumulative volume $Q(z,t)$ which is flown through a horizontal plane of unit size at depth z is given by:

$$Q(z,t) = -\int_0^t D\frac{\partial C(z,t')}{\partial z}dt \qquad \textit{Eq. 2-23}$$

that in the case of $C(z,0) = c_0$ becomes

$$Q(z,t) = -\int_0^t q(z,t)dt \qquad \textit{Eq. 2-24}$$

where

$$q(z,t) = -D\left[(c_b - c_t)\frac{1}{L} - \sum_{n=1}^{\infty}\frac{2}{\pi n}\int_0^L \left(c_o(z') - c_b\frac{z'}{L} - c_t\frac{(L-z')}{L}\right) sin(z'\lambda)dz'\, cos(z\lambda)\, e^{-\lambda^2 Dt}\right] = -D\left[(c_b - c_t)\frac{1}{L} - \sum_{n=1}^{\infty}\frac{L}{\pi n}k_n\, cos(z\lambda)\, e^{-\lambda^2 Dt}\right] \qquad \textit{Eq. 2-25}$$

is the instantaneous flux at time *t*.

Formulas above, as (Hantschel & Kauerauf, 2009) , can be used to determinate the gas volume in caprock. In fact if we measure the cumulative flux at z=0 and z=L for a caprock of height equal to L we can see that:

$$\frac{Q(0,t)}{c_t L} = -\frac{1}{c_t L}\int_0^t D\frac{\partial C(0,t')}{\partial z}dt = \frac{Dt}{L^2}c_t L - \frac{1}{6} - 2\sum_{n=1}^{\infty}\frac{[(-1)^n]}{\lambda^2 L^2}e^{-\lambda^2 Dt} \qquad \textit{Eq. 2-26}$$

$$\frac{Q(L,t)}{c_t L} = -\frac{1}{c_t L}\int_0^t D\frac{\partial C(L,t')}{\partial z}dt = \frac{Dt}{L^2}c_t L + \frac{1}{3} - 2\sum_{n=1}^{\infty}\frac{[1]}{\lambda^2 L^2}e^{-\lambda^2 Dt} \qquad \textit{Eq. 2-27}$$

And computing the difference for the flux for t→ ∞

$$\Delta Q = Q(L,t) - Q(0,t) = \frac{Lc_t}{2} \qquad \textit{Eq. 2-28}$$

We can so link the gas volume only to the height and to the concentration in reservoir.

2.4 SENSITIVITY ANALYSIS

Sensitivity analysis (SA) allows the identification of the parameter or set of parameters that have the greatest influence on the model output. Consequently it provides useful insight on which model input uncertainty contributes most to the variability of the model output. Sensitivity analysis has been widely used in many fields, such as risk assessment, economics and engineering. The application of sensitivity analysis can be summarized as: understanding the input–output relationship, determining how uncertainty in structural model parameters contribute to the overall variability of the model output, identifying the important and influential parameters that drive model outputs and magnitudes, and guiding future experimental designs.

SA methods are divided into two main categories:

- Local Sensitivity Analysis (LSA), that gives information about the output variation, round of a fixed valued, respect the input parameters.
- Global Sensitivity Analysis (GSA), that gives information about the output uncertainty, due to the input parameters uncertainty.

The GSA is a useful tool in the study of a mathematical model. In the framework of GSA we can distinguish two categories (Sudret B. , 2008):

- Regression methods: the regression coefficients are based on the linear regression of the input vector and the output. These are significant if the link between the inputs and the output is linear or approximatively linear.
- Variance based methods: These methods aim to the decomposition of the output variance as a summed of the single input contribution and of the mutual iteration between theme.

They are called also ANOVA technique (ANalysis Of VAriance). The Sobol Indices and Multiple Moment Based belong to these methods (Sobol I. , 2001) (Sobol I. , 1993) (Dell'Oca, Riva, & Guadagnini, 2017).

The next Sections are dedicated to the theoretical introduction of GSA Variance Based indexes that will be used later for the model object of the thesis.

2.4.1 Sobol's Indices

Sobol sensitivity analysis is intended to determine how much of the variability in model output is dependent upon each of the input parameters, either upon a single parameter or upon an interaction between different parameters. The decomposition of the output variance in a Sobol sensitivity analysis employs the same principal as the classical analysis of variance in a factorial design. It should be noted that Sobol sensitivity analysis is not intended to identify the cause of the input variability. It just indicates what impact it will have on model output. As a consequence, it cannot be used to determine the sources of variance.

In order to understand how the output variance can be attributed to individual input variables and the interaction between each of the input variables, the total-order, first-order, second-order, and higher-order sensitivity indices are calculated to accurately reflect the influence of the individual input, and the interaction between them.

Consider our model function $C = f(x)$, that rapresents the relation between our concentration output and the vector x of the N independent stochastic parameters (three in this case: cb, c0, D). Suppose that the latter are defined in the N-dimensional unit hypercube, IM. If $f(x)$, is integrable, we can estimate the sensitivity of $f(x)$ with to respect to different variables.

The function $f(x)$ can be expanded as follows (Sobol I. , 1993):

$$f = f_0 + \sum_{i=1}^{N} f_i(x_i) + \sum_{1 \le i \le j \le N}^{N} f_{ij}(x_i, x_j) + \cdots + f_{1,\ldots,N}(x_1, \ldots, x_N) \qquad \textit{Eq. 2-29}$$

The sum is Eq. 2-27 is an expansion into summands of different dimensions. f_0 is constant and the integrals of the summands $f_{i1}, \dots, i_s$, with resperct to any of their "own" variables are zero, i.e.

$$\int_0^1 f_{i1\dots,i_s}, (x_{i1}, \dots, x_{is}) dx_{ik} = 0 \qquad 1 \leq k \leq s \qquad \textit{Eq. 2-30}$$

It follows from this definition that

$$f_0 = \int_{k^n} f(x) dx \qquad \textit{Eq. 2-31}$$

And all the summands on the right-hand side of Eq. 2-26 are orthogonal. i.e., if $(i_1 \dots i_s) \neq (j_1 \dots j_s)$ then

$$\int_{k^n} f_{i1,\dots,i_s} f_{i1,\dots,i_s} dx = 0 \qquad \textit{Eq. 2-32}$$

Since at least one of the indices $i1, \dots, i_s$ $i1, \dots, i_s$ is not repeated and the integral wot respect to that variable vanishes because of By theorem 1 in (Sobol I. , 1993) , there exists a unique expansion of Eq.2-26 for any function $f(x)$ integrable in k^n. The principal property of Sobol representation are therefore:

1. The decomposition summation contains $\sum_{j=1}^{n} \binom{n}{j} = 2n - 1$ terms.
2. The constant f_0 is the mean value of function $f_0 = \int_{k^n} f(x) dx$
3. The summations are mutual orthogonal if $\int_{k^n} f_{i1,\dots,i_s}(x_i \dots x_{i_s}) \cdot f_{j1,\dots,j_s}(x_j \dots x_{j_s}) dx = 0$ for $(i_1 \dots i_s) \neq (j_1 \dots j_s)$

Assume now that the input parameters are independent and uniformly distributed random variables in [0,1]. The output $C = f(x)$,is also a random variable and its variance is:

$$D = Var[f(x)] = \int_{k^n} f^2(x)\varphi_d(x)dx - f_0^2 \qquad \textit{Eq. 2-33}$$

Where $\varphi_d(x)$ is the joint probability density function that for a uniform distribution is equal to the product of the probability distributions of the individual independent variables in x.
W we can then break it down the total variance as:

$$D = \sum_{i=1}^{n} D_i + \sum_{1\le i<j\le n}^{n} D_{ij} + \cdots + D_{1,2,..n} \qquad \textit{Eq. 2-34}$$

Where the partial contributions are defined as:

$$D_{i1,\dots,i_s} = \int f^2_{i1,\dots,i_s}(x_{i1},\dots,x_{i_s})dx_i \dots dx_{i_s}, \quad 1 \le i_1 < \cdots < i_s \le n, s = 1,\dots,n \qquad \textit{Eq. 2-35}$$

Sobol's Indices are defined as:

$$S_{i_1..i_s} = \frac{D_{i_1\dots i_s}}{D} \qquad \textit{Eq. 2-36}$$

By the definition e remembering the Eq. 2-27 the index must be satisfied the follow equation:

$$\sum_{i=1}^{n} S_i + \sum_{1\le i<j\le n}^{n} S_{ij} + \cdots + S_{ij} = 1 \qquad \textit{Eq. 2-37}$$

Every indices $S_{i_{1..}i_s}$ is a sensitivity value that describe the variance part respect the total variance for the relative stochastic variable.
The Index of the first order define the influence of every parameter taken alone. The superior indices evaluate the combination effect of the theme.
We can also define the total sensibility indices, that taken into account every variance contribution of each stochastic variables, as alone as combination:

$$S_i^T = \sum_{J_i} D_{i_1 \ldots i_s} \qquad \textit{Eq. 2-38}$$

Sobol's indices are good sensitivity value descriptors because we don't hypothesize linearity or monotony for the output function respect the input random variables.

2.4.2 Multiple-moment-based metrics for GSA

The follow considerations are taken by the (Dell'Oca, Riva, & Guadagnini, 2017). In contrast with traditional variance-based GSA methods, as the Sobol's indices of the capitol before, they quantify changes in pdf of the output through its first four statistical moments, mean $E[CON]$, variance $VAR[CON]$, skewness, $\gamma[CON]$, and kurtosis, $k[CON]$. The latter is an indicator of the behavior of the tails of the pdf of the concentration and is particularly useful in the context of risk analysis, while $\gamma[CON]$ quantifying the asymmetry of the pdf of y.
The authors introduced the following quantities for the general input x_i:

$$AMAE_{x_i} = \frac{1}{|y_0|} E[|y_0 - E[CON|x_i]|] \qquad \textit{Eq. 2-39}$$

$$AMAV_{x_i} = \frac{1}{V[CON]} E[|V[CON] - V[CON|x_i]|] \qquad \textit{Eq. 2-40}$$

$$AMAG_{x_i} = \frac{1}{\gamma[CON]} E[|\gamma[CON] - \gamma[CON|x_i]|] \qquad \textit{Eq. 2-41}$$

$$AMAk_{x_i} = \frac{1}{k[CON]} E[|k[CON] - k[CON|x_i]|] \qquad \textit{Eq. 2-42}$$

3 METHODOLOGY

This chapter illustrates the results for the problem stated previously: firstly the homogeneous case and secondly the heterogeneous.

Initially we consider the caprock as constituted by a homogeneous rock, taking into account a diffusion coefficient constant both in space and in time. This was one of the principal hypothesis to solve analytically the problem (Appendix A).

In the following discussion we will examine the problem initially assuming only one of the three input variables (diffusion coefficient, reservoir boundary condition and initial condition) as uniformly random distributed, keeping constant the others two. We calculate the analytical equations of mean and variance for concentration and instantaneous flux and we compare them with the solutions obtained from the Montecarlo simulations and from the numerical method (details in Appendix C). As in literature we compute the analytical mean and variance from the mean Calculus Theory (Appendix B). We can say therefore that the expressions computed in our work are valid in general, if the inputs are uniformly distributed.

At the end we will consider all three inputs as random variables uniformly distributed.

We calculate the analytical expression for the mean and the variance for the concentration and instantaneous flux. Finally we perform a global sensitivity analysis of the problem computing both the Sobol indices (Sobol I. , 2001) (Sobol I. , 1993) and the most recent AMAE, AMAV, AMAG and AMAK indices (Dell'Oca, Riva, & Guadagnini, 2017).

3.1 PROBLEM WITH STOCHASTIC INITIAL CONDITION

Initial gas concentration inside the caprock is typically uncertain. This problem is due to the local gas genesis: biogenesis and thermogenesis (shale gas). In order to take this into account, we assume a given mean and

uncertainty of measurement, and we use a probability distribution function for the initial condition. We computed the general analytical solutions considering the initial condition as a space dependent random variable. After, we take it constant in space and with a uniform distribution. We compute the analytical solutions and we compare them with Montecarlo simulations and numerical solutions (see Appendix C).

Analogously to Soong (1973), we consider our problem (Eq.2-16)where:

1. $c_0(z)$ is the mean value of the initial condition.
2. $c'_0(z) = c_0(z) - E[c_0(z)]$ is the fluctuation of the random variable
3. *f(x)* is the pdf of $c_0(z)$

we can split the homogeneous problem for D constant and constant Dirichlet boundary condition into two new problems: a stochastic one and a deterministic one.

The stochastic problem reads:

$$\frac{\partial C(z,t)}{\partial t} = D\frac{\partial^2 C(z,t)}{\partial z^2}, \quad 0 < z < L, t > 0 \qquad \textit{Eq. 3-1}$$

with boundary and initial condition given by:

$$C(0,t) = c_t, \ t > 0$$

$$C(L,t) = c_b, \ t > 0$$

$$C(z,0) = c_0(z) - E[c_0(z)] \ \ 0 < z < L$$

and the deterministic problem is:

$$\frac{\partial C(z,t)}{\partial t} = D\frac{\partial^2 C(z,t)}{\partial z^2}, \quad 0 < z < L, t > 0 \qquad \textit{Eq. 3-2}$$

$$C(0,t) = c_t, \ t > 0$$

$$C(L,t) = c_b, \qquad t > 0$$

$$C(z,0) = E[c_0(z)] \ \ 0 < z < L$$

The sum of solutions of these problems, for the linearity of the differential equation, is the solution of the general problem (J.C.Cortes, 2005).

The infiniteness of the series (Eq.2-18) in the deterministic solution suggests taking approximate process obtained by truncating the series. Given a positive integer N, we consider the stochastic process $C_N(z,t)$ defined by:

$$C_N(z,t) = \omega(z,t) + \sum_{n=1}^{N} k_n u_n(z,t) \qquad \textit{Eq. 3-3}$$

Where $\omega(z,t)$ is defined in Eq.2-19, k_n in Eq.2-20 and u_n in Eq.2-21.
Only the k_n terms are stochastic, while u_n are a deterministic function. Using for k_n the mean square random variable property we can compute the mean and variance of the concentration and of the instantaneous flux as:

$$E[C_N(z,t)] = \omega(z,t) + \sum_{n=1}^{N} sin\frac{n\pi z}{L} e^{-n^2 Dt\pi^2/L^2} \left[\frac{2}{L}\int_0^L E[c_0(z')]\, sin\frac{n\pi z'}{L} dz' - \frac{2}{L}\int_0^L \omega(z,t) sin\frac{n\pi z'}{L} dz'\right] \qquad \textit{Eq. 3-4}$$

and

$$VAR[C_N(z,t)] = VAR[\omega(z,t)] + J(i,j) \qquad \textit{Eq. 3-5}$$

where:

$$J(i,j) = \frac{4}{L^2}\int_0^L \int_0^L sin\frac{i\pi z'}{L} sin\frac{j\pi z''}{L} \Gamma_{ff}(z',z'')dz'\, dz'' \qquad \textit{Eq. 3-6}$$

where $\Gamma_{c_0 c_0}(z',z'')$ is the autocorrelation function that can be written as:

$$\Gamma_{c_0 c_0}(z',z'') = E[c_0(z')c_0(z'')] - E[c_0(z')]E[c_0(z'')] \qquad \textit{Eq. 3-7}$$

The mean and the variance of the flux $q_N(z,t)$ read, respectively:

$$E[q_N(z,t)] = -D\left[(c_b - c_t)\frac{1}{L} + \sum_{n=1}^{\infty}\frac{L}{\pi n} E[k_n]\, cos(z\lambda)\, e^{-\lambda^2 Dt} + + \sum_{n=1}^{\infty}\frac{2}{L}\left(E[c_0(z')] - c_b - \omega(z,t)\right) sin(z\lambda)\, u_n\right] = -D\left[(c_b - c_t)\frac{1}{L} + + \sum_{n=1}^{\infty}\frac{2}{\pi n}\int_0^L \left(E[c_0(z')] - c_b\frac{z'}{L} -\right.$$

$$c_t\,\frac{(L-z')}{L}\Big)\,sin(z'\lambda)dz'\,cos(z\lambda)\,e^{-\lambda^2 Dt}+\sum_{n=1}^{\infty}\frac{2}{L}\Big(E[c_0(z\,)]-c_b\,\frac{z'}{L}-$$

$$c_t\,\frac{(L-z')}{L}\Big)\,sin(z'\lambda)\,sin(z\lambda)\,e^{-\lambda^2 Dt}\Big] \qquad \textit{Eq. 3-8}$$

and

$$VAR[q_N(z,t)]=$$

$$-D[\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{4}{L^2}e^{-(i^2+i^2)Dt\pi^2/L^2}\,cos\frac{i\pi z}{L}cos\frac{j\pi z}{L}\,[JJ(i,j)]] \qquad \textit{Eq. 3-9}$$

where

$$JJ(i,j)=VAR[c_o-(-1)^n c_o]=E[c_0^2+c_0^2(-1)^{2n}-2c_0^2(-1)^n]-$$

$$E[c_o-(-1)^n c_o]*E\big[c_o-(-1)^j c_o\big]=E[c_0^2(z')][(-1)^{2n}-$$

$$(-1)^n+1]-(E[c_0(z')]-E[c_0(z')](-1)^n)^2\,. \qquad \textit{Eq. 3-10}$$

Note that $c_0(z)$ is a stochastic problem by hypothesis and hence k_n is a second random variable for each n ≥ 1 and the convergence of the $C_N(z,t)$ must be regarded in the mean square sense (J.C.Cortes, 2005). This, from the mean square calculus theory, is respected if k_n is a well-defined mean square integral, and so if $\Gamma_{c_0c_0}(z',z'')$ satisfies:

$$\frac{4}{L^2}\int_0^L\int_0^L sin\frac{i\pi z'}{L}sin\frac{j\pi z''}{L}\Gamma_{ff}(z',z'')dz'\,dz''<\infty \quad . \qquad \textit{Eq. 3-11}$$

Using the third property of the Mean square Riemann Integral we have:

$$\sum_{n=1}^{N}\|k_n\||u_n(z,t)|\le M\sum_{n=1}^{N}e^{-\frac{n^2Dt\pi^2}{L^2}}<\infty \qquad \textit{Eq. 3-12}$$

with

$$M=2*max\{\|c_0\|;0\le z\le L\} \qquad \textit{Eq. 3-13}$$

And as in (J.C.Cortes, 2005) it can be easily proved that the stochastic process $C_N(z,t)$ is an exact solution process for the problem in exam.

If c_0 is uniform distributed on the interval [a, b] and it doesn't depend on coordinate z, we can solve the integral in Eq. 3-4 and the mean concentration reads

$$E[C_N(z,t)] = c_t + (c_b - c_t)\frac{z}{L} + \sum_{n=1}^{N} sin\frac{n\pi z}{L} e^{-\frac{n^2 D t \pi^2}{L^2}} \left[\frac{2}{L}\left(\frac{b+a}{2}\right)\left(\frac{L^2}{n\pi}(1-(-1)^n)\right) - \frac{2}{n\pi}[c_t + (-1)^n(c_t - c_b)]\right] \quad \textit{Eq. 3-14}$$

and its corresponding variance:

$$VAR[C_N(z,t)] = VAR[\omega(z,t)] + \sum_{i=1}^{N}\sum_{j=1}^{N} u_i(z,t)u_j(z,t)COV\left[k_i, k_j\right] = VAR[\omega(z,t)] + \sum_{i=1}^{N}\sum_{j=1}^{N} u_i(z,t)u_j(z,t)\left[E\left[k_i k_j\right] - E[k_i]E\left[k_j\right]\right] = VAR[\omega(z,t)] + \sum_{i=1}^{N}\sum_{j=1}^{N} e^{-(i^2+i^2)Dt\pi^2/L^2} sin\frac{i\pi z}{L} sin\frac{j\pi z}{L}\ [J(i,j)] \quad \textit{Eq. 3-15}$$

where:

$$J(i,j) = \frac{4}{L^2}\int_0^L \int_0^L sin\frac{i\pi z'}{L} sin\frac{j\pi z''}{L}\Gamma_{ff}(z', z'')dz'\,dz'' = \frac{4}{L^2}\int_0^L \int_0^L sin\frac{i\pi z'}{L} sin\frac{j\pi z''}{L}(E[c_0^2] - E^2[c_0])dz'\,dz'' = \frac{4}{L^2}\int_0^L \int_0^L sin\frac{i\pi z'}{L} sin\frac{j\pi z''}{L}(\frac{(a-b)^2}{12})dz'\,dz'' = \frac{4}{ij\pi^2}\left(\frac{(a-b)^2}{12}\right)\left(1-(-1)^i\right)\left(1-(-1)^j\right). \quad \textit{Eq. 3-16}$$

The same quantities for the flux are:

$$E[q_N(z,t)] = -D\left[(c_b - c_t)\frac{1}{L} + \sum_{n=1}^{\infty} 2\left[\left(\frac{b+a}{2}\right) - c_t + (-1)^n\left(c_b - \left(\frac{b+a}{2}\right)\right)\right]\frac{cos\, z\mu_n}{L} e^{-{\mu_n}^2 Dt}\right] \quad \textit{Eq. 3-17}$$

and

$$VAR[q_N(z,t)] =$$

$$-D[\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{4}{L^2}e^{-(i^2+i^2)Dt\pi^2/L^2}\ cos\frac{i\pi z}{L}cos\frac{j\pi z}{L}\ [JJ(i,j)] \qquad \textit{Eq. 3-18}$$

where

$$JJ(i,j) = VAR[c_o - (-1)^n c_o] = E[c_0^2 + c_0^2(-1)^{2n} - 2c_0^2(-1)^n] -$$

$$E[c_o - (-1)^n c_o] * E\left[c_o - (-1)^j c_o\right] = \frac{(b^3-a^3)}{3(b-a)}[(-1)^{2n} - (-1)^n +$$

$$1] - \left(\left(\frac{b+a}{2}\right) - \left(\frac{b+a}{2}\right)(-1)^n\right)^2 \quad . \qquad \textit{Eq. 3-19}$$

3.1.1 Results

Here we illustrate the results obtained for the concentration and the flux taking into account a variable initial condition c_o. We make vary the initial condition within the interval of extremes a and by using the coefficient of variation (coefvar) and the mean μ:

$$c_o = [a,b] = [\mu - coefvar \cdot \mu\ , \mu + coefvar \cdot \mu] \qquad \textit{Eq. 3-20}$$

Figure 3-1 illustrates the profile mean concentration and its variance versus depth and their relative band of uncertainty (±2σ). Figure 3-1 shows the exact solution computed by using a large number of Montecarlo realizations of deterministic solution (Eq.2-22), the analytical solution obtained by the above formulas (Eq.3-14 and Eq.3-15), and a numerical solution computed by the Euler Finite Difference Method. Figure 3-2 displays the same quantities for the instantaneous flux. We consider the case where the initial condition mean is $c_o = 0.2$, uniform distribution coefficient of variation of c_o equal to 0.25 $[a,b] = [0.15,0.25]$; $c_t = 1$ and a simulation time of 2.5 million years . The diffusion coefficient is considered a constant on all space domain and equal to 10^{-12} m^2/s, that is a typical value for the shale rocks.

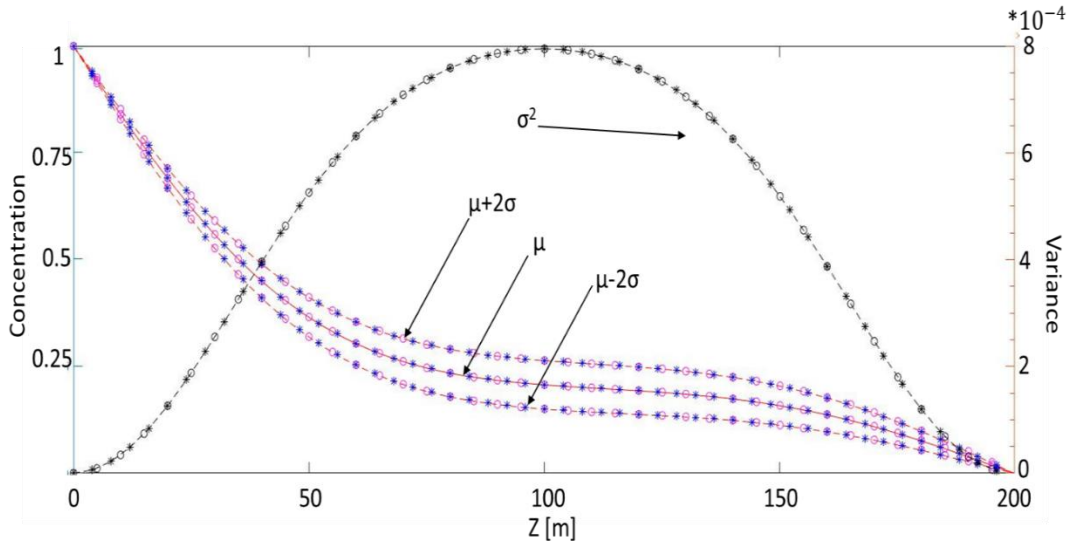


Fig 3-1 Comparison between the three instantaneous concentration solutions, their relative band of uncertainty (±2σ) and the variance (in black) at time t= 2.5 million years in function of the spatial distance with random uniform distributed initial condition. The pink points are the numerical solution (Appendix C), the blue stars are the analytical solution (Eq.3-15 Eq.3-16) while the continue red line are the exact Monte Carlo solution (Eq.2-18)

.

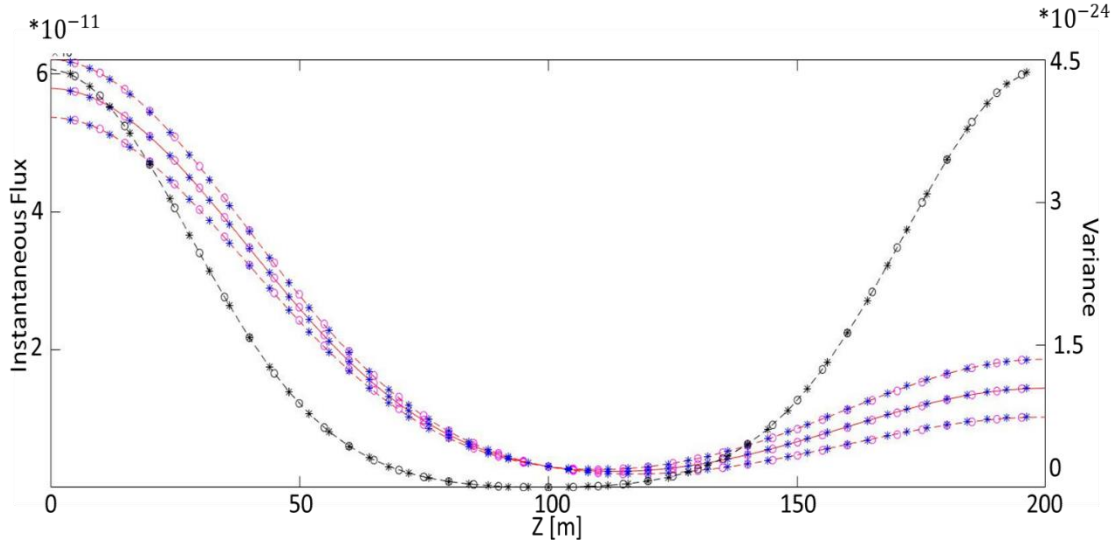


Fig 3-2 Comparison between the three instantaneous flux solutions, their relative band of uncertainty (±2σ) and the variance (in black) at time t= 2.5 million years in function of the spatial distance with random uniform distributed initial condition. The pink points are the numerical solution, the blue stars are the

analytical solution (Eq.3-17 Eq.3-18) while the continue red line are the exact Monte Carlo solution (eq. 2-25)

.

Figures 3-1 and 3-2 show that the three solutions: numerical, analytical and the one obtained by using Monte Carlo simulation, coincide. We observe that the concentration profile has a flex point when the mean concentration reaches the value of the initial concentration and at the same point we have the minimum flux. The position of the minimum mean flux, or, equivalently, of the flex point in the concentration profiles, depends on the observation time and the relative values of the initial and boundary concentration. In the example considered above we observe the minimum at about 110 m because the boundary condition doesn't influent solution yet and so it doesn't create a concentration gradient and consequently flux. Therefore, we have null uncertainty for the flux (i.e., the flux is null without uncertainty) and the same uncertainty of the initial condition for the concentration solution. After c.a. 140 m, for the same reason, the initial concentration and null boundary condition create a concentration gradient that give a flux.

In order to emphasize the above considerations, we observe the same quantities for a larger diffusion time. Figures 3-3 and 3-4 display mean and variance of concentration and flux profiles for an observation time of 5 million years.

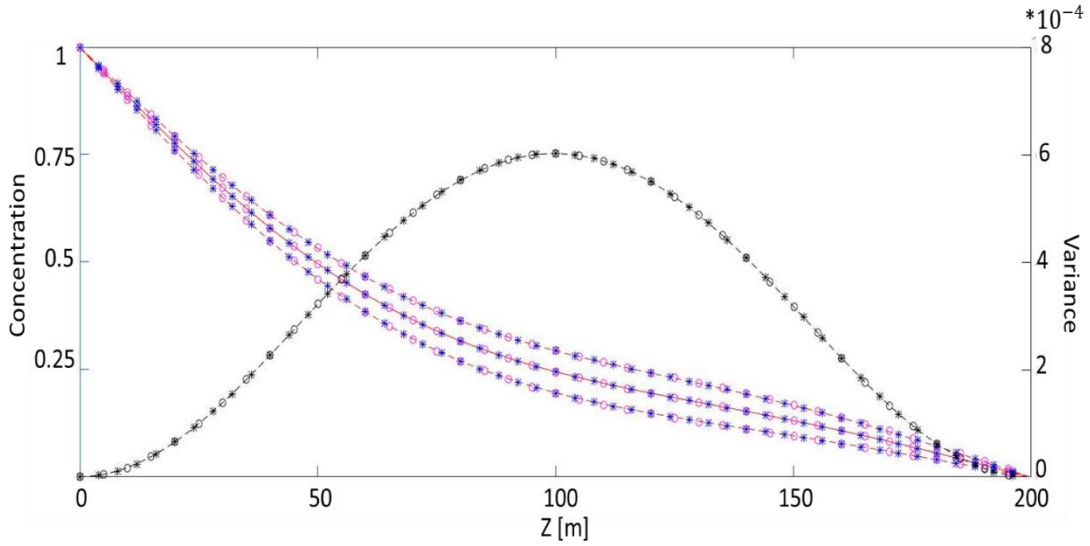


Fig 3-3 Comparison between the three concentration solutions, their relative band of uncertainty (±2σ) and the variance (in black) function of the spatial distant at a time t=5 million years with random uniform distributed initial condition. The pink points are the numerical solution, the blue stars are the analytical solution (Eq-3-14 and Eq.3-15) while the continue red line are the exact Monte Carlo solution (eq.2-18).

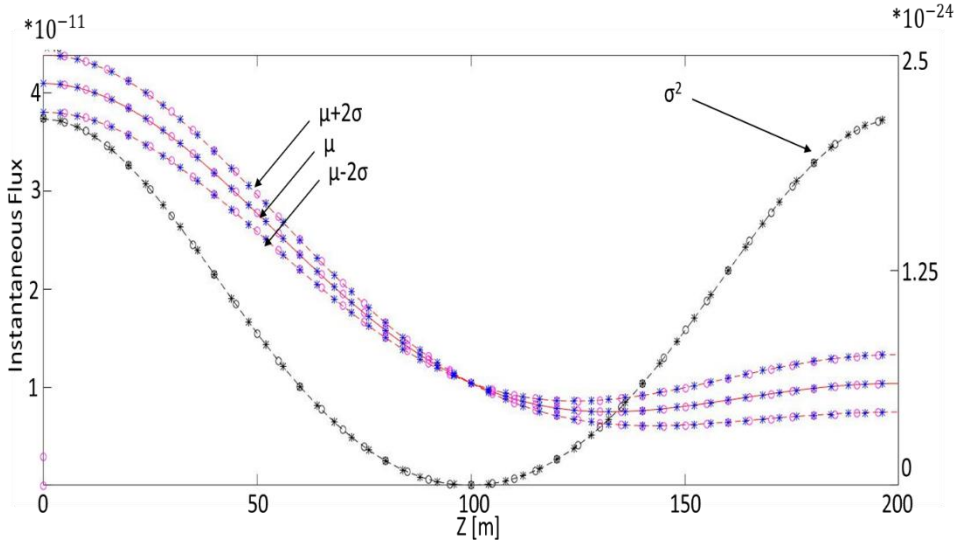


Fig 3-4 Comparison between the three instantaneous flux solutions, their relative band of uncertainty (±2σ) and the variance (in black) at time t= 5 million years in function of the spatial distance with random uniform distributed initial condition. The pink points are the numerical solution, the blue stars are the

analytical solution (Eq.3-17 and Eq.3-18) while the continue red line are the exact Monte Carlo solution (Eq.2-25).

Figure 3-4 show that at z= L/2 = 100 m we have again the minimum error flux solution and the minimum flux is shifted to the end of the domain. In fact, if we observed the analytical solution of flux variance (Eq. 3-18) we see that for, z=L/2:

$$VAR[q_N(z,t)] =$$

$$-D[\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{4}{L^2}e^{-(i^2+i^2)Dt\pi^2/L^2}\,cos\frac{i\pi}{2}cos\frac{j\pi}{2}\left[\frac{(b^3-a^3)}{3(b-a)}[(-1)^{2n}-(-1)^n+1]-\left(\left(\frac{b+a}{2}\right)-\left(\frac{b+a}{2}\right)(-1)^n\right)^2\right]=0 \quad . \qquad \textit{Eq. 3-21}$$

Figure 3-5 illustrates the variance of the gas concentration versus depth for diverse values of variation coefficient for initial condition. The red line is for a variation coefficient of 0.25, the blue line for 0.75 and light blue for 1.5. The maximum value of variance is always at z=L/2 as we expected our analytical solution.

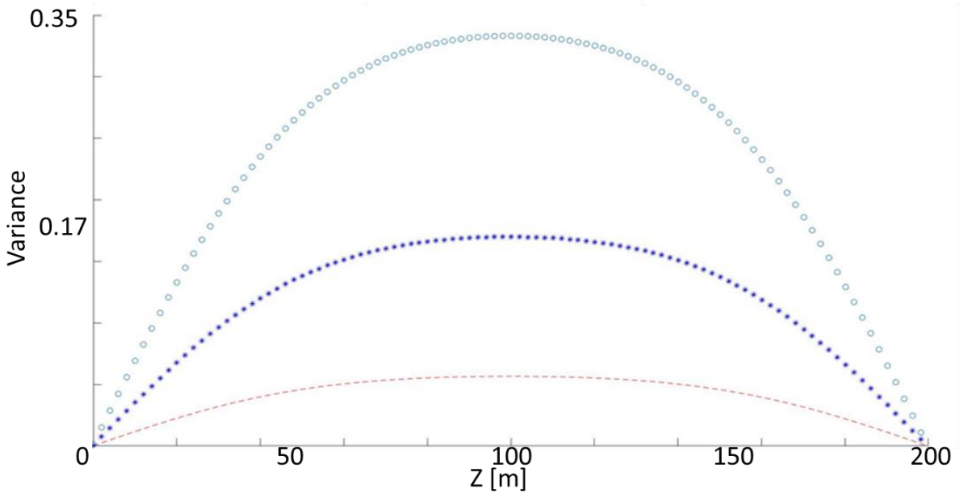


Fig 3-5 Variance profiles for different variation coefficients for initial condition. The red line is for coefvar=0.25, the blue line for 0.75 and light blue for 1.5.

In the follow graph, Fig 3-6, we can observe the profile concentration in function of time at a fixed position.

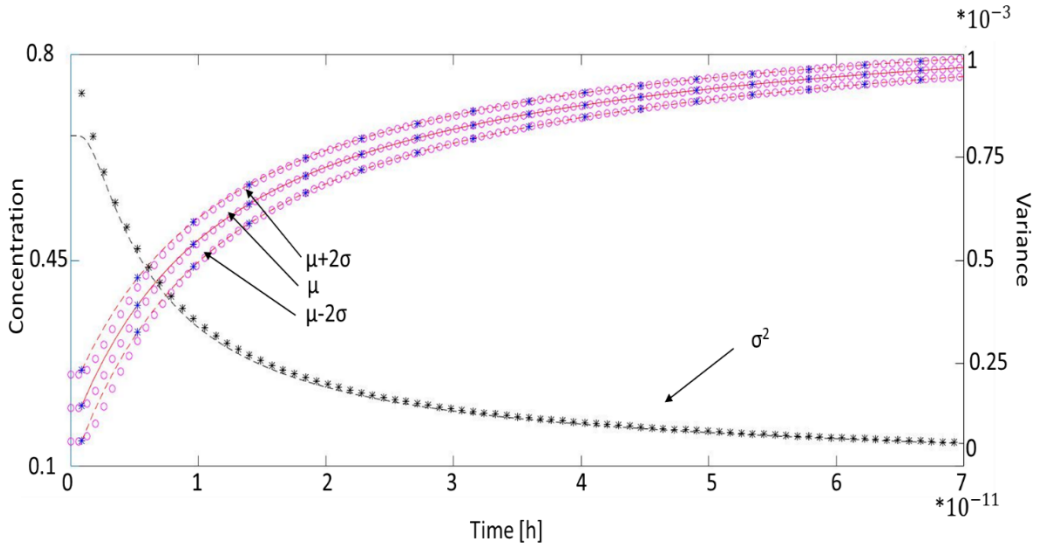


Fig 3-6 Comparison between the three concentration solutions, their relative band of uncertainty (±2σ) and of the variance (black graphs) at z=25 m for a maximum diffusion time of 8 million years with random uniform distributed initial condition. The pink points are the numerical solution, the blue stars are the analytical solution computed above while the red continue line is the exact Monte Carlo solution

As we can see the numerical Euler method solution begin exactly from the initial concentration while the others two not. This is an error that we found until $0.2*10^{-11}$ h and it is due at the truncation of the series in the analytical mean and in the deterministic solution; see Appendix D for details.

3.2 PROBLEM WITH STOCHASTIC DIFFUSION COEFFICIENT

Here we consider the problem where the boundary conditions and the initial condition are deterministic, while the diffusion coefficient is a second order random variable for each $x \in [0,L]$. The function g(D) is the density function of the random variable D defined in the domain Φ. This is the most important problem to really understand the caprock behavior. Possibility of being able to distinguish with a single effective diffusion coefficient the caprock and its

relative uncertainty is in fact the objective of many researches both in the experimental and modeling fields.

In the model we consider the case where:

1) D is a mean square continuous second-stochastic-problem [(Soong, 1973),pag.90]
2) $D' = D - E[D]$ is the diffusion coefficient fluctuation around the mean value $E[D]$

As we did before we split the problem for $c_0(z)$ deterministic and constant Dirichlet boundary condition into two new problem

the stochastic problem

$$\frac{\partial C(z,t)}{\partial t} = (D - E[D])\frac{\partial^2 C(z,t)}{\partial z^2}, \quad 0 < z < L, t > 0 \qquad \textit{Eq. 3-22}$$

with boundary and initial conditions given by

$$C(0,t) = c_t, \qquad t > 0$$

$$C(L,t) = c_b, \qquad t > 0$$

$$C(z,0) = c_0(z) \quad 0 < z < L$$

and the deterministic problem

$$\frac{\partial C(z,t)}{\partial t} = E[D] \ \frac{\partial^2 C(z,t)}{\partial z^2}, \quad 0 < z < L, t > 0 \qquad \textit{Eq. 3-23}$$

with boundary and initial conditions:

$$C(0,t) = c_t, \qquad t > 0$$

$$C(L,t) = c_b, \qquad t > 0$$

$$C(z,0) = c_0(z) \ \ 0 < z < L.$$

As in (J.C.Cortes, 2005) it is easily proved that the stochastic process $C(z,t)$ (Eq. 2-18) is an exact solution process for the problem in exam (Appendix B). Then, k_n is a determinist series and u_n the stochastic part, so using for k_n the mean square random variable property we can compute the mean and the variance of the concentration (J.C.Cortes, 2005) as well as the mean and the variance of the instantaneous flux.

If *D* is uniformly distributed on the interval [a,b] and it doesn't depend on z we obtain for the mean value:

$$E[C_N(z,t)] = \omega(z,t) + \sum_{n=1}^{N} k_n E[u_n(z,t)] \quad = \omega(z,t) +$$

$$\sum_{n=1}^{N} k_n \sin\frac{n\pi z}{L} E\left[e^{-\frac{n^2 Dt\pi^2}{L^2}}\right] = c_t + (c_b - c_t)\frac{z}{L} +$$

$$\sum_{n=1}^{N} \frac{2}{L}\int_0^L (c_0(z) - c_t - (c_b - c_t)\frac{z}{L}) \sin\frac{n\pi z'}{L} dz' \sin\frac{n\pi z}{L} \int_a^b e^{-\frac{n^2 Dt\pi^2}{L^2}} \frac{1}{b-a} dD$$

$$= c_t + (c_b - c_t)\frac{z}{L} +$$

$$\sum_{n=1}^{N} \frac{2}{L}\int_0^L (c_0 - c_t - (c_b - c_t)\frac{z}{L}) \sin\frac{n\pi z'}{L} dz' \sin\frac{n\pi z}{L} \int_a^b e^{-\frac{n^2 Dt\pi^2}{L^2}} \frac{1}{b-a} dD$$

$$= c_t + (c_b - c_t)\frac{z}{L} - \sum_{n=1}^{N} \frac{2}{n\pi}[c_o - c_t + (-1)^n (c_b -$$

$$c_o)] \sin(\frac{n\pi z}{L}) \frac{1}{(b-a)} \left[\frac{L^2}{n^2 t\pi^2}\left(e^{-n^2 at\pi^2/L^2} - e^{-n^2 bt\pi^2/L^2}\right)\right] \qquad \textit{Eq. 3-24}$$

and for the variance:

$$VAR[C_N(z,t)] = VAR[\omega(z,t)] +$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} k_i k_j \sin\frac{i\pi z}{L} \sin\frac{j\pi z}{L} \left[\int_\Phi e^{-(i^2+i^2)Dt\pi^2/L^2} g(D) dD - \right.$$

$$\left.\left(\int_\Phi e^{-(i^2)Dt\pi^2/L^2} g(D) dD\right)\left(\int_\Phi e^{-(j^2)Dt\pi^2/L^2} g(D) dD\right)\right] =$$

$$VAR[\omega(z,t)] + \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{2}{i\pi}\left[c_o - c_t + (-1)^i (c_b - c_o)\right] \frac{2}{j\pi}\left[c_o - \right.$$

$$\left. c_t + (-1)^j (c_b - \right.$$

$$c_o)\Big] sin\frac{i\pi z}{L} sin\frac{j\pi z}{L}\left[\frac{1}{(b-a)}\left[\frac{L^2}{(i^2+j^2)t\pi^2}\left(e^{-(i^2+j^2)at\pi^2/L^2} - e^{-(i^2+j^2)bt\pi^2/L^2}\right)\right] - \frac{1}{(b-a)}\left[\frac{L^2}{i^2t\pi^2}\left(e^{-i^2at\pi^2/L^2} - e^{-i^2bt\pi^2/L^2}\right)\right]\frac{1}{(b-a)}\left[\frac{L^2}{j^2t\pi^2}\left(e^{-j^2at\pi^2/L^2} - e^{-j^2bt\pi^2/L^2}\right)\right]\right] \quad \textit{Eq. 3-25}$$

Successively we compute the mean and the variance considering a random diffusion coefficient for the instantaneous flux which read respectively:

$$E[q_N(z,t)] = E[D]\left[(c_b - c_t)\frac{1}{L} - \sum_{n=1}^{\infty}\frac{L}{\pi n}k_n\, cos(z\lambda)\, e^{-\lambda^2 E[D]t}\right] = \frac{b+a}{2}\left[(c_b - c_t)\frac{1}{L} - \sum_{n=1}^{\infty}\frac{L}{\pi n}k_n\, cos(z\lambda)\, e^{-\lambda^2\frac{b+a}{2}t}\right] \quad \textit{Eq. 3-26}$$

and

$$VAR[q_N(z,t)] = \sum_{i=1}^{N}\sum_{j=1}^{N}k_i k_j\, cos\frac{i\pi z}{L} cos\frac{j\pi z}{L}\left[\left(E\left[e^{-(i^2+i^2)Dt\pi^2/L^2}D^2\right]\right) - E\left[e^{-\frac{i^2Dt\pi^2}{L^2}}D^2\right]E\left[e^{-\frac{j^2Dt\pi^2}{L^2}}D^2\right]\right] \quad \textit{Eq. 3-27}$$

where

$$E\left[e^{-(i^2+i^2)Dt\pi^2/L^2}D^2\right] = \frac{1}{b-a}\left[\left(-\frac{e^{-(i^2+i^2)bt\pi^2/L^2}}{-(i^2+i^2)t\pi^2/L^2}b^2 + \frac{e^{-(i^2+i^2)at\pi^2/L^2}}{-(i^2+i^2)t\pi^2/L^2}a^2\right) - \left(\frac{e^{-(i^2+i^2)bt\pi^2/L^2}}{(-(i^2+i^2)t\pi^2/L^2)^2}2b - \frac{e^{-(i^2+i^2)at\pi^2/L^2}}{(-(i^2+i^2)t\pi^2/L^2)^2}2a\right) + \left(-\frac{e^{-(i^2+i^2)bt\pi^2/L^2}}{(-(i^2+i^2)t\pi^2/L^2)^3}2 + \frac{e^{-(i^2+i^2)at\pi^2/L^2}}{(-(i^2+i^2)t\pi^2/L^2)^3}2\right)\right] \quad \textit{Eq. 3-28}$$

and

$$E\left[e^{-i^2Dt\pi^2/L^2}D^2\right] = \frac{1}{b-a}\left[\left(\frac{e^{-\frac{i^2at\pi^2}{L^2}}}{\frac{i^2t\pi^2}{L^2}}a - \frac{e^{-\frac{i^2bt\pi^2}{L^2}}}{\frac{i^2t\pi^2}{L^2}}b\right) - \left(\frac{e^{-\frac{i^2bt\pi^2}{L^2}}}{\left(\frac{i^2t\pi^2}{L^2}\right)^2} - \frac{e^{-\frac{i^2at\pi^2}{L^2}}}{\left(\frac{i^2t\pi^2}{L^2}\right)^2}\right)\right] \quad . \quad \textit{Eq. 3-29}$$

3.2.1 Results

Here we illustrate some results obtained by taking a variable D uniformly distributed in the interval of extremes a and b, by considering its mean value μ and the variation coefficient

$$D = [a, b] = [\mu - coefvar \cdot \mu\,, \mu + coefvar \cdot \mu\,]\ . \qquad \textit{Eq. 3-30}$$

The following figures show the concentration and the instantaneous flux plotting the exact solution (computed by Montecarlo realizations of the analytical solution Eq.2-22), the analytical solution given in Eq.3-24 and Eq.3-25 and the numerical solution (computed by the Euler Finite Difference Method) and their relative band of uncertainty (±2σ). We consider the case of $c_o = 0$ and $c_t = 1$ for a time t=2.5 million years and a coefficient of variation for the uniform distribution of D of 0.25. The diffusion coefficient is considered a constant on the space domain with mean equal to μ =10-12 m^2/s that is a typical value for the shale rocks.

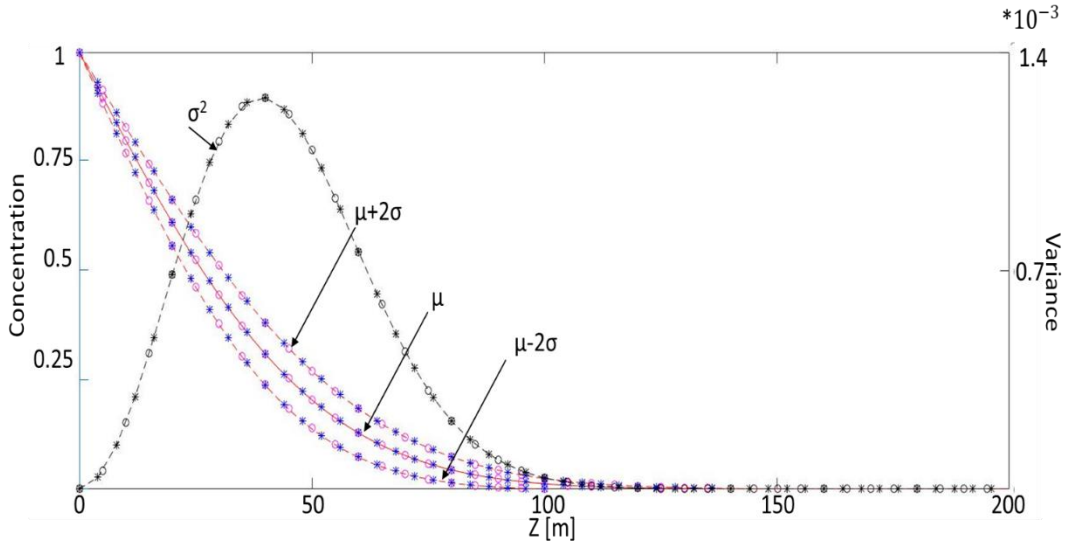


Fig 3-7 Comparison between the three concentration solutions, their relative band of uncertainty (±2σ) and of the variance (black graphs) for 2.5 million years in all space domain with random uniform distributed diffusion coefficient. The pink points are the numerical solution, the blue stars are the analytical solution computed above (Eq.3-24 and Eq.3-25) while the continue red line is the exact Monte Carlo solution.

Figure 3-7 show that the three solutions (the analytical, the Montecarlo and the numerical) coincide.

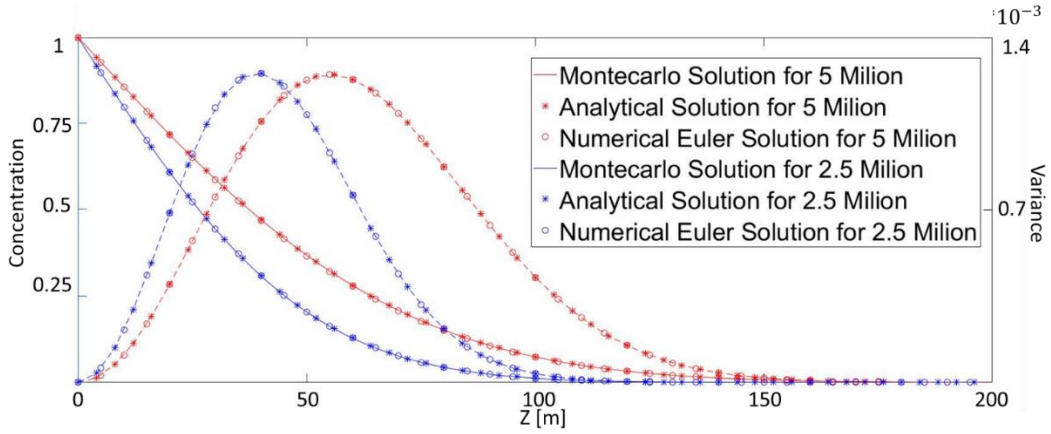


Fig 3-8 Comparison between the solution concentration profile and the respectively variances for 2.5 and 5 million years in all space domain with random uniform distributed diffusion coefficient.

Figure 3-8 compare the bell for a longer time (5 million years) with the previous one. We can observe that the maximum variance has the same value, while the position is shifted due to the larger diffusion space.

Figure 3-9 illustrate the variance of the solution with diverse coefficient of variation. Note that the maximum coefficient of variation taken is 1 because, according with Equation (8-27) a larger value could imply negative D values that is physically impossible. Figure 3-9 also shows that the peak of bell translates in z direction as the coefficient of variation increases.

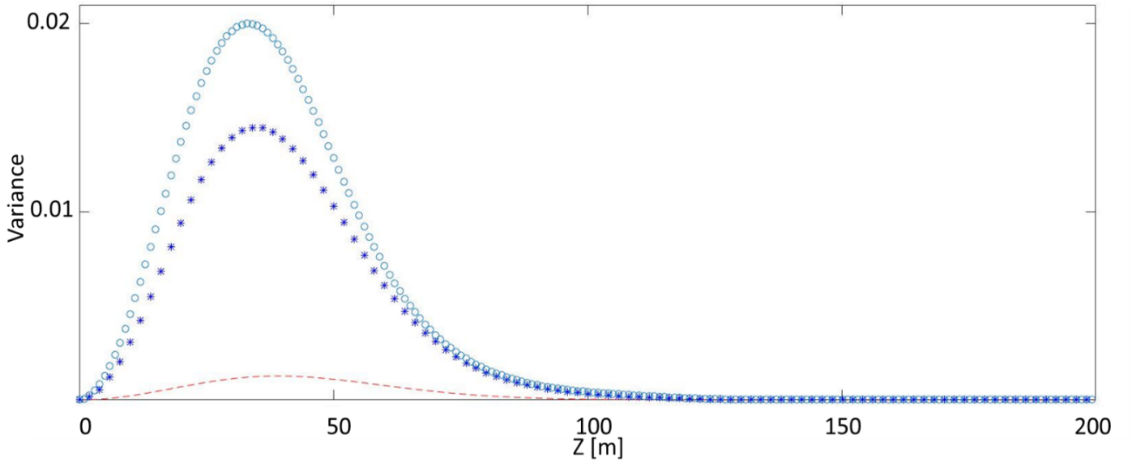


Fig 3-9 Variance profiles for different variation coefficients. The red line is for coefvar=0.25, the blue line for 0.75 and light blue for 1. Values over the unity for the coefvar are not allowed because they could bring negative diffusion coefficient that they aren't physically possible

Figure 3-10 shows the flux along the caprock at a certain time (2,5 million years). As we can see now the variance of the flux solution has a non-zero value at z=0, a maximum at the same z point of the maximum variance of the concentration and as expected 0 variance after the maximum diffusion point.

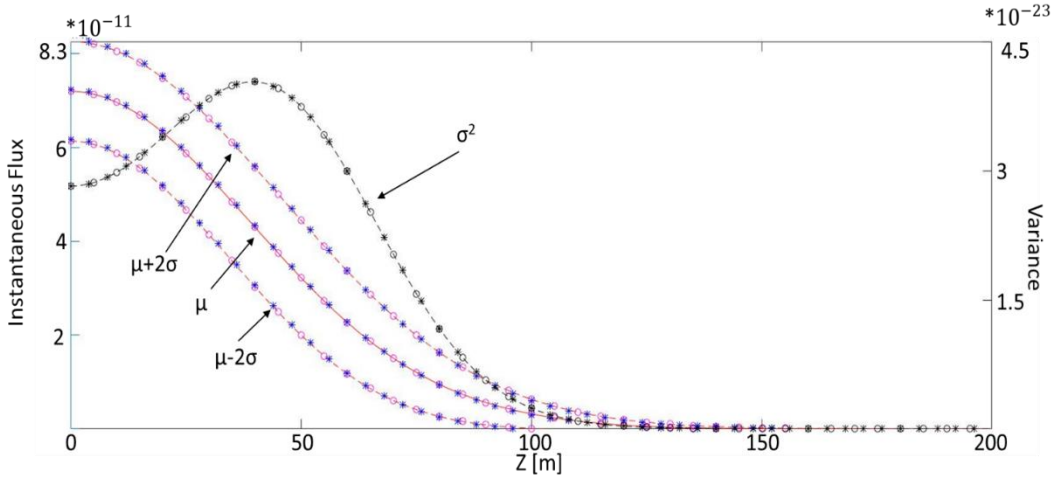


Fig 3-10 Comparison between the three instantaneous flux solutions, their relative band of uncertainty (±2σ) and of the variance (black graphs) for 2.5 million years in all space domain with random uniform distributed diffusion coefficient. The pink points are the numerical solution, the blue stars are the analytical solution computed above (Eq.3-26 and Eq.3-27) while the continue red line is the exact Monte Carlo solution.

As before, we finally consider the change of the concentration and of the flux at fixed points in function of time. In Figure 3-11 and Figure 3-12 we observe concentration and instantaneous flux at z=25m at t=8 million years. We can see that the concentration trend increases until it tends to reach an asymptote. Instead, the instantaneous flux, after an initially increase, it decreases and so it tends to stabilize.

We can see that initially the both variances increase. After a certain time, different between the concentration and the flux, it decreases and it tends to stabilize.

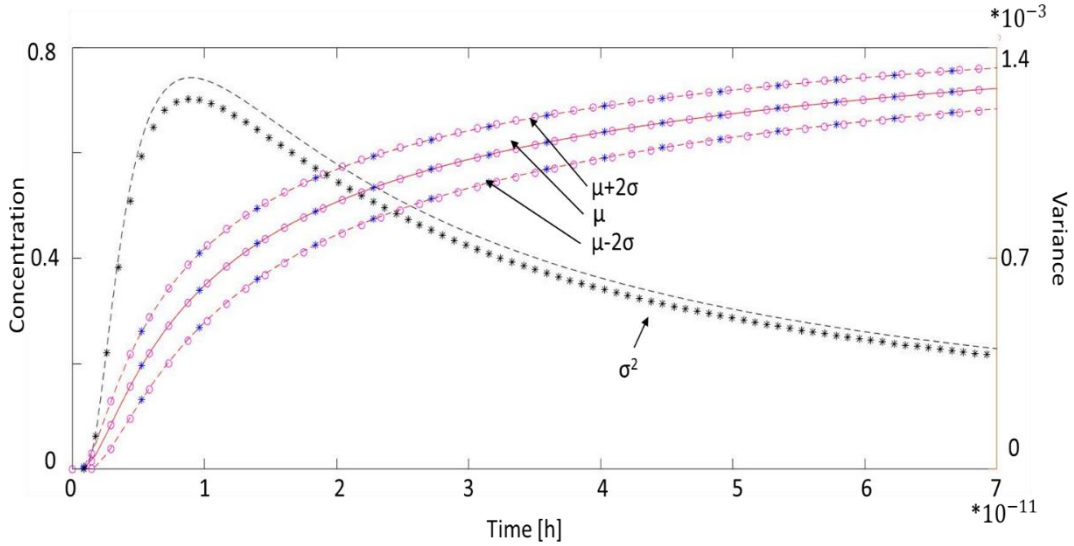


Fig 3-11 Comparison between the three concentration solutions, their relative band of uncertainty (±2σ) and of the variance (in black) at z=25 m for a diffusion time of 8 million years with random uniform distributed diffusion coefficient. The pink points are the numerical solutions, the blue stars are the analytical solutions Eq.3-24 and Eq.3-25 while the continue line is the exact Monte Carlo solution.

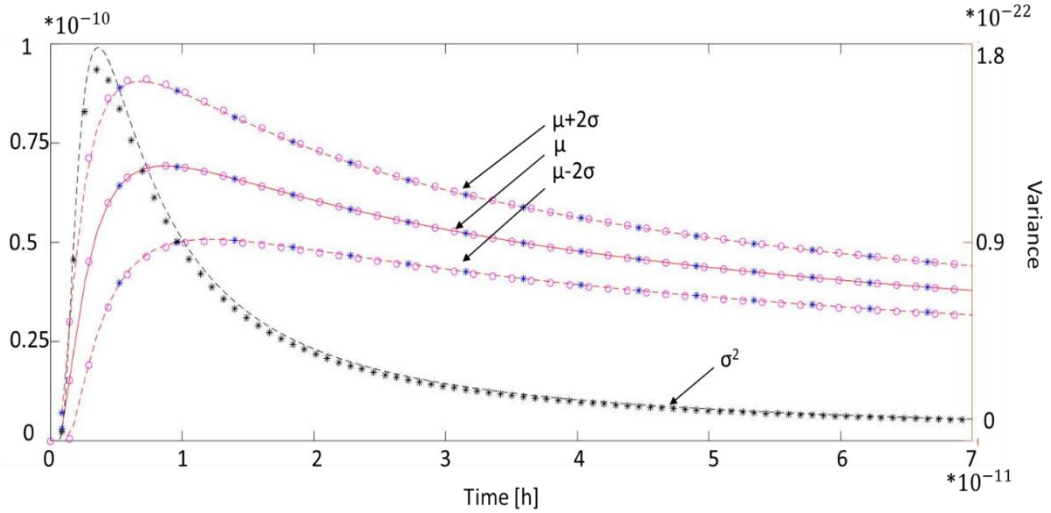


Fig 3-12 Comparison between the three instantaneous flux solutions, their relative band of uncertainty (±2σ) and of the variance (in black) at z=25 m for a diffusion time of 8 million years with random uniform distributed diffusion coefficient. The pink points are the numerical solutions, the blue stars are the analytical solutions Eq. 3-26 and Eq.3-27 while the continue red line is the exact Monte Carlo solution.

Figure 3-13 and Figure 3-14 illustrate the concentration and the instantaneous flux for z=175m. We can see that both the solutions increase. The variances display the same behaviors. This different behavior respect the result for z=25m is due to the much bigger distance from the not null boundary condition.

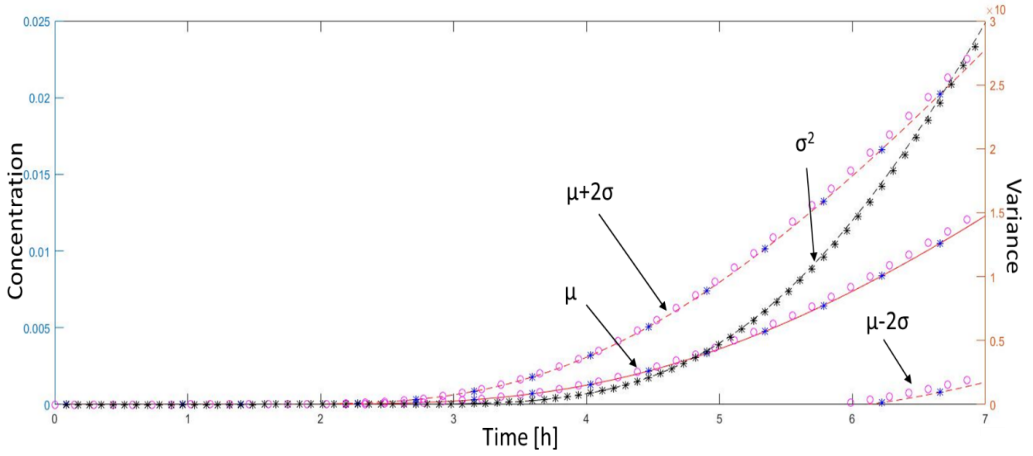


Fig 3-13 Comparison between the three concentration solution, their relative band of uncertainty (±2σ) and of the variance (in black) at z=175 m for a diffusion time of 8 million years with random uniform distributed diffusion coefficient. The points are the numerical solutions, the stars are the analytical solutions (Eq.3-24 and Eq.3-25) while the continue line is the exact Monte Carlo solution.

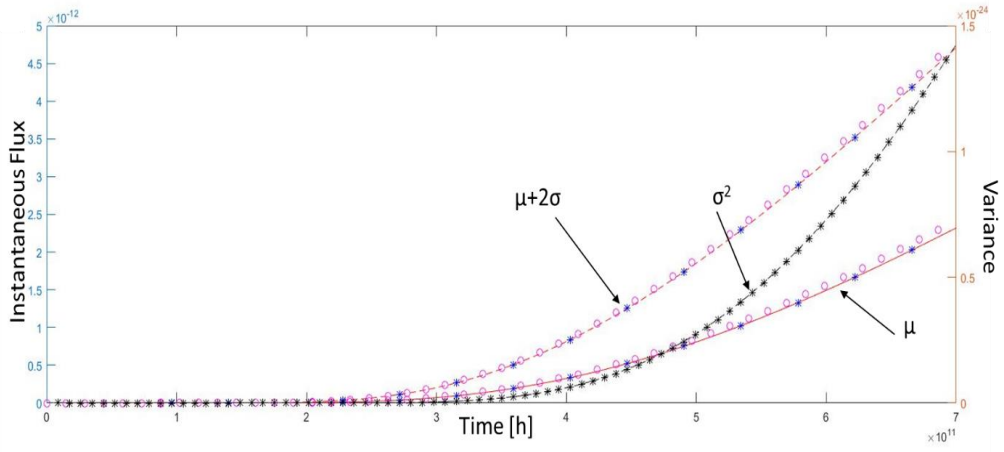


Fig 3-14 Comparison between the three instantaneous flux solutions and of the variance (black graphs) at z=175 m for a maximum diffusion time of 8 million years with random uniform distributed diffusion coefficient. The points are the numerical solution, the stars are the analytical solution computed above (Eq.3-26 and Eq.3-27) while the continue line is the exact Monte Carlo solution.

These considerations are confirmed in Figure 3-15 where is shown the analytical solution of the concentration for a larger time.

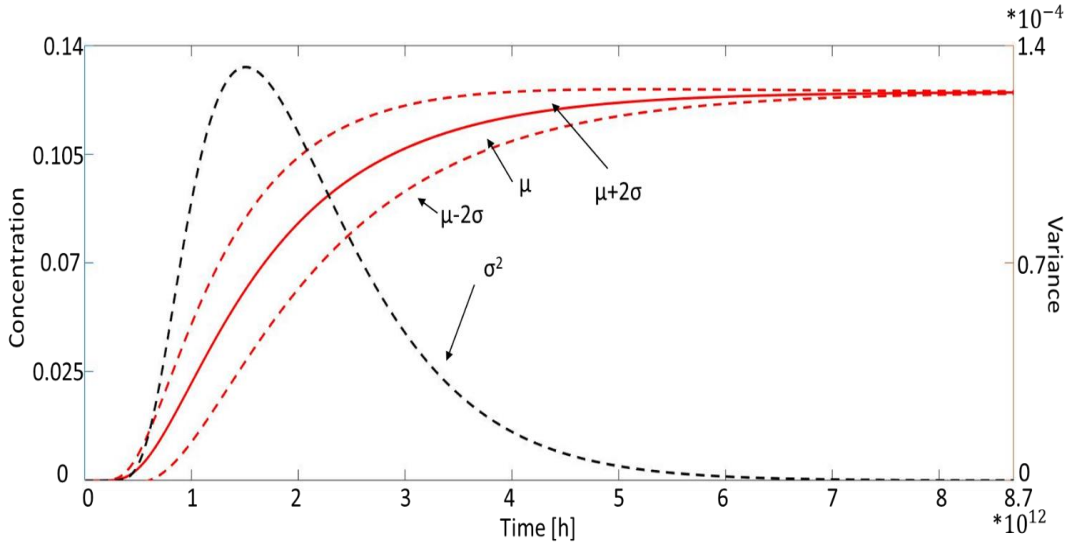


Fig 3-15 Analytical concentration solution and variance (black graph) for 100 million years diffusion time with stochastic diffusion coefficient at z=175m.

We can note that for infinite time, at z=175, the concentration tends to a constant deterministic value. This limit is predictable from the analytical solution (Eq.3-23):

$$\lim_{t\to\infty} E[C(175,t)] = \lim_{t\to\infty}(c_t + (c_b - c_t)\frac{z}{L} - \sum_{n=1}^{N}\frac{2}{n\pi}[c_o - c_t + (-1)^n(c_b - c_o)]\, sin(\frac{n\pi z}{L})\frac{1}{(b-a)}\left[\frac{L^2}{n^2 t\pi^2}\left(e^{-n^2 at\pi^2/L^2} - e^{-n^2 bt\pi^2/L^2}\right)\right] = c_t - (c_b - c_t)\frac{z}{L} \quad . \qquad \textit{Eq. 3-31}$$

Moreover we can observe that the uncertainly in the solution due to the stochastic diffusion coefficient, after certain time, vanishes. This is a key information, because the diffusion coefficient is commonly uncertain in real problems. Further investigations should be done for diverse distributions of heterogeneity that could bring different behaviors.

3.3 PROBLEM WITH STOCHASTIC BOUNDARY CONDITION

Here we consider the case where the boundary condition at z=0, c_t, is a mean square random variable and consequently $\omega(z)$ is a mean square random variable.

This is the case when, in the reservoir model, the value of the possible past concentration inside the reservoir is considered constant in time and affected by an uncertainty.

If we consider the case where:

1) $E[\omega(z)] = \bar{c_t} + (c_b - \bar{c_t})\frac{z}{L}$
2) $\omega(z)$ is a mean square continuos second-stochastic-problem [(Soong, 1973),pag.90]
3) $\omega(z)' = \omega(z) - E[\omega(z)] = c_t' - c_t'\frac{z}{L}$

We can see that only k_n terms are stochastic, while u_n are a deterministic function. Analogously to the previous case we compute the mean value of the concentration

$$E[C_N(z,t)] = \overline{\omega(z,t)} +$$

$$\sum_{n=1}^{N} sin\frac{n\pi z}{L} e^{-n^2 Dt\pi^2/L^2} \left[\frac{2}{L}\int_0^L c_0(z) sin\frac{n\pi z'}{L} dz' -\right.$$

$$\left.\frac{2}{L}\int_0^L \overline{\omega(z,t)} sin\frac{n\pi z'}{L} dz'\right] = \frac{(b+a)}{2} + \frac{c_b - \frac{(b+a)}{2}}{L} * z + \sum_{n=1}^{\infty}\frac{2}{L}\int_0^L \left(c_o(z) -\right.$$

$$\left. c_b\frac{z'}{L} - \frac{(b+a)}{2}\frac{(L-z')}{L}\right) sin(z'\lambda) dz' \, sin(z\lambda)\, e^{-\lambda^2 Dt} \qquad \textit{Eq. 3-32}$$

and the variance:

$$VAR[C_N(z,t)] = E\left[c_t' - c_t'\frac{z}{L} - \sum_{n=1}^{N} c_t'\frac{2}{n\pi} sin\frac{n\pi z}{L} e^{-n^2 Dt\pi^2/L^2}\right]^2 =$$

$$\frac{(b-a)^2}{12} + \frac{(b-a)^2}{12}\left(\frac{z}{L}\right)^2 + \sum_{n=1}^{N}\frac{(b-a)^2}{12}\left(\frac{2}{n\pi} sin\frac{n\pi z}{L} e^{-\frac{n^2 Dt\pi^2}{L^2}}\right)^2 +$$

$$2\sum_{n=1}^{N}\sum_{i=1}^{n-1}\left[\frac{(b-a)^2}{12}\frac{2}{n\pi} sin\frac{n\pi z}{L} e^{-\frac{n^2 Dt\pi^2}{L^2}} \frac{2}{i\pi} sin\frac{i\pi z}{L} e^{-\frac{i^2 Dt\pi^2}{L^2}}\right] - 2\frac{(b-a)^2}{12}\frac{z}{L} +$$

$$2\frac{z}{L}\sum_{n=1}^{N}\frac{(b-a)^2}{12}\frac{2}{n\pi} sin\frac{n\pi z}{L} e^{-\frac{n^2 Dt\pi^2}{L^2}} -$$

$$2\sum_{n=1}^{N}\frac{(b-a)^2}{12}\frac{2}{n\pi} sin\frac{n\pi z}{L} e^{-\frac{n^2 Dt\pi^2}{L^2}} \qquad \textit{Eq. 3-33}$$

For the instantaneous flux we obtain:

$$E[q_N(z,t)] = D\left[(c_b - E[c_t])\frac{1}{L} - \sum_{n=1}^{\infty}\frac{L}{\pi n} k_n\, cos(z\lambda)\, e^{-\lambda^2 Dt}\right] \textit{Eq. 3-34}$$

and

$$VAR[q_N(z,t)] = E[\frac{D^2(c_t')^2}{L^2} + D^2\sum_{n=1}^{N}\left(c_t'\frac{2L}{n^2\pi^2} cos\frac{n\pi z}{L} e^{-\frac{n^2 Dt\pi^2}{L^2}}\right)^2 +$$

$$2\sum_{n=1}^{N}\sum_{i=1}^{n-1}\left[(c_t')^2\frac{2L}{n^2\pi^2} cos\frac{n\pi z}{L} e^{-\frac{n^2 Dt\pi^2}{L^2}} \frac{2L}{i^2\pi^2} cos\frac{i\pi z}{L} e^{-\frac{i^2 Dt\pi^2}{L^2}}\right] +$$

$$\frac{D^2}{L}\sum_{n=1}^{N}\left((c_t')^2\frac{2L}{n^2\pi^2} cos\frac{n\pi z}{L} e^{-\frac{n^2 Dt\pi^2}{L^2}}\right)] = \frac{D^2\frac{(b-a)^2}{12}}{L^2} +$$

$$D^2\sum_{n=1}^{N}\frac{(b-a)^2}{12}\left(\frac{2L}{n^2\pi^2} cos\frac{n\pi z}{L} e^{-\frac{n^2 Dt\pi^2}{L^2}}\right)^2 +$$

$$2\sum_{n=1}^{N}\sum_{i=1}^{n-1}\left[\frac{(b-a)^2}{12}\frac{2L}{n^2\pi^2}cos\frac{n\pi z}{L}e^{-\frac{n^2Dt\pi^2}{L^2}}\frac{2L}{i^2\pi^2}cos\frac{i\pi z}{L}e^{-\frac{i^2Dt\pi^2}{L^2}}\right]+$$

$$\frac{D^2}{L}\sum_{n=1}^{N}\left(\frac{(b-a)^2}{12}\frac{2L}{n^2\pi^2}cos\frac{n\pi z}{L}e^{-\frac{n^2Dt\pi^2}{L^2}}\right) \qquad \textit{Eq. 3-35}$$

3.3.1 Results

As previously the random value of c_t is taken from a uniform distribution of extremes a and b, that we compute starting from its mean value μ and the coefficient of variation:

$$c_t = [a,b] = [\mu - coefvar\cdot\mu\,,\mu + coefvar\cdot\mu\,]\ . \qquad \textit{Eq. 3-36}$$

We plot the exact solution (computed by Montecarlo realizations of (Eq 2.17)), the analytical solution obtained given in Eq.3-32 and Eq.3-33 and the numerical solution computed by the Euler Finite Difference Method (see Appendix C) of concentration and of instantaneous flux and their relative band of uncertainty (±2σ). In Figure 3-16 and Figure 3-17 we consider the case of $c_o = 0$ and the mean of c_t: $\mu = 1$ at time t= 2.5 million years and a coefficient of variation for the uniform distribution of $c_t = [0.75 - 1.25]$ of 0.25. The diffusion coefficient is considered a constant on the domain and equal to D =10^{-12} m^2/s that is a typical value for the shale rocks.

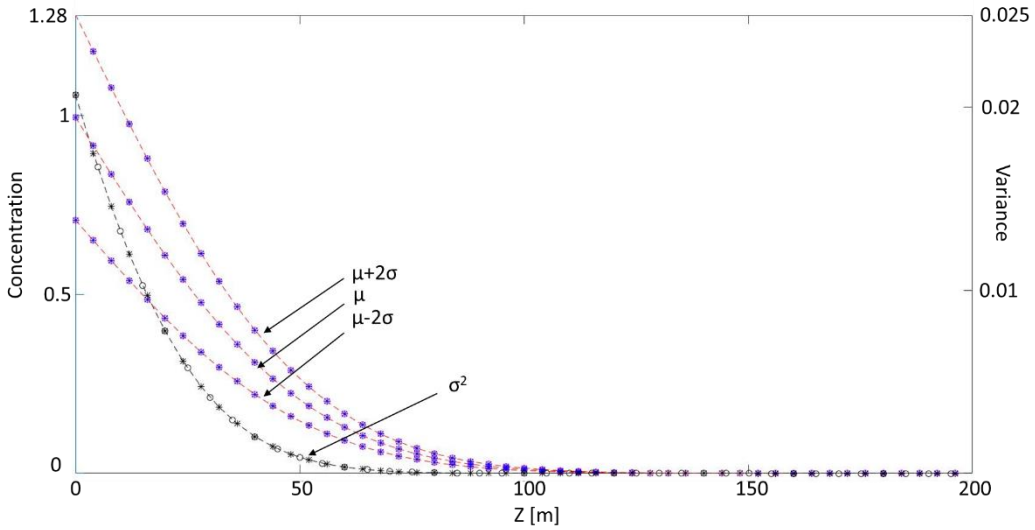


Fig 3-16 Comparison between the three concentration solutions, their relative band of uncertainty (±2σ) and of the variance (in black) for t=2.5 million in function of z, taking into account a random uniform distributed boundary condition in z=0. The pink points are the numerical solutions, the blue stars are the analytical solutions given in Eq.3-32 and Eq. 3-33 while the continue red line is the exact Monte Carlo solution.

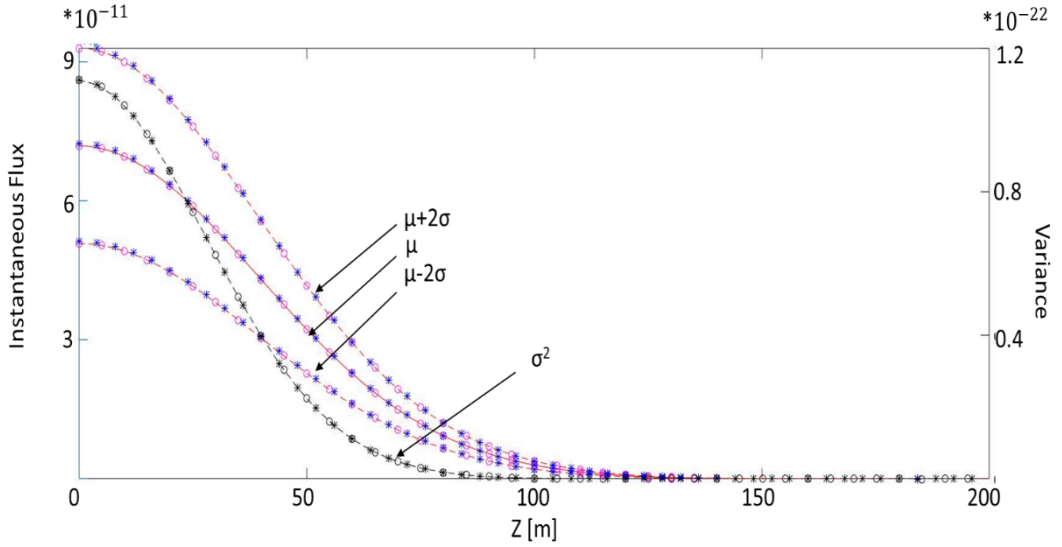


Fig 3-17 Comparison between the three instantaneous flux solutions, their relative band of uncertainty (±2σ) and of the variance (black graphs) for 2.5 million years in all space domain with random uniform distributed boundary condition in z=0. The pink points are the numerical solution, the blue stars are the analytical solution computed above Eq.3-34 and Eq.3-35 while the continue red line is the exact Monte Carlo solution.

Figures 3-16 and 3-17 show that the variances of the concentration and of the instantaneous flux, have an exponential behavior with a maximum is for z=0, where the stochastic variable is defined.

Figure 3-18 shows the variance profiles for different variation coefficients.

We observe that increasing the coefficient of variation increase the variance values without changing of shape.

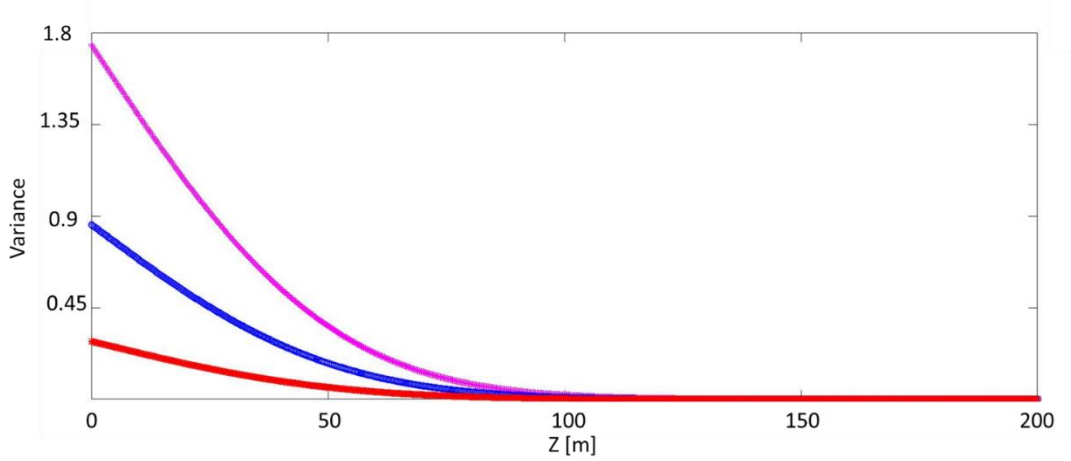


Fig 3-18 Variance profiles for different variation coefficients. The red line is for coefvar=0.25, the blue line for 0.75 and magenta for 1.

Figures 3-19 and 3-20 show the concentration solution for z=25m and z=175 m and for a time t= 8 million years.

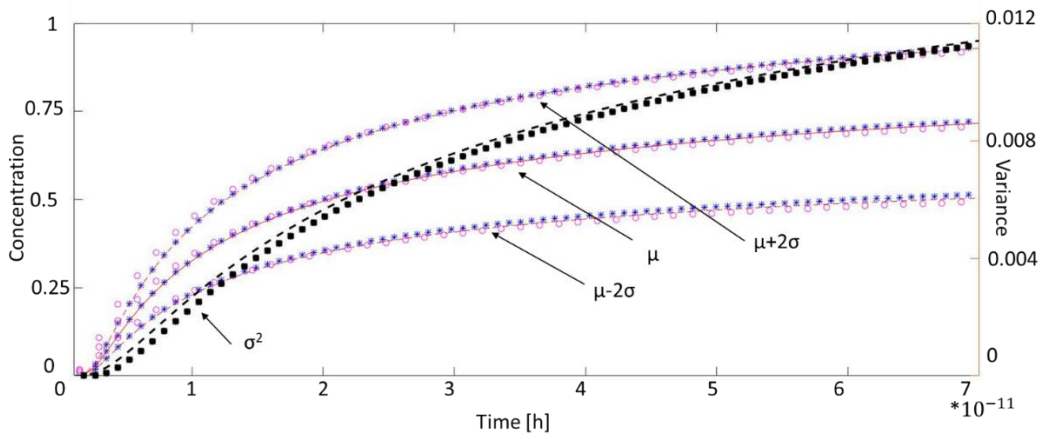


Fig 3-19 Comparison between the three concentration solutions (in blue pink and red) and of the variance (in black) at z=25 m for a diffusion time t=8 million years with random uniform distributed boundary condition at z=0. The points are the numerical solutions, the stars are the analytical solutions Eq.3-32 and Eq.3-33 .while the continue line is the exact Monte Carlo solution.

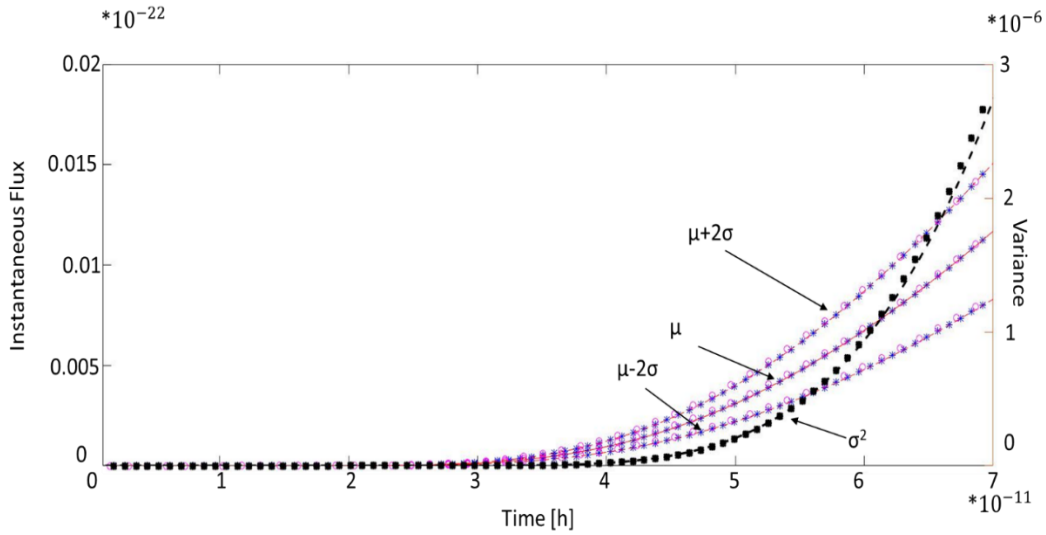


Fig 3-20 Comparison between the three instantaneous flux solutions (in blue pink and red) and of the variance (in black) at z=175 m for a diffusion time t= 8 million years with random uniform distributed diffusion coefficient. The points are the numerical solution, the stars are the analytical solutions Eq.3-34 and Eq.3-35 while the continue line is the exact Monte Carlo solution.

We can observe that the variance increases monotonously in time as predicted by the Eq. 3-35. To confirm this trends for longer times we observe the concentration and its variance for time t=20 million years. Figure 3-21 displays the Montecarlo solution and the analytical solution.

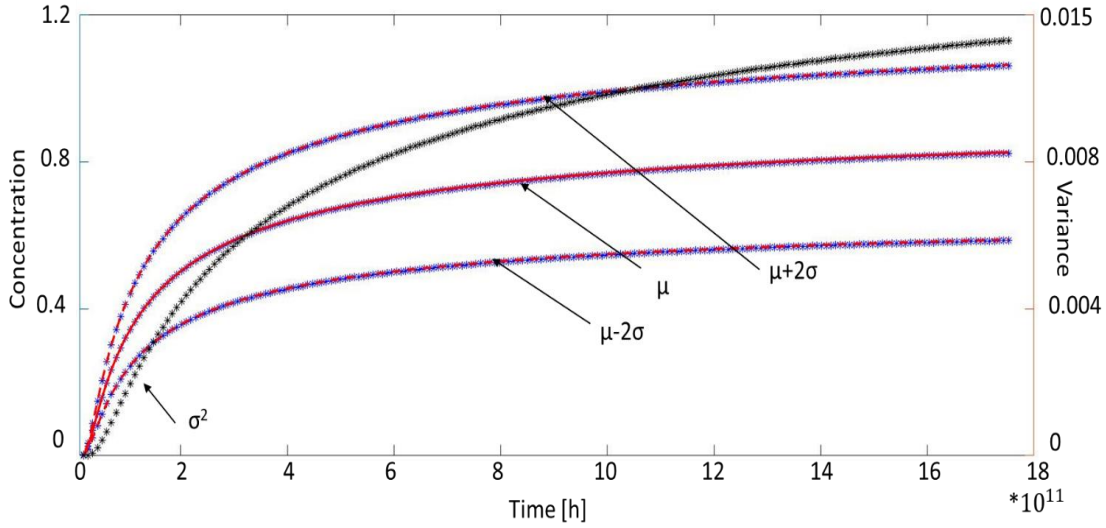


Fig 3-21 Comparison between the three concentration solutions (in red and blue) and of the variance (in black) at z=25 m for diffusion time t=20 million years with random uniform distributed boundary condition

The stars are the analytical solutions (Eq.3-32 and Eq.3-33) while the continue line is the exact Monte Carlo solution.

3.4 COMPARISON OF THE VARIANCE RESULTS FOR SINGLE STOCHASTIC VARIABLE

In the following we compare the results of three above problems: stochastic boundary condition, stochastic initial condition and stochastic diffusion coefficient. Firstly, we compare the temporal evolution of the variances, displayed in Figure 3-22.

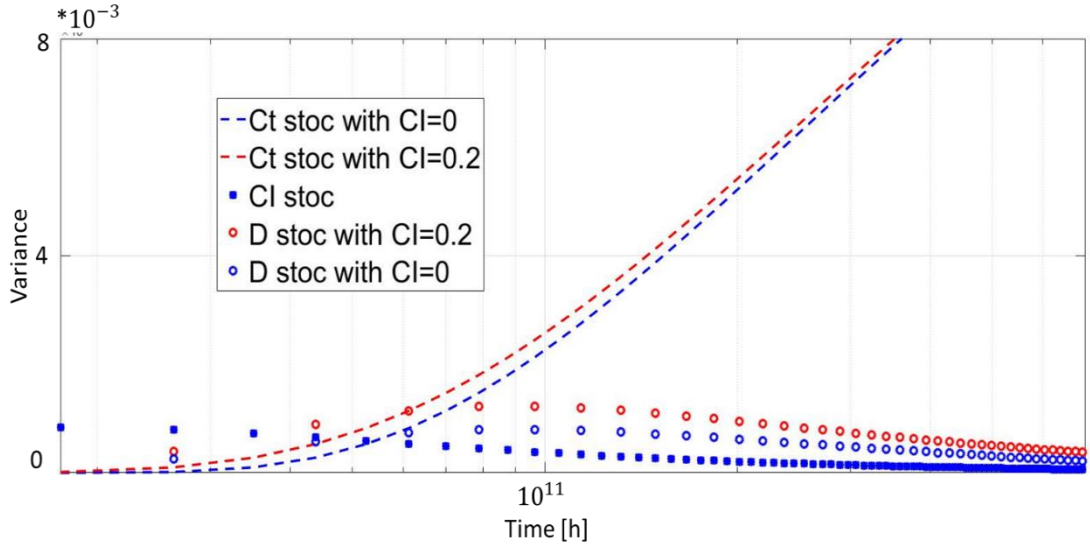


Fig 3-22 Comparison between the variance of concentration in time (in log scale) at z=25 m and maximum time diffusion of 30 millions years for the three cases above. There are also in red the variances when the initial condition is not null.

Figure 3-22 shows that the variance for Ct stochastic is monotonously growing, while in the other two cases, even if we have an initial increment for the D stochastic problem, after a given time it tends to decrease. If we observe the same figure, without the stochastic initial condition that have a similar behaviour, but for the variance of the instantaneous flux in Figure 3-23:

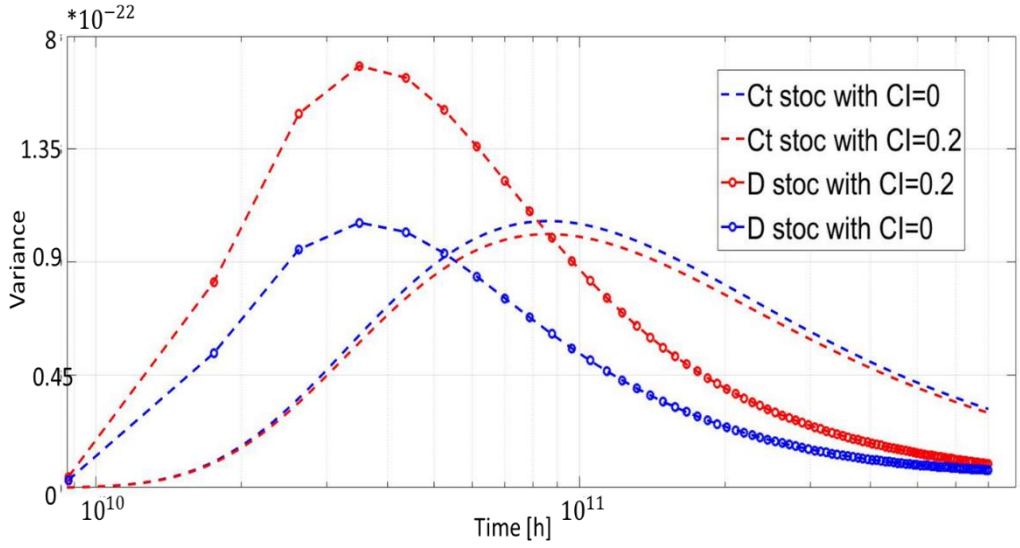


Fig 3-23 Comparison between the variance of the instantaneous flux in time (in log scale) at z=25 m and diffusion time t= 8 million years for the two cases whit Ct and D stochastic. There are also in red the variances when the initial condition is not null.

We can observe that the behaviors are very different for the solution with the boundary condition stochastic. In time while the concentration's variance of this problem increases the instantaneous flux's variance, after a peak, decrease. The presence, of homogeneous or not, initial condition doesn't change the shape in both cases.

3.5 Diffusion Problem taking boundary condition, Diffusion Coefficient and Initial Condition as stochastic independent variables

The last problem that we analyzed for the homogeneous case, is when all the three parameters: boundary condition, the initial condition and the diffusion coefficient are independent second order random variable for each z ϵ [0,L]. These have a probability density functions defined respectively as f (c_t), s (c_0) and g(D) with the respectively domains. In this case the most important study

is the relative importance of the stochastic variables in the solution and thus we will perform a sensitivity analysis.

As before, the general problem can be divided into a stochastic problem:

$$\frac{\partial C(z,t)}{\partial t} = D' \frac{\partial^2 C(z,t)}{\partial z^2}, \quad 0 < z < L, t > 0 \qquad \textit{Eq. 3-37}$$

with boundary and initial condition given by:

$$C(0,t) = c'_t, \qquad t > 0$$
$$C(L,t) = c_b, \qquad t > 0$$
$$C(z,0) = c'_0(z) \quad 0 < z < L$$

and a deterministic problem that reads:

$$\frac{\partial C(z,t)}{\partial t} = E[D] \ \frac{\partial^2 C(z,t)}{\partial z^2}, \quad 0 < z < L, t > 0 \qquad \textit{Eq. 3-38}$$

with the following boundary and initial condition:

$$C(0,t) = E[c_t], \qquad t > 0$$
$$C(L,t) = c_b, \qquad t > 0$$
$$C(z,0) = E[c_0(z)] \ \ 0 < z < L.$$

As in the previous section the solution can be written by an infinitive series (see…….) that we approximate. Given a positive integer N, the concentration can be represented by the stochastic process $C_N(z,t)$ defined by:

$$C_N(z,t) = \omega(z,t) + \sum_{n=1}^{N} k_n u_n(z,t) \, . \qquad \textit{Eq. 3-39}$$

We have to observe that the $u_n(z,t)$ contains only the diffusion coefficient random variable in the exponent (Eq.2-21) , k_n (Eq. 2-20) and $\omega(z,t)$ (Eq.2-19) contain the other two random variables but are linear .

We take these variables uniformly distributed in a given interval, with c_0 and D constant in space. The intervals of the distribution are computed, as previously, as:

$$c_t = [a_t, b_t] = [\mu_{ct} - coefvarc_t \cdot \mu_{ct}\,, \mu_{ct} + coefvarc_t \cdot \mu_{ct}] \qquad \textit{Eq. 3-40}$$

$$D = [a_D, b_D] = [\mu_D - coefvarD \cdot \mu_D\,, \mu_D + coefvarD \cdot \mu_D] \qquad \textit{Eq. 3-41}$$

$$c_0 = [a_0, b_0] = [\mu_{c0} - coefvarc_0 \cdot \mu_{c0}\,, \mu_{c0} + coefvarc_0 \cdot \mu_{c0}]. \qquad \textit{Eq. 3-42}$$

By using the properties of the mean and the variance (see Appendix B Eq.B22-Eq.B25) we compute the mean and variance of the concentration.

$$E[C_N(z,t)] = \frac{(b_t+a_t)}{2} + (c_b - \frac{(b_t+a_t)}{2})\frac{z}{L} + \sum_{n=1}^{N} 2\left[\frac{(b_0+a_0)}{2} - \frac{(b_t+a_t)}{2} + (-1)^n\left(c_b - \frac{(b_0+a_0)}{2}\right)\right]\frac{\sin z\lambda}{\lambda L}\ sin\left(\frac{n\pi z}{L}\right)\frac{1}{(b_D-a_D)}\left[\frac{L^2}{n^2 t\pi^2}\left(e^{-\frac{n^2 a_D t\pi^2}{L^2}} - e^{-\frac{n^2 b_D t\pi^2}{L^2}}\right)\right] \qquad \textit{Eq. 3-43}$$

$$VAR[C_N(z,t)] = VAR\left[C_{N_{ct}}(z,t)\right] + VAR\left[C_{N_{ct}}(z,t)\right]VAR\left[C_{N_D}(z,t)\right] + VAR\left[C_{N_{ct}}(z,t)\right]E\left[C_{N_D}(z,t)\right]^2 + VAR\left[C_{N_D}(z,t)\right]E\left[C_{N_{ct}}(z,t)\right]^2 + VAR\left[C_{N_{c0}}(z,t)\right]VAR\left[C_{N_D}(z,t)\right] + VAR\left[C_{N_{c0}}(z,t)\right]E\left[C_{N_D}(z,t)\right]^2 + VAR\left[C_{N_D}(z,t)\right]E\left[C_{N_{c0}}(z,t)\right]^2 \qquad \textit{Eq. 3-44}$$

where these quantities have been previously computed:

$$VAR\left[C_{N\_ct}(z,t)\right] = Eq\ 3\text{-}33$$

$$VAR\left[C_{N\_c0}(z,t)\right] = Eq\ 3\text{-}4$$

$$VAR\left[C_{N\_D}(z,t)\right] = Eq\ 3\text{-}25$$

$$E\left[C_{N\_c0}(z,t)\right] = \ Eq\ 3\text{-}5$$

$$E\left[C_{N\_D}(z,t)\right] = Eq\ 3\text{-}24$$

$$E\left[C_{N_{ct}}(z,t)\right] = Eq\ 3\text{-}28.$$

3.5.1 Concentration GSA

In the follow section we show the global sensitivity analysis results. We take into account the First Sobol's indices for the concentration deriving the analytical expression and we compare it with the numerical solution. We analyze the case where the intervals of the random variables are:

$c_t = [a_t, b_t] = [0.75, 1.25]$

$D = [a_D, b_D] =, 1.25^{-12}]m^2/s$

$c_0 = [a_0, b_0] = [0.15\ , 0.25].$

3.5.1.1 Analytic Total order Sobol's Indices

We have seen in 2.4.1 as the Sobol's indices can be a very useful indicator of the relative importance of the stochastic parameter. They quantify the total contribution of the uncertainty of each input parameter to the variance of the output. The total

Sobol's indices expressions can be computed by the definition. We consider into the numerator of the Eq.2-38 for every input only the parts of the total variance where it compares. We obtain so:

$$S_{C0} = \frac{VAR[C_{N\_c0}(z,t)] + VAR[C_{N\_c0}(z,t)]E[C_{N\_D}(z,t)]^2 + VAR[C_{N\_c0}(z,t)]VAR[C_{N\_D}(z,t)]}{VAR[C_N(z,t)]} \quad \textit{Eq. 3-45}$$

$$S_D = \frac{\begin{matrix} VAR[C_{N_D}(z,t)]E[C_{N_{c0}}(z,t)]^2 + VAR[C_{N_D}(z,t)]E[C_{N_{ct}}(z,t)]^2 \\ +VAR[C_{N\_ct}(z,t)]VAR[C_{N\_D}(z,t)] + VAR[C_{N\_c0}(z,t)]VAR[C_{N\_D}(z,t)] \end{matrix}}{VAR[C_N(z,t)]} \quad \textit{Eq. 3-46}$$

$$S_{Ct} = \frac{VAR[C_{N_{ct}}(z,t)] + VAR[C_{N_{ct}}(z,t)]E[C_{N_D}(z,t)]^2 + VAR[C_{N_{ct}}(z,t)]VAR[C_{N_D}(z,t)]}{VAR[C_N(z,t)]} \quad \textit{Eq. 3-47}$$

3.5.1.2 Analytic First order Sobol's Indices

We can for the homogenous problem, compute also the analytical expression of the first order Sobol's Indices. From the definition of the First order Sobol's indices, they quantify the contribution of the single input to the variance of the output, without the contribution of the mutual effect of the inputs. The first order Sobol's indices for the concentration, considering the definition above and the Eq.2-36 are:

$$S_{C0} = \frac{VAR[C_{N_{c0}}(z,t)] + VAR[C_{N_{c0}}(z,t)]E[C_{N_D}(z,t)]^2}{VAR[C_N(z,t)]} \quad \textit{Eq. 3-48}$$

$$S_D = \frac{VAR[C_{N_D}(z,t)]E[C_{N_{c0}}(z,t)]^2 + VAR[C_{N_D}(z,t)]E[C_{N_{ct}}(z,t)]^2}{VAR[C_N(z,t)]} \quad \textit{Eq. 3-49}$$

$$S_{Ct} = \frac{VAR[C_{N_{ct}}(z,t)] + VAR[C_{N_{ct}}(z,t)]E[C_{N_D}(z,t)]^2}{VAR[C_N(z,t)]} \quad \textit{Eq. 3-50}$$

Figure 3-24 show the First order Sobol index relative of every random input using the analytic expression computed in Eq.3-52, Eq.3-53 and Eq.3-54. The variation coefficient considered is 0.25 for all stochastic variables.

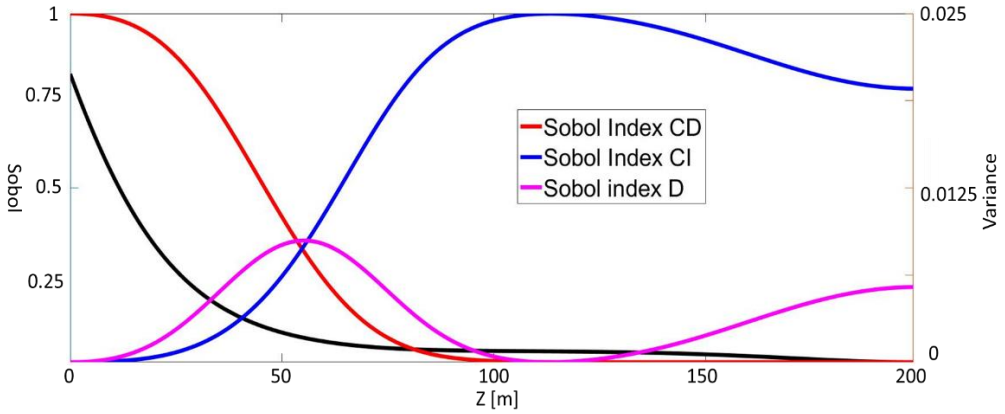
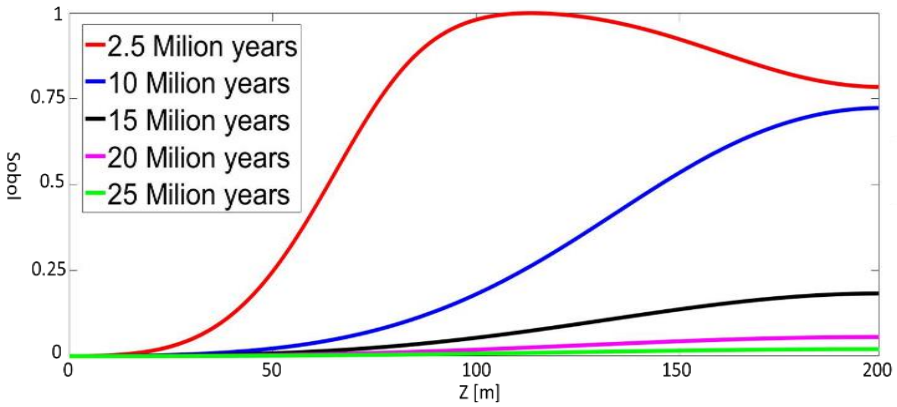
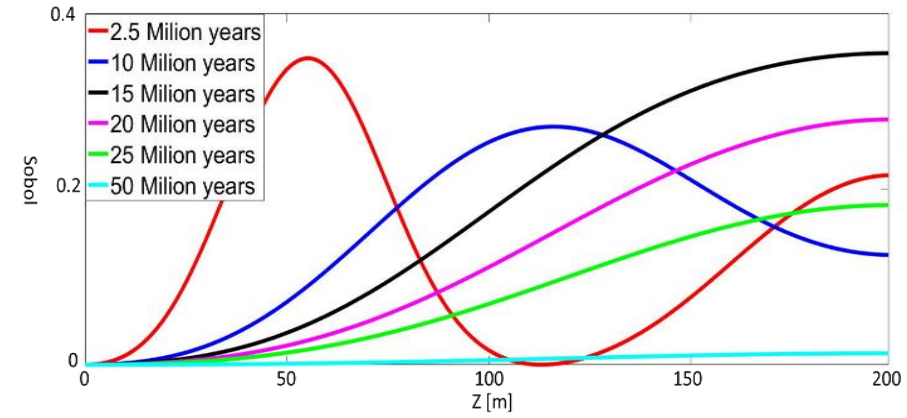


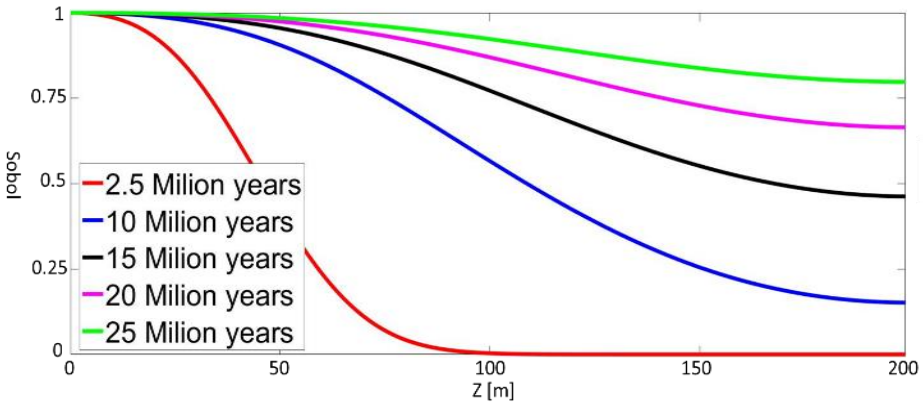
Fig 3-24 Analytical results for the first order Sobol indices Eq. 3-52 Eq.3-53 Eq3-54 for the problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years. The black line is the total variance of the concentration solution.



(a)



(b)



(c)

Fig 3-25 Sobol indices analytical results for different diffusion time. In (a) Sobol index for initial condition, n (b) for the diffusion coefficient, in (c) for boundary condition.

Fig.3-24 shows that initially the uncertainty due to the stochastic boundary condition does not influence significantly the solution uncertainty, but, after a given time it has the predominant influence.

The Sobol index of the diffusion coefficient, after that the boundary condition effect arrives in z=L, increases until a certain time (c.a. 15 million years in example Fig.3-25-b). After this time, the contribution to the total variance from the uncertainty of D decreases and tends to zero. This behavior has been seen also before, in Fig.3-15. Indeed, the Initial condition plays an important role for short time, but once we see the effect of the boundary condition it decreases.

We can note that the uncertainty about the diffusion coefficient is directly linked to the flux amplitude and its relative velocity. As expected, when we haven't concentration gradient and consequent null flux, uncertainty related to the diffusion coefficient does not play a role.

3.5.1.3 Analytic Second order Sobol's Indices

From the Eq. 2-36 we can compute the second order Sobol's indices as:

$$S_{CI,D} = \frac{VAR[C_{N\_c0}(z,t)]VAR[C_{N\_D}(z,t)]}{VAR[C_N(z,t)]} \qquad \textit{Eq. 3-51}$$

$$S_{Ct,D} = \frac{VAR[C_{N\_ct}(z,t)]VAR[C_{N\_D}(z,t)]}{VAR[C_N(z,t)]} \qquad \textit{Eq. 3-52}$$

$$S_{CI,Ct} = \frac{0}{VAR[C_N(z,t)]} = 0 \qquad . \qquad \textit{Eq. 3-53}$$

Second order Sobol's indices quantify the contribution to the total variance of the output of varying one single parameter and two parameters simultaneously, adding the effect of their iteration to the effect of their individual variation.

Fig 3-26 show the Second order Sobol index relative of every random input using the analytic expression computed in Eq.3-55 Eq.3-54. The variation coefficient considered is 0.25 for all stochastic variables.

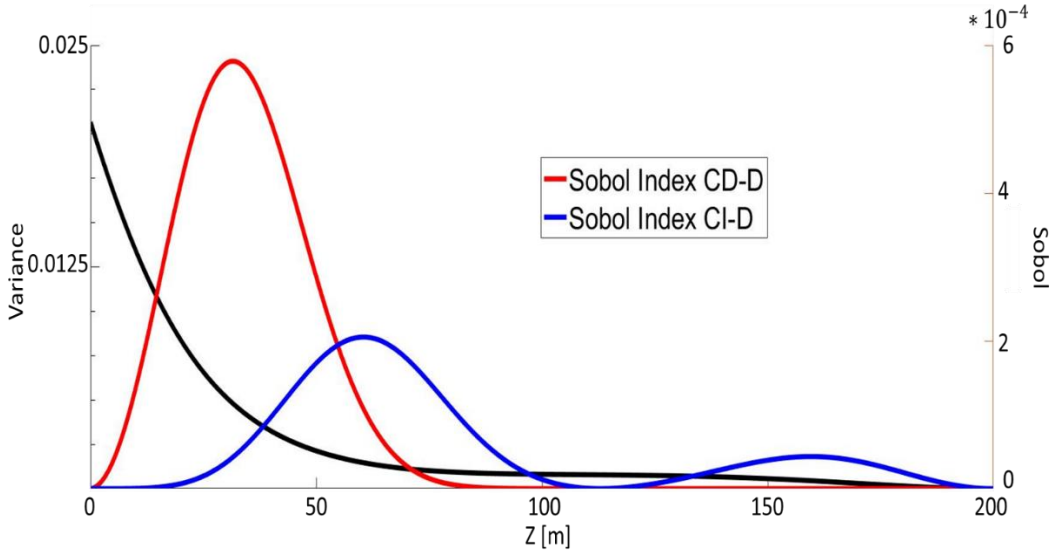


Fig 3-26 Analytical second order Sobol indices Eq.3-55 Eq.3-56 for the problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years. The black line is the total variance of the problem.

We observe that the contribution of the iteration between the inputs for the total variance of the output is very small.

In order to study the effect of this interaction for longer time Figure 3-27 illustrate the second order Sobol's indices Eq.3-51, Eq.3-52 and Eq.3-53 for a diffusion time of 20 million years.

We see that increasing the diffusion time we observe that the index referred to the iteration between the initial condition and the diffusion coefficient tends to annul. The maximum of the index of the iteration between the boundary condition and the diffusion coefficient instead tend to decrease. It has the middle position between the z=0 and the most distant point influenced by the boundary condition.

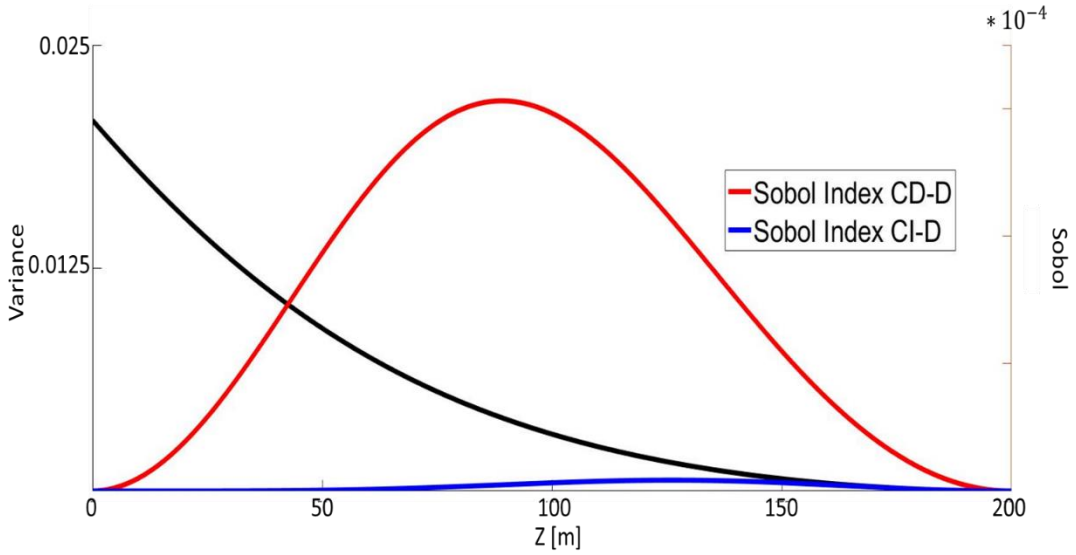


Fig 3-27 Analytical second order Sobol indices Eq.3-55 and Eq.3-56 for the problem studied with all input parameters randomly uniformly distributed, with coefvar 0.25 for all of them (Eq.3-51,Eq.3-52 Eq.3-53). Diffusion time is 20 million years. The black line is the total variance of the problem.

3.5.1.4 Numerical First-order Sobol's Indices

The calculation of the Sobol indices is normally performed numerically. In our case, we have a series of triads of the stochastic parameters taken random (our inputs). Remember that we assume all random variables independent, so each parameter is sampled from its marginal distribution independently from the others. For each triad, we compute with the model the concentration solution that is our output, for a certain time and space z.

As in (Saltelli, et al., 2008) we create a matrix of 4 columns (for the three inputs and the output) and n lines where n corresponds to the number Monte Carlo simulations. We create for each input a matrix of n x 2 (where the 2 columns are the input in exam e the relative output) in ascending order for the input and so we obtain 3 scatters plot: for stochastic boundary condition Fig 3-28, stochastic Initial condition Fig 3-29 and for stochastic diffusion coefficient Fig 3-30. Linear regression are performed over for n=32000 realizations.

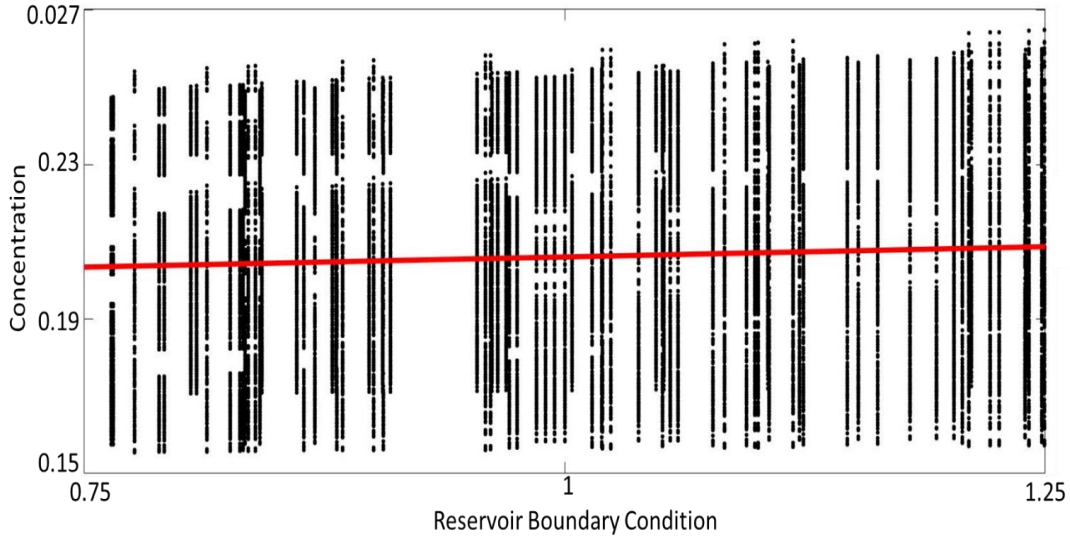


Fig 3-28 Scatter plot (for c.a. 3000 Montecarlo Simulation) about the stochastic boundary condition and the linear regression of the points solution.

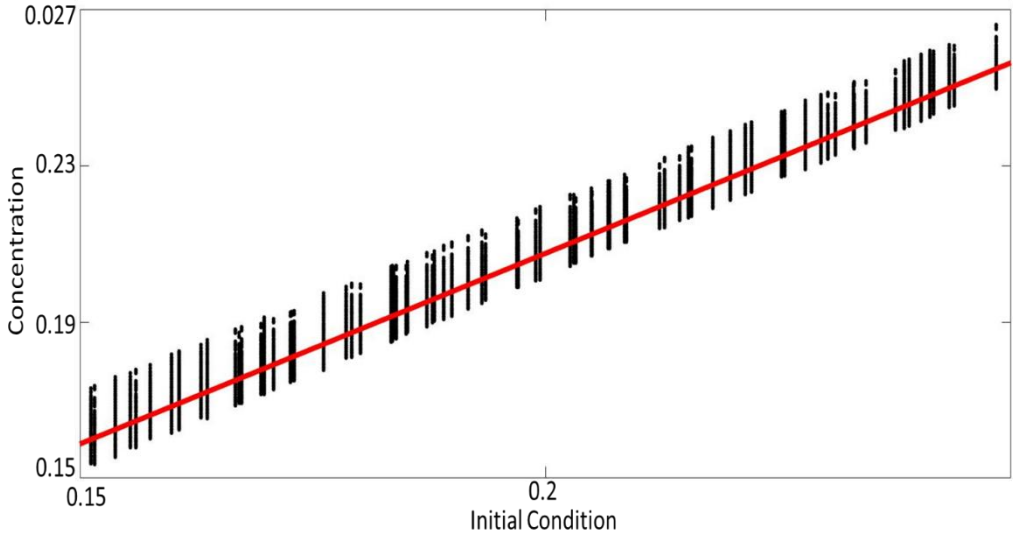


Fig 3-29 Scatter plot (for c.a. 3000 Montecarlo Simulation) about the stochastic Initial condition and the linear regression of the points solution.

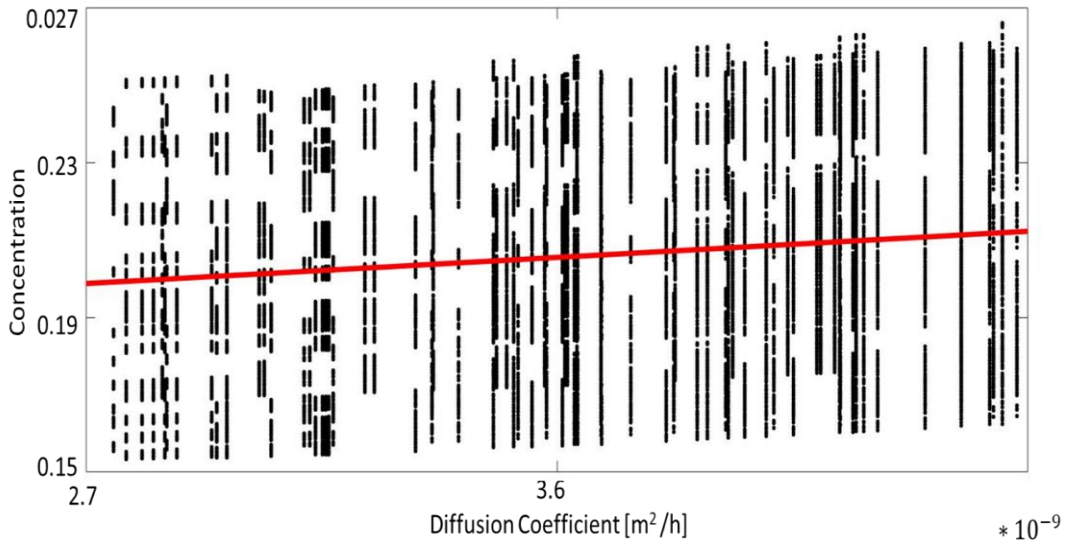


Fig 3-30 Scatter plot (for c.a. 3000 Montecarlo Simulation) about the stochastic diffusion coefficient and the linear regression of the points solution.

In order to compute numerically the Sobol indices, we perform a linear regression on the data given by:

$$C(z,t) = b_0 + \sum_{j=1}^{3} b_{z_j} z_j \quad , \qquad \textit{Eq. 3-54}$$

where the coefficients b_0, b_{z_j} are determined by least-square computation, based on the squared differences between the C-values produced by the regression model and the actual model output produced by Monte Carlo simulation and the subscript j.

For example we compute the indices in z=L/2:

$$C\left(\frac{L}{2}, 2.5 \cdot 10^6 y\right) = 0.1785 + 0.0000075 * D \qquad \textit{Eq. 3-55}$$

$$C\left(\frac{L}{2}, 2.5 \cdot 10^6 y\right) = 0.01249 + 0.9752 * C_0 \qquad \textit{Eq. 3-56}$$

$$C\left(\frac{L}{2}, 2.5 \cdot 10^6 y\right) = 0.1946 + 0.01089 * Ct \qquad \textit{Eq. 3-57}$$

If now we normalized the coefficient b_{z_j} on the standard deviation of the input and of the output we obtain the standardized regression coefficients, that are generally more used respect to b_i

$$\beta_D = \frac{\sigma_D}{\sigma_{CON}} b_D = \frac{4.9568*10^{-10}}{0.0271} b_D = 1.3718 * 10^{-13} \quad \textit{Eq. 3-58}$$

$$\beta_{Ci} = \frac{\sigma_{Ci}}{\sigma_{CON}} b_{Ci} = 0.9907 \quad \textit{Eq. 3-59}$$

$$\beta_{ct} = \frac{\sigma_{ct}}{\sigma_{CON}} b_{Ct} = 0.0573 \quad \textit{Eq. 3-60}$$

If we sum all the squared standardized regression coefficients we obtain the model coefficient of determination:

$$R_Y^2 = \Sigma_{j=1}^{3} \beta_j^{\,2} = 0.9848. \quad \textit{Eq. 3-61}$$

This number is the fraction of linearity of our model. In fact, if the model was perfectly linear we will obtain the unity. More precisely, this number, is equal to the fraction of the variance of the original data which come from our model which is explained by the regression model of Equation (Saltelli, et al., 2008). The square of beta coefficients is the first order Sobol's index:

$$S_j = \beta_j^{\,2} \quad \textit{Eq. 3-62}$$

We compute the Sobol index on the whole domain for 2.5 and 8 million years diffusion time in Fig 3-31 and Fig3-33.

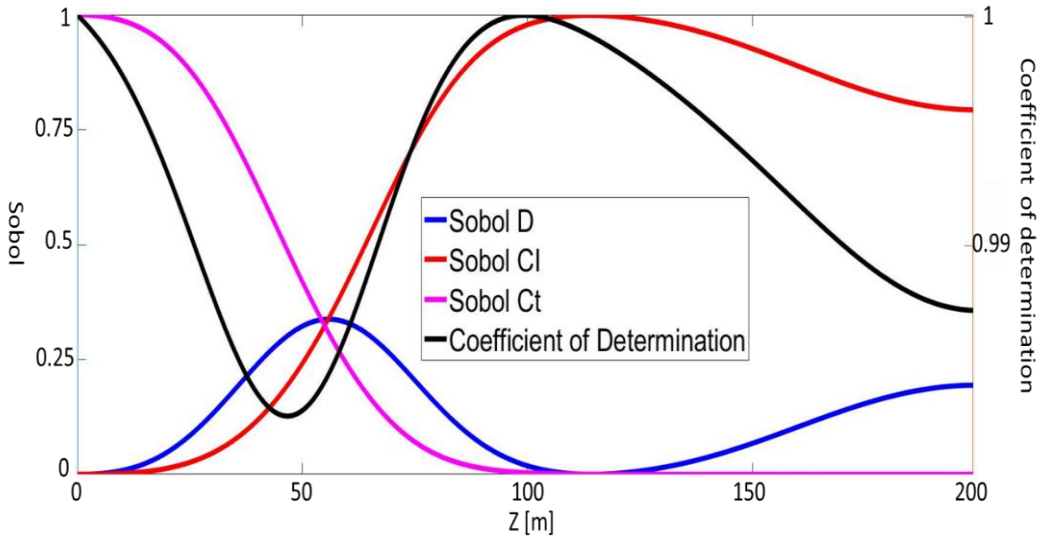


Fig 3-31 Numerical Sobol indices for the problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years.

We compare the numerical solution with the analytical solution for the first order Sobol's indices computed for 2.5 and 5 million years we can observe that for a number of combination major of 48000 we obtain already a good convergence between the two solutions:

Simulations	$S_D^{anl} - S_D^{num}$	$S_{Ct}^{anl} - S_{Ct}^{num}$	$S_{Ci}^{anl} - S_{Ci}^{num}$
20000	0.0381	-0.0136	0.0088
40000	0.0134	-0.0022	0.0015
48000	-0.0019	-0.0001	-0.002

Tab 2 Comparison between Analytical and Numerical Sobol's Indices

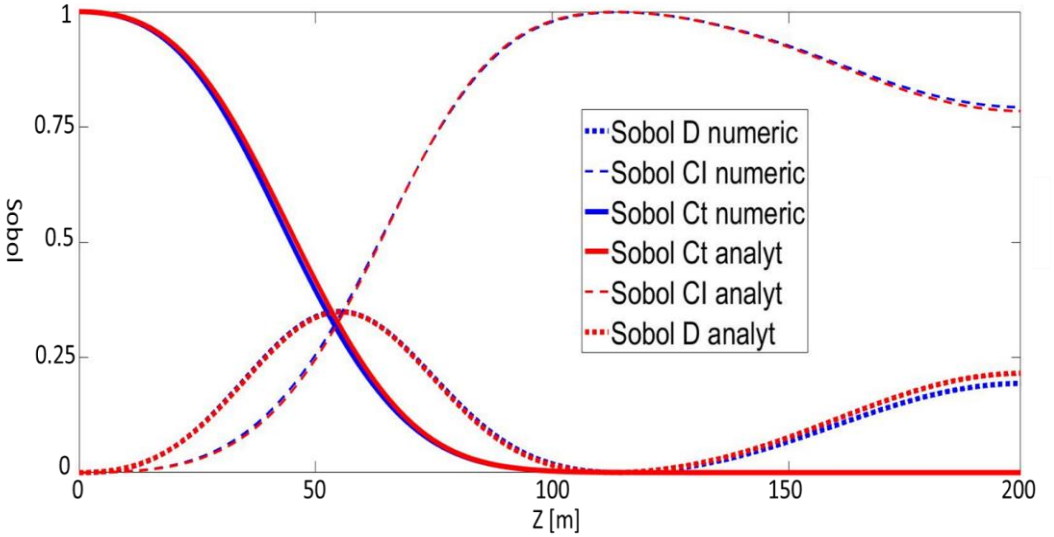


Fig 3-32 Comparison between the two different methods to compute the Sobol indices with coefvar 0.25 for all of the inputs. The diffusion time is 2.5 million years. In blue the numerical method, in red the analytical Eq.3-49 Eq.3-50 and Eq.3-51.

The computational time between the two solution is completely different. The analytical Sobol index is computed almost instantly, instead the Sobol index computational time with the linear regression, using the procedure above, need almost half hour.

To better understand the behavior of the diffusion solution in Fig 3-33 we observe the Sobol's indices also for a 8 million years diffusion time.

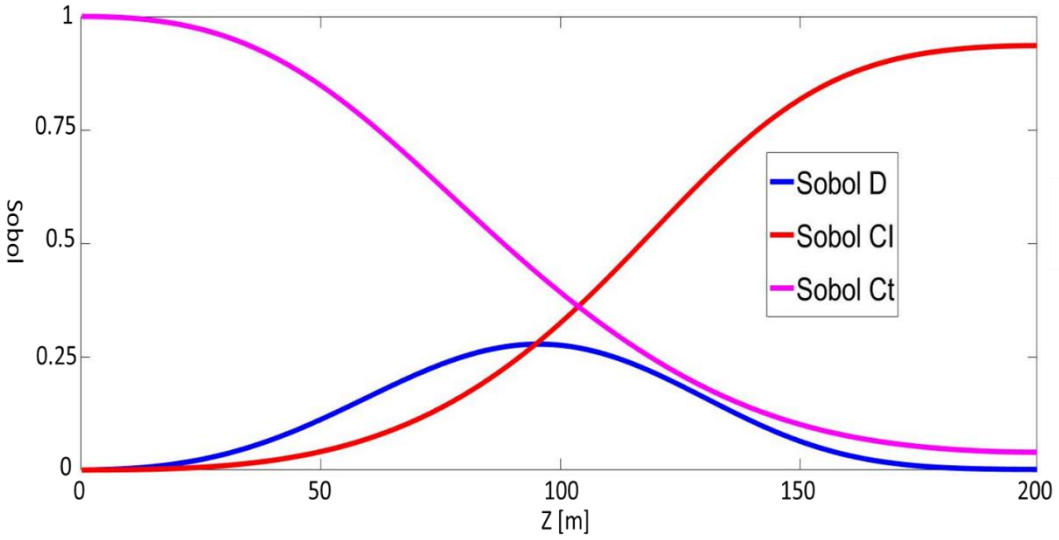


Fig 3-33 Numerical Sobol indices for the problem with all input parameters random and with a coefvar 0.25s. The diffusion time is 8 million years.

We can observe how the point where the initial condition and the boundary condition have the same weight is shifted to the right. This because, in time, boundary condition tends to take the principal influence in the space domain and in the same time, the initial condition, tends to vanish in time. Another observation is that the uncertainty due to diffusion coefficient, as already observed Fig 3-15, play a less role for the total uncertainty with increasing time.

3.5.1.5 Numerical Second-order Sobol's Indices

Two factors are said to interact when their effect on output cannot be expressed as a sum of their single effects. Interactions may imply, for instance, that extreme values of the output Y are uniquely associated with particular combinations of model inputs, in a way that is not described by the first-order effects above mentioned. Interactions represent important features of models, and are more difficult to detect than first-order effects.

We can see the first Sobol's indices as the ratio between the variance of the conditioned mean of the output by the input:

$$S_i = \frac{Var_i\left(E(C|X_i)\right)}{Var(C)} \qquad \textit{Eq. 3-63}$$

where X_i is the input random variable.

In the same way we define the second order Sobol's index as the ratio between the variance of the output mean conditioned by two inputs:

$$S_{ij} = \frac{Var_{ij}\left(E\left(C|X_i, X_j\right)\right)}{Var(C)} - \frac{Var_i\left(E(C|X_i)\right)}{Var(C)} - \frac{Var_i\left(E\left(C|X_j\right)\right)}{Var(C)} \qquad \textit{Eq. 3-64}$$

and thus (Saltelli, et al., 2008):

$$S_{ij} = \frac{Var_{ij}\left(E\left(C|X_i, X_j\right)\right)}{Var(C)} - S_i - S_j \quad . \qquad \textit{Eq. 3-65}$$

The function S_{ij} captures that part of the response of the concentration to X_i, X_j that cannot be written as a superposition of distinct effect of constrain X_i or X_j separately.

3.5.1.6 Multiple-moment-based metrics for Global Sensitivity Analysis

Here we use another new Global Sensitivity Analysis (GSA) method based on Multiple-moments metrics (Dell'Oca, Riva, & Guadagnini, 2017). These indices can be obtained numerically computing the single parts that are in the formulas. The first four conditioned and unconditioned statistical moments that are taken into account are: the mean of the output its variance, the skewness and the kurtosis.

To obtain them, we create a matrix of 4 columns (for the three inputs and the output) and as many lines as Monte Carlo simulations "n". Then for each input we obtain a matrix of n x 2 elements (the input in exam and the output) in ascending order for the input and so we obtain 3 scatters plots (as Fig 3-28,

Fig 3-29 and Fig 3-30) with the input conditioned value on the x-axe and the output in y-axe. We can finally obtain the conditioned moments wanted, computing these on the output. The procedure is similar with the Sobol numeric computing procedure.

In Figure 3-34 we observe the AMAE indices (Eq.2-39) along the space domain for 2.5 million years of time diffusion.

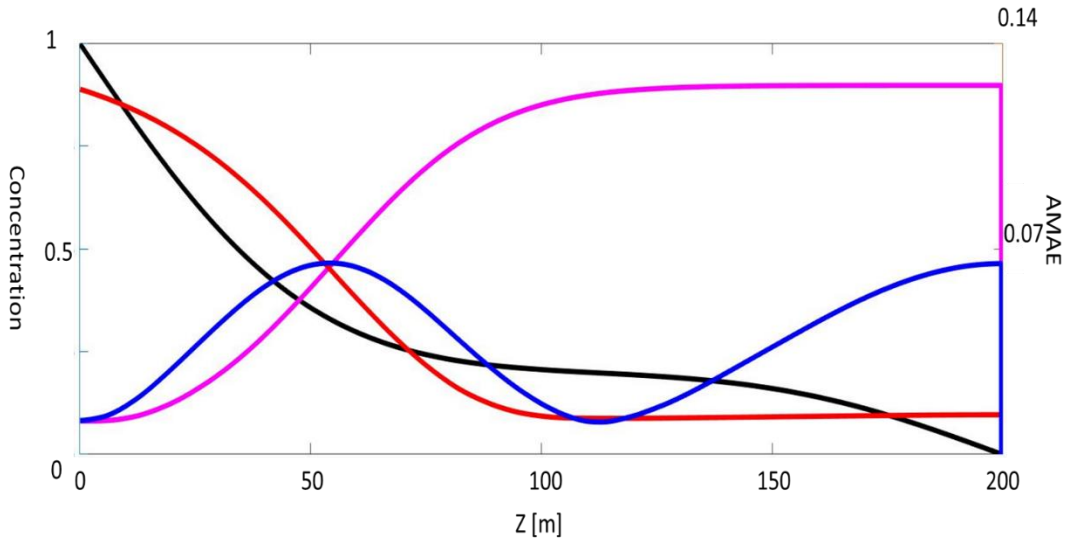


Fig 3-34 Numerical AMAE Eq.2-39 indices for the problem with all input random and with coefvar 0.25 for all of them. The diffusion time is 2.5 million years. The purple line is referred to initial condition, red line to boundary condition and the blue line to diffusion coefficient. Black Line is the relative concentration solution.

We can note that the main contribution in the concentration solution until z < c.a. 55 m is given by the boundary condition as expected, while for larger z, the initial condition and the diffusion coefficient assume influence most the concentration. The unusual big contribute of D could be seen because it is direct linked with the difference between the stationary condition and the example condition. Moreover, we observe that we have two maximums for the AMAE-D: the first where the boundary condition have the same influence of initial condition in the reach the stationary condition and the second (at z=L) where the initial condition have the maximum influence in the gradient generation having to react to the null boundary condition.

In Figure 3-35 we observe AMAV index along the space domain for 2.5 million years of time diffusion.

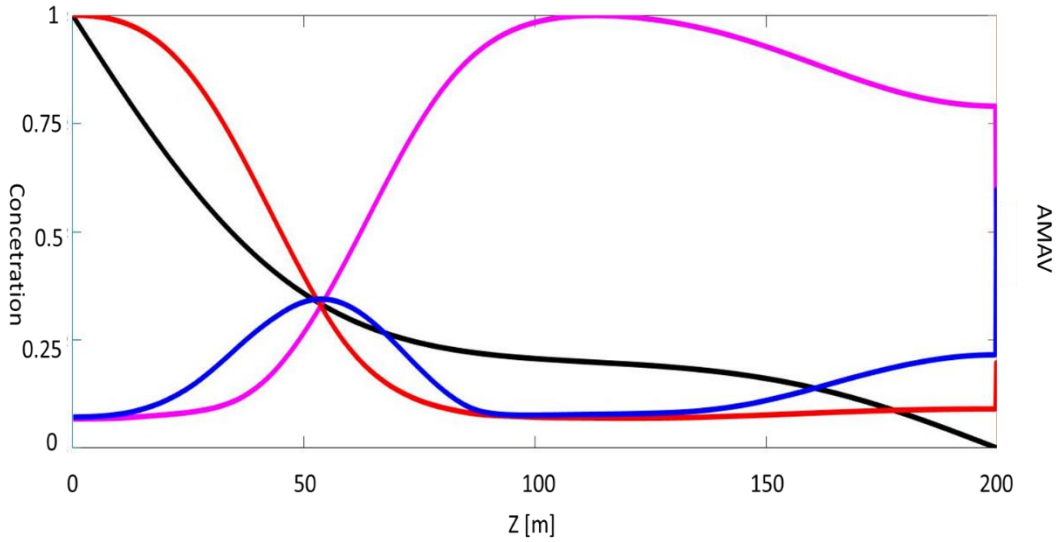


Fig 3-35 Numerical AMAV indices Eq.2-40 for the problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years. The purple line is referred to initial condition, red line to boundary condition and the blue line to diffusion coefficient.

We can note how, that the AMAV for the boundary condition, diffusion coefficient and for the Initial condition are similar to the Sobol's Indices profile seen before. To validate these considerations, we show in Figure 3-36 the same indices for a larger diffusion time (20 million years), where we can hypothesize stationarity. In Figure 3-36 we observe how, as already seen for Sobol's indices, the initial condition and the diffusion coefficient don't influence the mean solution and the variance.

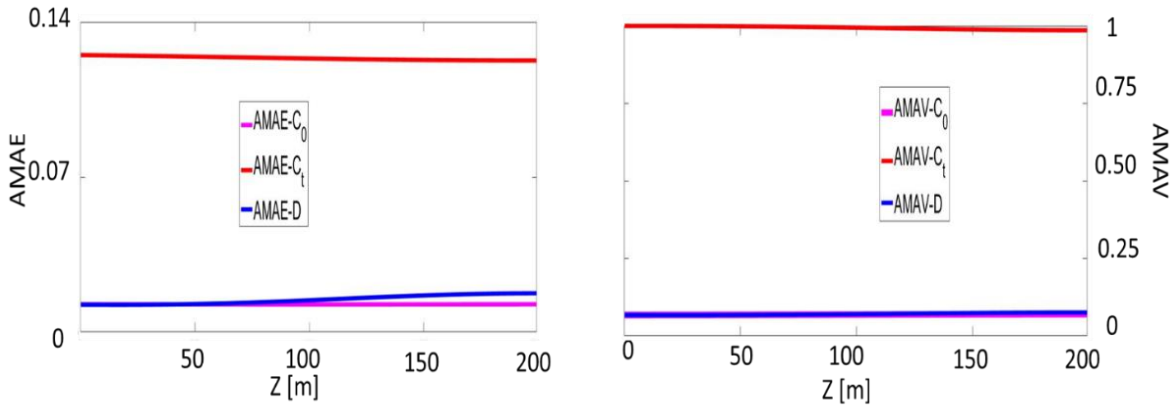


Fig 3-36 AMAE and AMAV indices computed numerically for the three inputs random with coefvar 0.25.. The diffusion time is 2.5 million years

Much evidence could be seen also observing Fig 3-37 where we observe the AMAE (a, b) and AMAV (c, d) indices plotted respect to the concentration for a time diffusion of 2.5 and 8 million years and the stationarity concentration. Here we have that initial condition is null and so it doesn't influence the solution. We see that where the distance is much bigger between the concentration and stationarity concentration we have that the mean solution is influenced by the diffusion coefficient much more respect than the second case. Another comparison is that the boundary condition influenced the solution constantly, (giving the gradient), but not the variance (and so the uncertainty). The diffusion coefficient instead influences mean and variance in function of the gradient that is present.

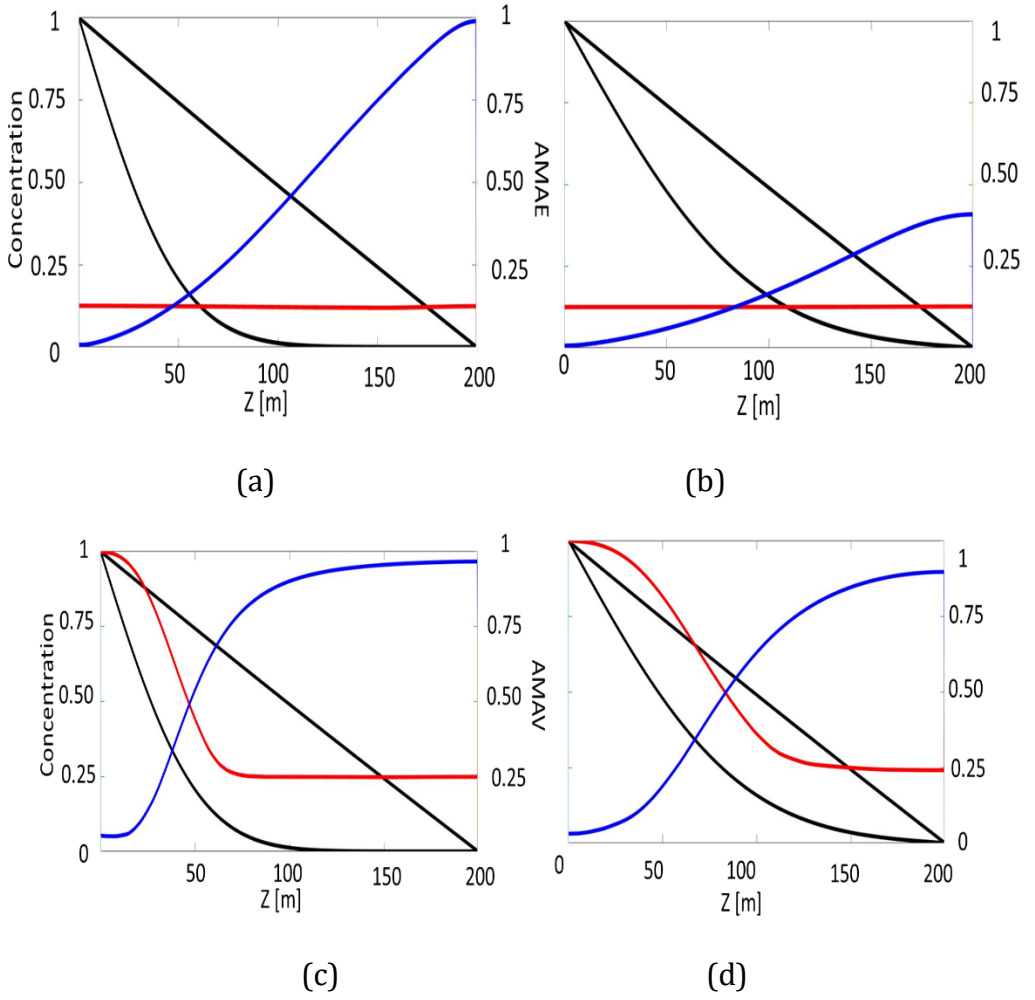


Fig 3-37 Numerical AMAE and AMAV indices for the problem with every input random and with coefvar 0.25 for the diffusion coefficient and the boundary condition. The blue lines are referred to diffusion

coefficient while the red line to the boundary condition. The initial condition is zero. The diffusion time is 2.5 million years for the (a) and (c) and 8 million years for (b) and (d). The black lines are the concentration profile in object and the stationary concentration profile (the diagonal).

In Figure 3-38 and 3-39 we observe the fluctuation of the conditioned mean of the concentration by the three inputs. On the x-axe we have the normalized inputs on its mean. The fluctuations in Fig 3-38 are due to the interval width that is taken into consideration in the calculation of the statistics.

The analytical solution has been obtained using:

$$E[C_N(z,t)|x_i] = \int_{a_{x_j}}^{b_{x_j}} \int_{a_{x_k}}^{b_{x_k}} E[C_N(z,t)]\, dx_j dx_k \qquad \textit{Eq. 3-66}$$

Where x_i, x_j, x_k are in rotation the random inputs.

We observe that while for z=L/4 the three inputs have the same influence, and as their value increase also the expected concentration increases, in z=L/2 the initial condition mostly influence the expected concentration. In fact, in this position, as we can also observe in the concentration solution above, the boundary condition in z=0 hasn't influenced much the solution yet, as well as the boundary condition in z=L.

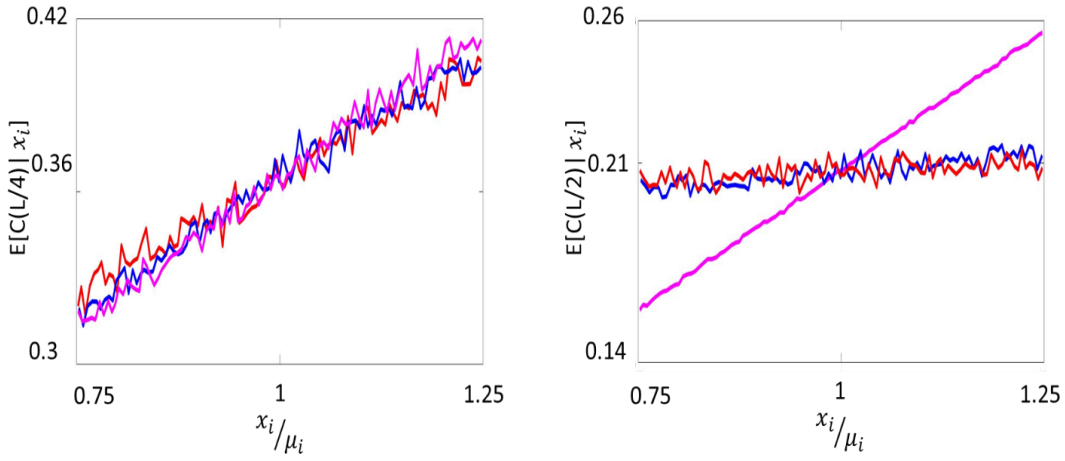


Fig 3-38 Numerical conditioned mean for the problem with every random inputs and with coefvar 0.25 for every of theme. x_i could be the Initial condition (purple line), boundary condition (red line) or the diffusion coefficient (blue line). On the x ax there is the input normalized on its mean. The initial condition is 0.2. On the left is in z=L/4, on the right z=L/2.

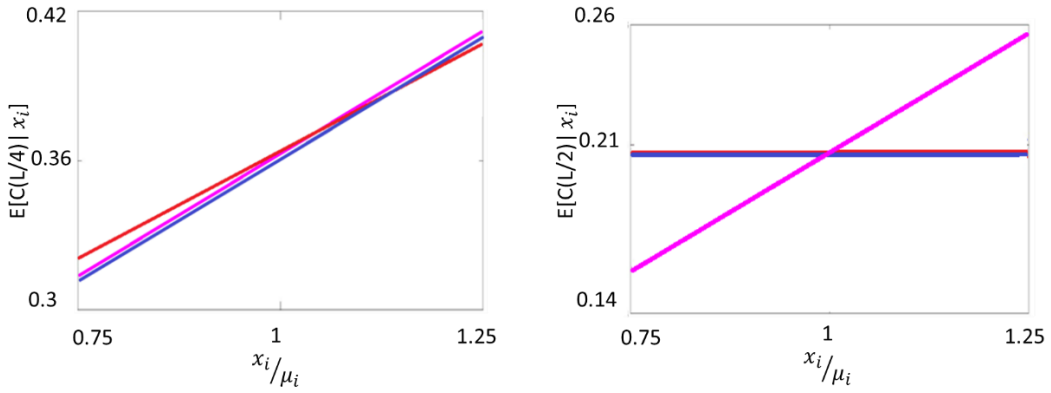


Fig 3-39 Analytical conditioned mean for the problem with every random inputs and with coefvar 0.25 for every of theme. x_i could be the Initial condition (purple line), boundary condition (red line) or the diffusion coefficient (blue line). On the x ax there is the input normalized on its mean. The initial condition is 0.2. On the left is in z=L/4, on the right z=L/2.

In Figure 3-40 and 3-41 we observe the fluctuation of the conditioned variance of the concentration by the three inputs. The analytical solution has been obtained using:

$$VAR[C_N(z,t)|x_i] = \int_{a_{x_j}}^{b_{x_j}} \int_{a_{x_k}}^{b_{x_k}} VAR[C_N(z,t)]dx_j dx_k \qquad \textit{Eq. 3-67}$$

Where x_i, x_j, x_k are in rotation the random inputs.

We see in Fig 3-41 that for z=L/4 the conditioned variance (and consequently the uncertainty due to that input) for the diffusion coefficient and for the initial condition increase with the increase of the input value. The boundary condition instead doesn't increase the variance with his increasing. In Z=L/2 we observe that the boundary condition doesn't influenced the variance solution because the same reason already seen in conditioned mean. The initial condition and the boundary condition doesn't change the variance with their changing.

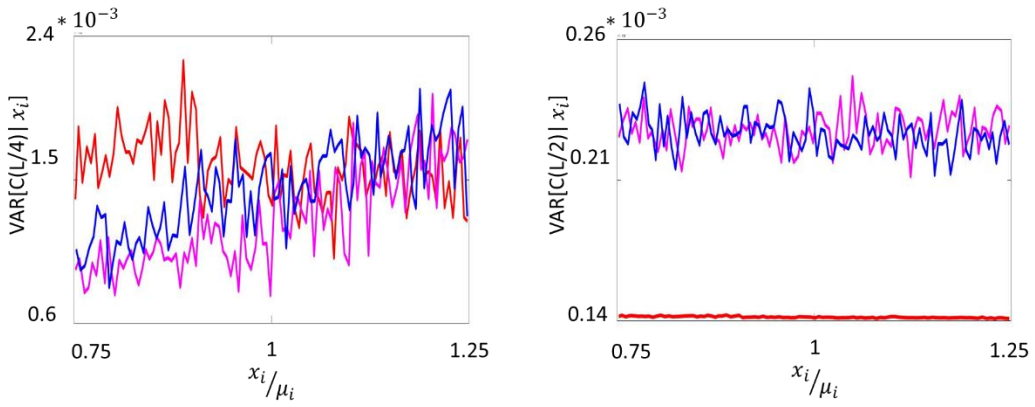


Fig 3-40 Numerical conditioned variance for the problem with every random inputs and with coefvar 0.25 for every of theme. x_i could be the Initial condition (purple line), boundary condition (red line) or the diffusion coefficient (blue line). On the x ax there is the input normalized on its mean. The unconditioned variance on the left is 0.0023, on the right is 0.00023. The initial condition is 0.2. On the left is in z=L/4, on the right z=L/2.

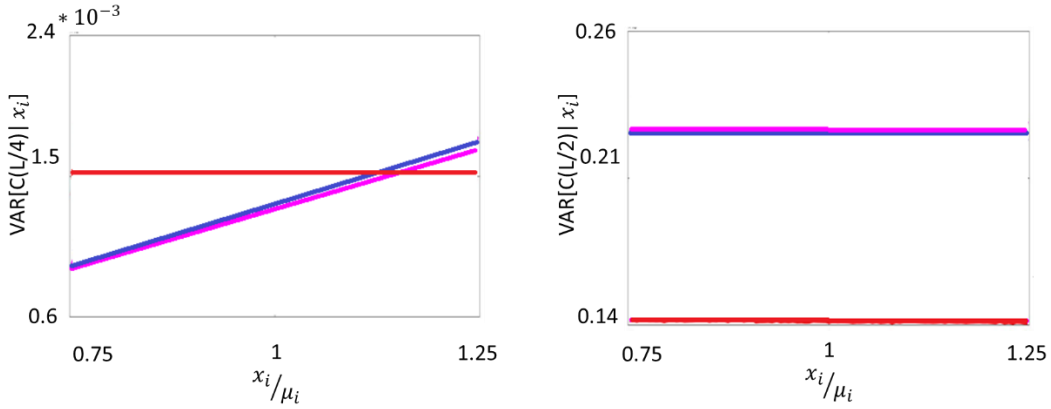


Fig 3-41 Analytical conditioned variance for the problem with every random inputs and with coefvar 0.25 for every of theme. x_i could be the Initial condition (purple line), boundary condition (red line) or the diffusion coefficient (blue line). On the x ax there is the input normalized on its mean. The unconditioned variance on the left is 0.0023, on the right is 0.00023. The initial condition is 0.2. On the left is in z=L/4, on the right z=L/2.

We computed also the third and fourth AMA-indeces, AMAG and AMAK, linked with the the third and fourth moments of the output, or rather the skewness, and kurtosis, respectively. In Figure 3-42 we observe the AMAG profiles for the three inputs. As we can see it is always null. The AMAG, being null, say that the solution is symmetric respect the mean.

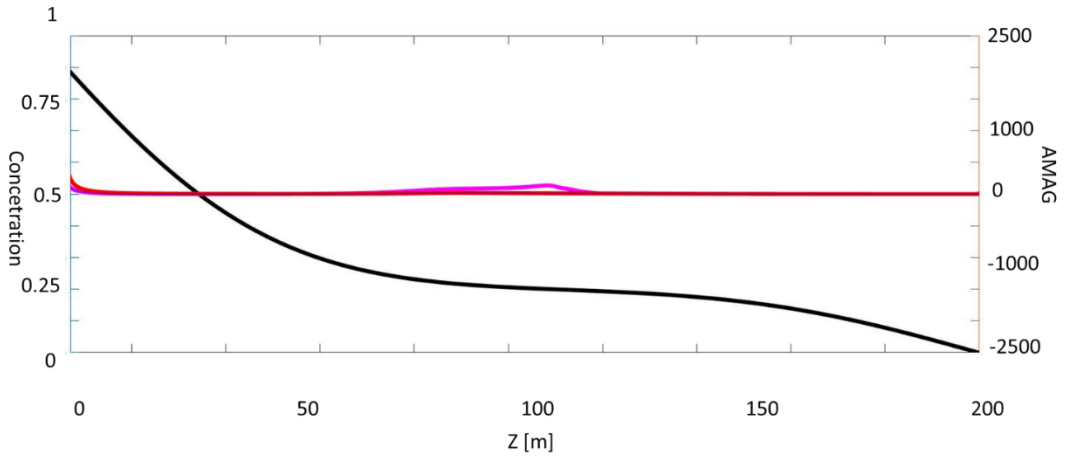


Fig 3-42 Numerical AMAG indices for the problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years. The purple line is referred to initial condition, red line to boundary condition and the blue line, overlapped under the red line, to diffusion coefficient.

In Figure 3-43 we observe the fluctuation of the conditioned skewness of the concentration by the three inputs This give us the information that the solution is symmetric respect his mean and that the inputs doesn't influenced it. We note that the AMAG indices are zero in each point (z=L/4 and z=L/2) and so they don't depend by the different distance from the stationarity solution.

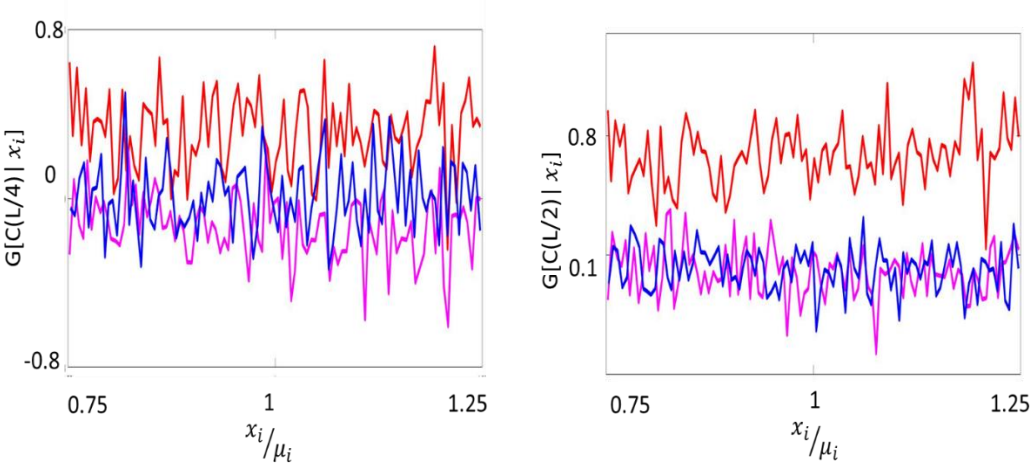


Fig 3-43 Numerical conditioned Skewness for the problem with every random inputs and with coefvar 0.25 for every of theme. x_i could be the Initial condition (purple line), boundary condition (red line) or the diffusion coefficient (blue line). On the x ax there is the input normalized on its mean.. The initial condition is 0.2. On the left is in z=L/4, on the right z=L/2.

Fig 3-44 shows the AMAK indices. We can note the maximum for the AMAK indices at z=111 m. AMAK index give information about the distance of the solution from the gaussian distribution. The abnormal value in that point, that

is also the point doesn't influenced by the boundary conditions and so where the flux is null, could be explain with the fact that there the initial condition is invariant.

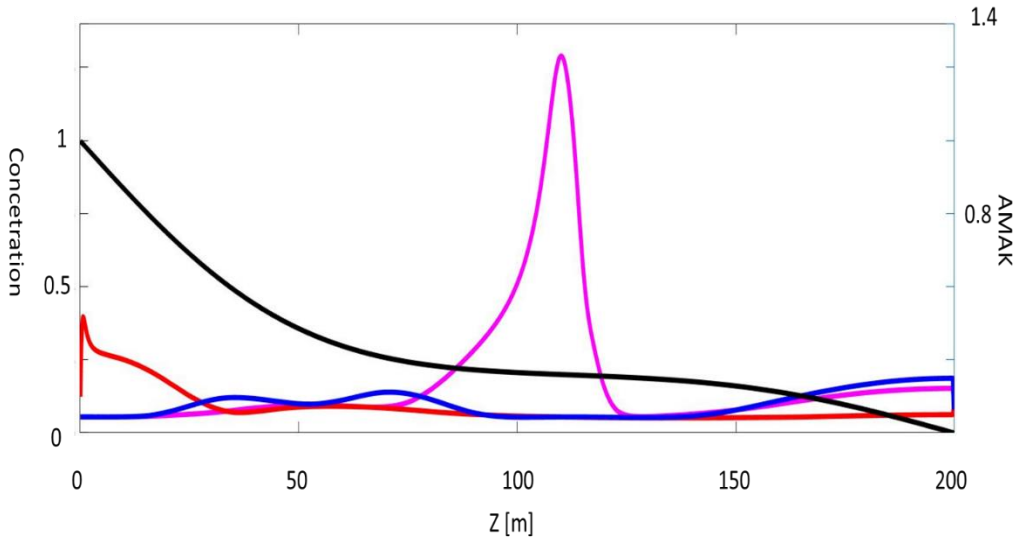


Fig 3-44 Numerical AMAK indices for the problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years. The purple line is referred to initial condition, red line to boundary condition and the blue line to diffusion coefficient

Fig 3-45 better shows the behavior of the skewness and the kurtosis in the point where we have the maximum. We note that the third and fourth moment conditioned by the initial condition are very different respect others two.

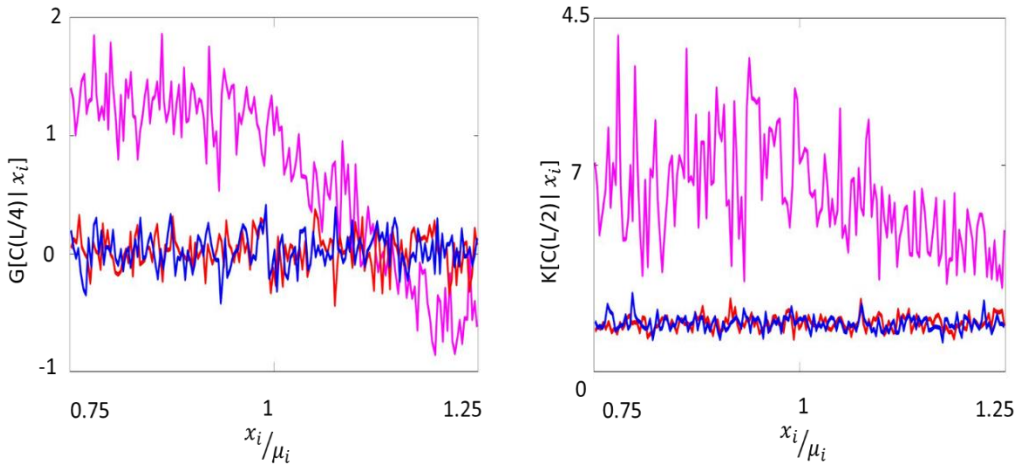


Fig 3-45 Numerical conditioned Skewness and Kurtosis for the problem with every random inputs and with coefvar 0.25 for every of theme. x_i could be the Initial condition (purple line), boundary condition (red line)

or the diffusion coefficient (blue line). On the x ax there is the input normalized on its mean. The initial condition is 0.2 and z=111 m.

We observe in Fig 3-46 the probability density function of the normalized concentration solution "Z" (mean is zero and variance is one) in z=L/2. We can observe the symmetry of the pdf, result that could be seen also in Fig 2-42 where all the three AMAG indices are zero. Another characteristic that confirm the useful of AMA indices is the good fit of the distribution with the normal distribution that it could be seen also in the AMAK.

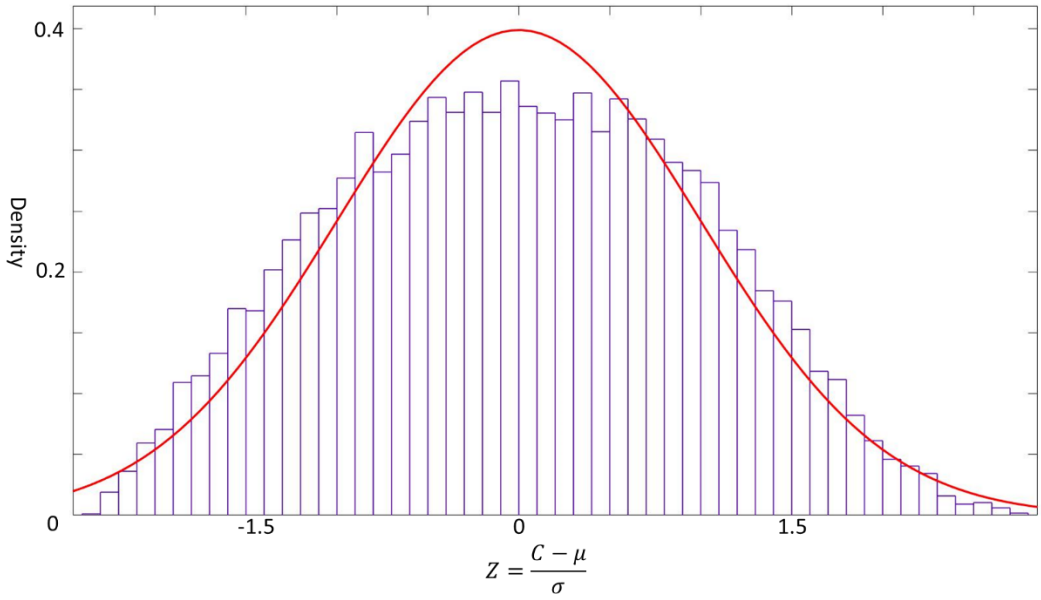


Fig 3-46 Pdf of the concentration solution in z=50 m normalized. The diffusion time is 2.5 million years. Red line is the normal distribution fit.

Fig 3-47 shows the probability density function of the normalized concentration solution "Z" (mean is zero and variance is one) in z=111 m. This is the point where AMAK value relative to the initial condition has a maximum and we can observe how the pdf is similar at a uniform pdf and it confirm that this point maintains the characteristics of the initial concentration distribution. In fact it isn't influenced by the boundary condition and it has zero concentration gradient and so also the diffusion coefficient doesn't influence the density function.

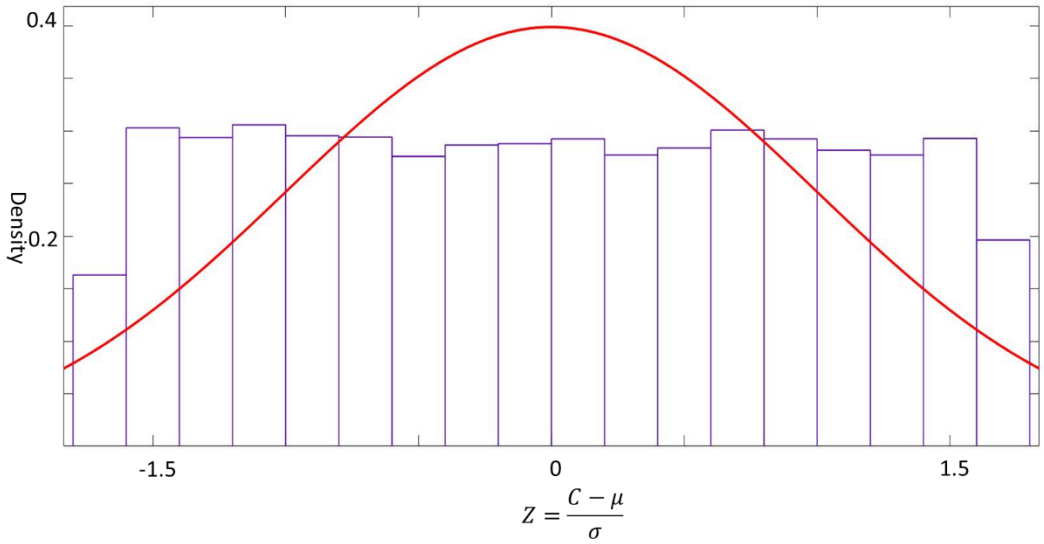


Fig 3-47 Pdf of the concentration solution in z=111 m normalized. The diffusion time is 2.5 million years. Red line is the normal distribution fit.

3.5.2 Instantaneous Flux GSA

In the follow section we show the results obtained with the global sensitivity analysis for the Instantaneous flux solution and the relative comments.

3.5.2.1 Numerical First-order Sobol's Indices

The procedure used to compute the follow graphs is the same used above for the concentration data.

In Figure 3-48 we observe the Sobol's indices along the space. We observe how the diffusion coefficient paly the principal role in uncertainty analysis. This due to the further diffusion coefficient that multiply the derivative of the concentration solution formula, as we seen above. The point where the boundary condition and initial condition Sobol's indices have the same value (very less) is the same of the maximum diffusion coefficient Sobol's index and the same where the AMAK concentration index has the maximum, that is Z= 111 m. This is more distance respect the same point in the concentration analysis (c.a. 55 m).

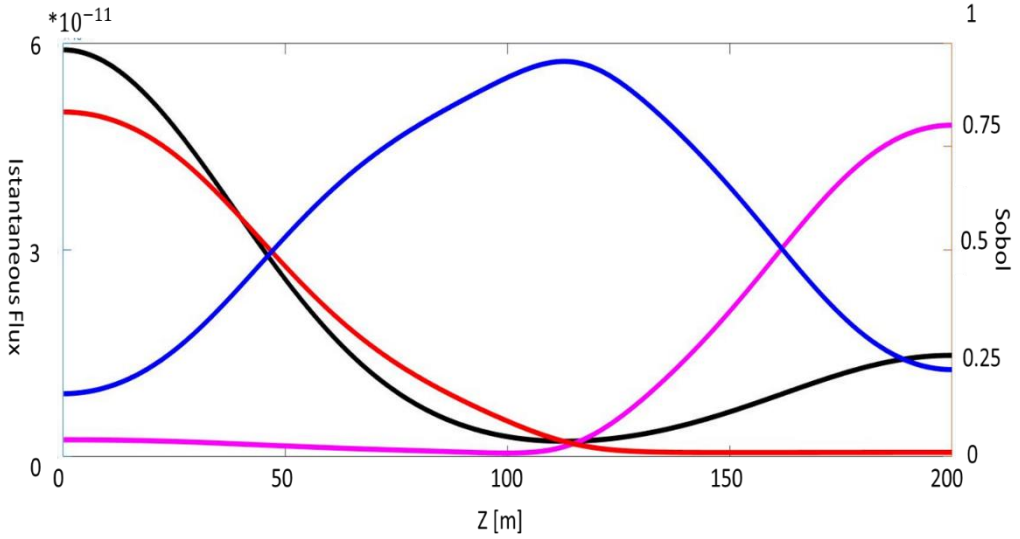


Fig 3-48 Numerical Sobol indices for the instantaneous flux problem with every input random and with coefvar 0.25. The diffusion time is 2.5 million years. The purple line is referred to initial condition, red line to boundary condition and the blue line to diffusion coefficient. Black line is the instantaneous flux solution.

3.5.2.2 Multiple-moment-based metrics for GSA

Fig 3-49 shows the AMAE indices for the instantaneous flux solution with time diffusion of 2,5 million years. We observe that as already seen before in many occasions, the point Z=111 m is that where boundary condition and Initial condition change the relative importance also for the mean solution.

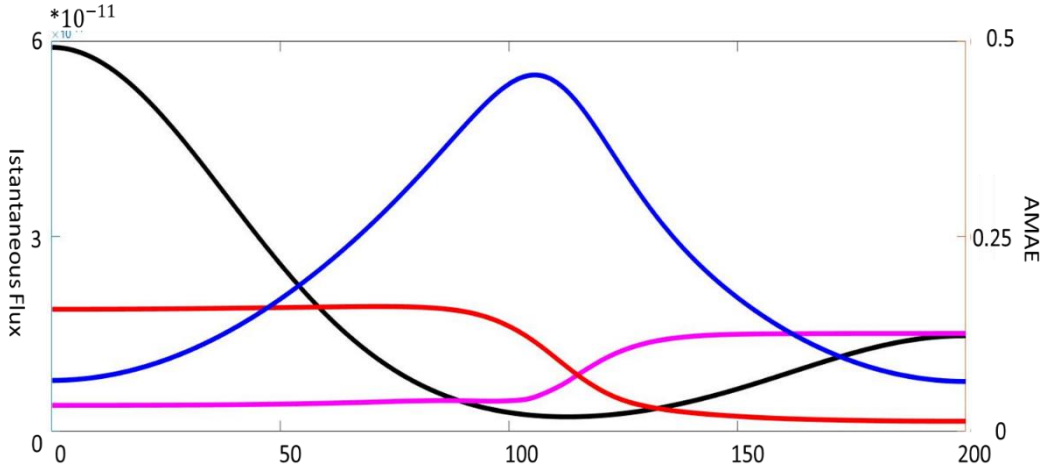


Fig 3-49 AMAE indices Eq. 2-39 for the instantaneous flux problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years. The purple line is referred to initial

condition, red line to boundary condition and the blue line to diffusion coefficient. Black line is the instantaneous flux solution.

We observe and we obtain the same information obtained from Sobol's index by Fig 3-50, that shows the AMAV coefficient. We can make another consideration that, also if the relative importance of the diffusion coefficient is the highest, that is also the point where the total uncertainty is smallest and so diffusion coefficient lost its importance.

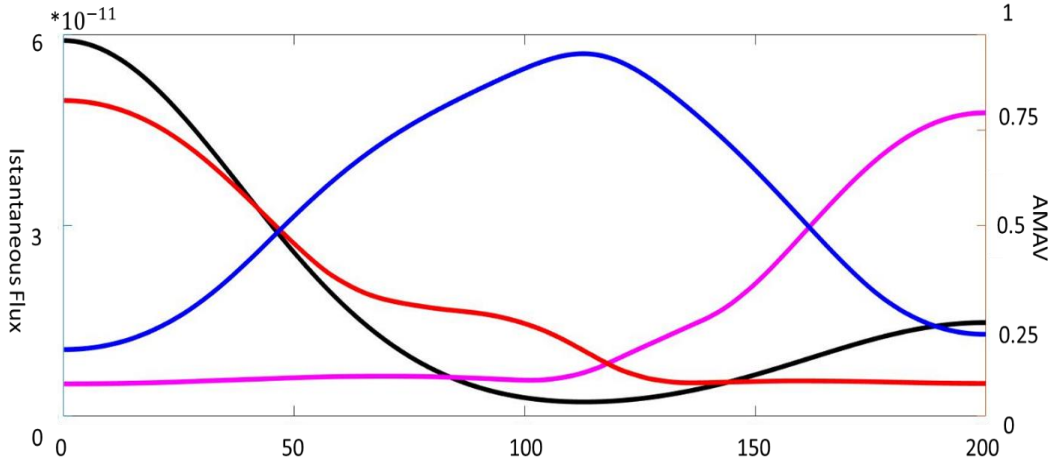


Fig 3-50 AMAV indices Eq.2-40 for the instantaneous flux problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years. The purple line is referred to initial condition, red line to boundary condition and the blue line to diffusion coefficient. Black line is the instantaneous flux solution.

In Fig 3-51 we see the AMAG indices along the space caprock. We see that the AMAG relative to the diffusion coefficient maintains a certain constant value (c.a. 1) while the others two have a minimum c.a. where the solution has its minimum.

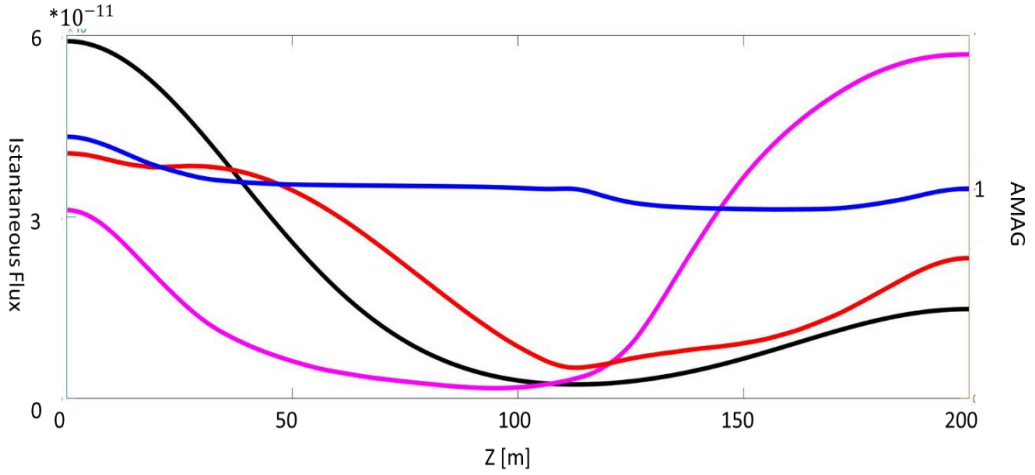


Fig 3-51 AMAG indices Eq.2-41 for the instantaneous flux problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years. The purple line is referred to initial condition, red line to boundary condition and the blue line to diffusion coefficient. Black line is the instantaneous flux solution.

We close this part finally observing in Fig 3-52 the AMAK indices. We can observe how the diffusion coefficient have a strange behavior, having a little relative maximum, in proximity of the minimum of the instantaneous flux.

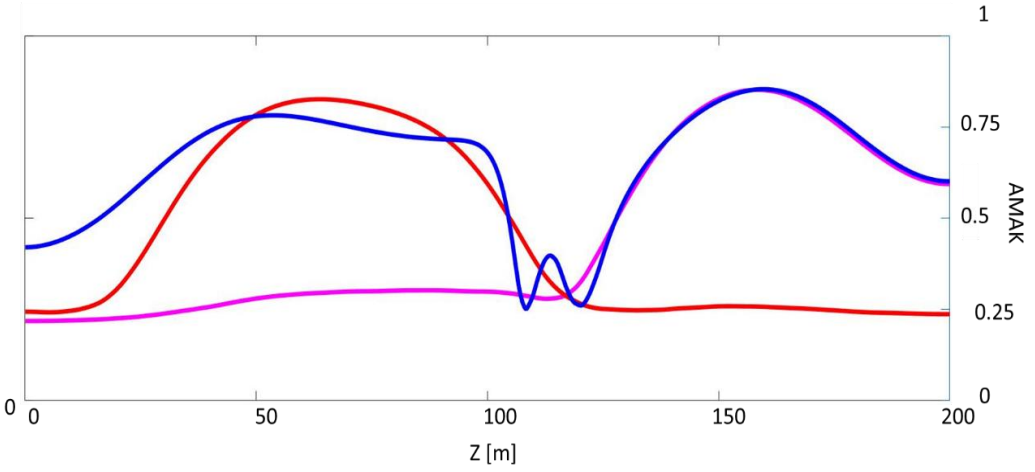


Fig 3-52 AMAK indices Eq.2-42 for the instantaneous flux problem with every input random and with coefvar 0.25 for all of theme. The diffusion time is 2.5 million years. The purple line is referred to initial condition, red line to boundary condition and the blue line to diffusion coefficient. Black line is the instantaneous flux solution.

Fig 3-53 shows the probability density function relative to the normalized flux solution in z=L/4=50 m. We observe how, respect the distribution of the

concentration in the same point seen in Fig 2-46, now the symmetry is not maintained. This could be seen also in Fig 2-51.

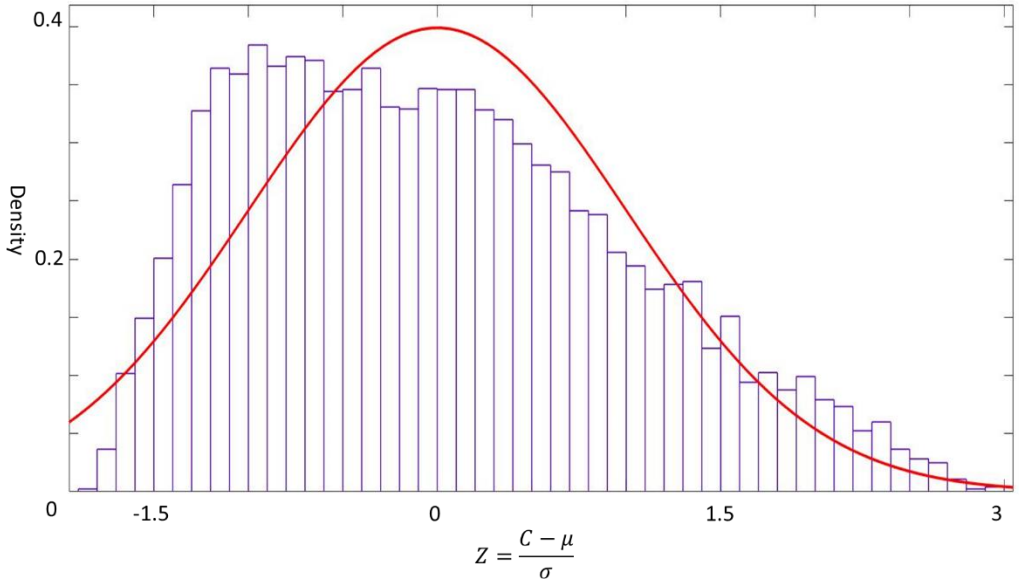


Fig 3-53 Pdf of the flux solution in z=50 m normalized. The diffusion time is 2.5 million years. Red line is the normal distribution fit

Fig 3-54 shows the probability density function relative to the normalized flux solution in z=111 m, the point in proximity of the minimum of the instantaneous flux. We can see the bigger asymmetry of the pdf respect that in z=50 m in Fig 2-52 and also the increase difference between the distribution and the Gaussian.

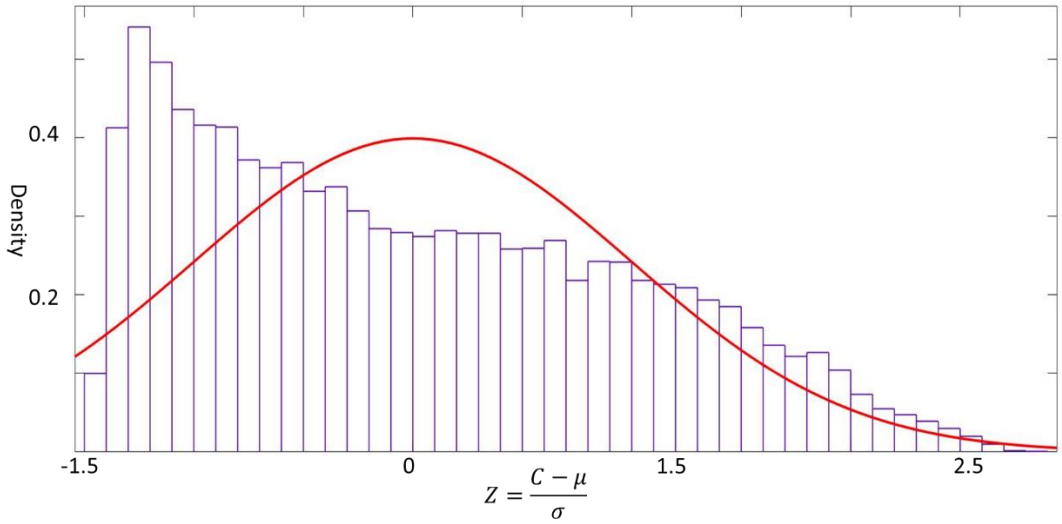


Fig 3-54 Pdf of the flux solution in z=111 m normalized. The diffusion time is 2.5 million years. Red line is the normal distribution fit.

4 CONCLUSION

In this thesis we study the diffusion problem within a stochastic framework. We analyze the homogeneous case, taking into account the concentration and instantaneous flux both analytically and numerically. We considered initial condition, boundary condition and diffusion coefficient as random inputs.

First, we let them vary one by one keeping constant the others two, then their joint effect on the statistical moments of the variable of interest. For every case, analytical expressions of mean and variance of the concentration and instantaneous flux are computed. In order to quantify the relative impact of each random input on the solutions, we also performed a global sensitivity analysis using two different methods: the Sobol's index (Sobol I. , 2001) (Sobol I. , 1993) and multiple-moment-based metrics (Dell'Oca, Riva, & Guadagnini, 2017).

Our analytical solutions of mean and variance of the target variables (i.e., gas concentration and flux) can be obtained with negligible CPU time, compared to the Montecarlo method and to the numerical solution of the PDEs satisfied by the moments of the target variables.

Key results of our study are briefly summarized below.

(1) From the case where the initial condition is considered random, we observe that the concentration profile has a flex point when the mean concentration reaches the value of the initial concentration and at the same point we have the minimum flux. The position of the minimum mean flux, or, equivalently, of the flex point in the concentration profiles, depends on the observation time and the relative values of the initial and boundary concentration. This is also the position of the max concentration variance and of the minimum instantaneous flux variance.

(2) From the case where the diffusion coefficient is considered random, we observe that the uncertainly in the concentration solution, after certain time, vanishes. Variance of the concentration solution has a bell shape while flux solution has a non-zero value in reservoir (z=0), a maximum at

the same z point of the maximum variance of the concentration and as expected 0 variance after the maximum diffusion point.

(3) From the case where the boundary condition is considered random, we note that the variances of the concentration have an exponential behavior with a maximum for z=0, where the stochastic variable is defined.

(4) From the case where all the three variables are considered random, we made also a global sensitivity analysis. By observing the spatial profiles of the Sobol indices we identified some critical points. They are the points where the behaviors of the moments have a minimum or a maximum. They are: the point that has the same concentration value of the initial condition (when this last isn't null) when one boundary condition is bigger than initial and the other is smaller and the point of maximum expansion of the diffusion from the boundary condition (when the initial condition is null). The first coincide also with the null flux point, where the concentration gradient is zero and so also the flux uncertainty is zero. Uncertainty due to the initial condition influences the solution only at the beginning. In time, the boundary condition uncertainty and diffusion coefficient uncertainty take always more influence.

The stochastic approach studied for the diffusion problem could be useful to understand the behavior of the gas transport through law sedimentary rock. The GSA method permits to understand the roles and the iterations of multiple sources of uncertainty and so it permits to predict the bigger output uncertainties about gas migration through a caprock.

Two main possible future application fields where the model can be useful are: the characterization of hydrocarbon origin by the isotope ratio and the shale gas field.

(1) Compositional data and stable isotope ratios are critical datasets for the study of hydrocarbon generation, expulsion, and migration. These geochemical parameters are also important for studies of leakage detection for geological carbon sequestration. Diffusion, gas–liquid partitioning and

adsorption, the three most crucial processes to affect gas transport through shale and other low-permeability formations, are known to cause stable isotope fractionation. A considerable number of studies have been conducted on stable isotope fractionation associated with diffusion and gas–liquid partitioning. However, significantly fewer data are reported in the literature that specifically address isotope fractionation during gas transport through a shale caprock. The main study about this process has done to characterize the generation of the hydrocarbon: thermogenesis or biogenesis. Many authors give different models to understand from the isotopic ratio and other ratios, the origin of the gas and if there was mixing as in (Alain Prinzhofer, 1997). It suggests testing a mixing hypothesis with bacterial methane to use a diagram displaying ethane/methane ratios versus $\delta^{13}C$. Leakage from reservoirs and source rocks must be considered as an important possible vector for the degassing of carbon compounds through the atmosphere. Considering that the isotopes help to trace such flux, this has consequences both for gas exploration and for the understanding of the carbon cycle in the atmosphere.

(2) Another useful idea to using the model proposed is in the shale gas field. Shale gas is a natural gas (predominantly methane) found in shale rock. Natural gas produced from shale is often referred to as ‘unconventional’ and this refers to the type of rock type in which it is found. ‘Conventional’ oil and gas refers to hydrocarbons which have previously sought in sandstone or limestone, instead of shale or coal which are now the focus of unconventional exploration. However, the techniques used to extract hydrocarbons are essentially the same. What has changed are advancements in technology over the last decade which have made shale gas development economically viable. The fact that the processes of movement of gases in poorly permeable media is not yet clear. The different roles that Fick’s diffusion, Knudsen diffusion and surface diffusion play in passing the gas from the matrix to the fractures make it very difficult to obtain an effective diffusion coefficient to be used in the models.

APPENDIX A – DIFFUSION SOLUTION PROOF

Using the separation of variables technique, the generic solution for the problem Eq.2-16 can be written as:

$$C(z,t) = u(z)G(t) \qquad \textit{Eq. A 1}$$

Thus, equation Eq.2-16:

$$\frac{\partial u(z)G(t)}{\partial t} = D\,\frac{\partial^2 u(z)G(t)}{\partial z^2}, \quad 0 < z < L, t > 0 \qquad \textit{Eq. A 2}$$

And so

$$u(z)\frac{\partial G(t)}{\partial t} = DG(t)\frac{\partial^2 u(z)}{\partial z^2}, \quad 0 < z < L, t > 0 \qquad \textit{Eq. A 3}$$

The two terms must be equal to the same constant which is conveniently takes as - λ^2 with $\lambda = \frac{n\pi}{L}$.

$$\frac{1}{DG(t)}\frac{\partial G(t)}{\partial t} = \frac{1}{u(z)}\frac{\partial^2 u(z)}{\partial z^2} = -\lambda^2, \quad 0 < z < L, t > 0 \qquad \textit{Eq. A 4}$$

Therefore, we obtain for $\lambda \neq 0$:

$$G(t) = e^{-\lambda^2 Dt} \qquad u(z) = A_1\, sin(z\lambda) + B_1\, cos(z\lambda) \qquad \textit{Eq. A 5}$$

Leading to a solution of the form:

$$C(z,t) = u(z)G(t) = (A_1\, sin(z\lambda) + B_1\, cos(z\lambda))e^{-\lambda^2 Dt} \qquad \textit{Eq. A 6}$$

where A_1 and B_1 are constants of integration. Since the diffusion partial differential equation is a linear equation, the most general solution is obtained by summing solution of this type, so that we have:

$$C(z,t) = \sum_{n=1}^{\infty}(A_1\, sin(z\lambda) + B_1\, cos(z\lambda))e^{-\lambda^2 Dt} \qquad \textit{Eq. A 7}$$

Firstly, we consider the homogeneous boundary conditions and we obtain that B_1 must be 0 and using the initial condition:

$$c_0(z) = \sum_{n=1}^{\infty}(A_1\, sin(z\lambda)) \qquad \textit{Eq. A 8}$$

But any function, defined on the interval 0<x<l has a unique representation of the form above and the coefficients in this are given by:

$$A_1 = \frac{2}{L}\int_0^L c_0(z')\, sin(z'\lambda)dz' \qquad \textit{Eq. A 9}$$

We now consider our non-homogeneous problem and we seek a transformation of the form

$$C(z,t) = F(z,t) + A(t) * z + B(t) \qquad \textit{Eq. A 10}$$

where $F(z,t)$ is the homogeneous problem.
By taking into account the boundary conditions of *C(z,t)* and the respective homogeneous problem *F(z,t)*, we obtain:

$$A(t) = \frac{c_b(t) - c_t(t)}{L}\ , \quad B(t) = \ c_t(t) \qquad \textit{Eq. A 11}$$

And consequently

$$C(z,t) = F(z,t) + \frac{c_b(t)-c_t(t)}{L} * z + c_t(t) \qquad \textit{Eq. A 12}$$

However, in doing so, we change not only the original equation but also the initial condition. They become, respectively:

$$\frac{\partial F(z,t)}{\partial t} = \frac{\partial D(z,t)\partial F(z,t)}{\partial z^2} - \frac{\partial c_b(t)}{\partial t}\frac{z}{L} + \frac{\partial c_t(t)}{\partial t}\frac{(L-z)}{L} \qquad \textit{Eq. A 13}$$

$$F_o(z,0) = c_o(z) - c_b(t)\frac{z}{L} - c_t(t)\frac{(L-z)}{L} \qquad \textit{Eq. A 14}$$

In addition, if $c_b(t)$ and $c_t(t)$ are constant in time we obtain:

$$F(z,t) = \sum_{n=1}^{\infty}\frac{2}{L}\int_0^L F_o(z',0)\,sin(z'\lambda)dz'\,sin(z\lambda)\,e^{-\lambda^2 Dt} =$$
$$\sum_{n=1}^{\infty}\frac{2}{L}\int_0^L \left(c_o(z) - c_b\frac{z'}{L} -$$
$$c_t\frac{(L-z')}{L}\right) sin(z'\lambda)dz'\,sin(z\lambda)\,e^{-\lambda^2 Dt} \qquad \textit{Eq. A 15}$$

And using the expression above:

$$C(z,t) = F(z,t) + \frac{c_b-c_t}{L} * z + c_t = \sum_{n=1}^{\infty}\frac{2}{L}\int_0^L \left(c_o(z) - c_b\frac{z'}{L} -$$
$$c_t\frac{(L-z')}{L}\right) sin(z'\lambda)dz'\,sin(z\lambda)\,e^{-\lambda^2 Dt} + \frac{c_b-c_t}{L} * z + c_t \qquad \textit{Eq. A 16}$$

APPENDIX B – MEAN SQUARE CALCULUS

General treatment of the mean square solutions of random differential equations requires some basic operational tools such as, the algebra of the mean square limits, the derivative of the product of two stochastic processes or the chain rule for the composition of stochastic process. Apart from special cases where the independence of the stochastic processes factors is assumed, to our knowledge, general results for the derivative of a product or the chain rule are not available. Difficulties with the mean square operational calculus of the product result from the fact that the mean square norm is not sub multiplicative.

We make here a review of some important results in the mean square calculus. Let (Ω, F, P) be a probability space. A random variable Y:Ω → R is said to be continuous if its distribution function F_Y is continuous and differentiable almost everywhere. In this case, its density function is defined by

$$g_Y(y) = \frac{dF_Y(y)}{dy} \qquad \textit{Eq. B 1}$$

Moreover, if Y satisfies the additional property

$$\int_R y^2 g_Y(y) dy < \infty \qquad \textit{Eq. B 2}$$

Then Y is said to be a second-order random variable and the integral is the expectation $E[Y^2]$ of Y^2 . (Soong, 1973)

Consider a stochastic process X(t), t ϵ T. X(t) is called a second order stochastic process if for every set t_1, t_2… the random variables $X(t_1)$, $X(t_2)$, $X(t_3)$…are elements of L_2-space.

L_2-space is a linear vector space where the inner product, the norm and the distance are defined as:

- $E\{X_1X_2\} = \langle X_1, X_2 \rangle$
- $\|X(t)\| = \langle X, X \rangle^{1/2}$
- $d(X_1, X_2) = \|X_1 - X_2\|$

A second order stochastic process is characterized by

$$\|X(t)\|^2 = E[X^2(t)] < \infty, \quad t \in T \qquad \textit{Eq. B 3}$$

and its covariance function is the deterministic function $\Gamma_{XX}(t, s)$defined by:

$$\Gamma_{XX}(t, s) = E[X(t)X(s)] - E[X(t)]E[X(s)] \qquad \textit{Eq. B 4}$$

Now the first step in the development of a calculus for stochastic processes is to define the convergence of a sequence of a random variable. There are four modes of convergence: convergence in distribution, convergence in probability, convergence in mean square and almost sure convergence. The development of the mean square calculus is based on the concept of convergence in mean square.

A sequence of random variables $\{X_n\}$ converges in mean square to a random variable X as n goes to → infinity if

$$\lim_{n\to\infty} \|X_n - X\| = 0 \qquad \textit{Eq. B 5}$$

And so we have also (proof in[(Soong, 1973),p.76]):

$$\lim_{n\to\infty} E\{X_n\} = E\{X\} \qquad \textit{Eq. B 6}$$

The concept of mean square convergence can be extended from a random sequence to a second-order stochastic process X(t) where t is continuous over a finite interval. This extension leads to the notion of continuity and, in this case, continuity in mean square.

By [(Soong, 1973),p.90] a second order stochastic process X(t), t ϵ T, is mean square continuous at t if, and only if, $\Gamma(t,s)$ is continuous.

By the concept of mean square continuity follow the concept of mean square differentiation; a second order stochastic process X(t) has a mean square derivative at t if

$$\lim_{\tau\to 0}[X(t+\tau)-X(t)]/\tau = \dot{X(t)} \qquad \textit{Eq. B 7}$$

Some properties are associated with the mean square derivative of a second order stochastic process (Soong, 1973):

- Mean square differentiability of X(t) at tϵT implies mean square continuity of X(t) at t, since t+ τ ϵ T.
- The mean square derivative $\dot{X(t)}$ of X(t) at t ϵ T, if it exists, is unique.
- If X(t) and Y(t) are mean square differentiable at t ϵ T then the mean square derivative of aX(t) + bY(t) exists at t and

$$\frac{d[aX(t)+bY(t)]}{dt} = \dot{aX(t)} + \dot{bY(t)} \qquad \textit{Eq. B 8}$$

- If an ordinary function $f(t)$ is differentiable at t ϵ T and X(t) is mean square differentiable at t ϵ T then $f(t)X(t)$ is mean square differentiable at t and
-

$$\frac{d[f(t)X(t)]}{dt} = \dot{\frac{d[f(t)]}{dt}}X(t) + \frac{dX(t)]}{dt}f(t) \qquad \textit{Eq. B 9}$$

$$E\{\dot{X(t)}\} = \frac{dE\{X(t)\}}{dt} \qquad \textit{Eq. B 10}$$

We define now the mean square Riemann Integral; let X(t) be a second order stochastic process defined on [a,b]. Let *f(t,u)* be an ordinary function defined on the same interval for t and Riemann integrable for every u ϵ U. We form the random variable:

$$Y_n(u) = \sum_{k=1}^{n} f(t'_k, u)X(t'_k)(t_k - t_{k-1}) \qquad \textit{Eq. B 11}$$

If, for u ϵ U,

$$\lim_{n\to\infty} Y_n(u) = Y(u) \qquad \textit{Eq. B 12}$$

exists for some sequence of a subdivisions p_n (collection of all finite partitions of interval [a,b]) , the stochastic process Y(u) is called the mean square Riemann integral of $f(t,u)X(t)$ over [a,b] and is denoted by:

$$Y(u) = \int_a^b f(t,u)X(t)dt \qquad \textit{Eq. B 13}$$

end it exists if, and only if, the ordinary double Riemann Integral

$$\int_a^b \int_a^b f(t,u)f(s,u)\Gamma_{XX}(t,s)dt\,ds \qquad \textit{Eq. B 14}$$

exists and is finite.
If Y(u) exists, it easily shown [(Soong, 1973),p.104] that:

$$E\{Y(u)\} = \int_a^b f(t,u)E\{X(t)\}dt \qquad \textit{Eq. B 15}$$

And the correlation function of Y(u) in terms of that of X(t) is so:

$$\Gamma_{YY}(u,v) = \int_a^b \int_a^b f(t,u) f(s,u) \Gamma_{XX}(t,s) dt\, ds \qquad \textit{Eq. B 16}$$

Some properties are associated with the mean square Riemann Integrals of a second order stochastic process [(Soong, 1973),pag.101]:

- Mean square continuity of X(t) on [a, b] implies mean square Riemann Integrability of X(t) on [a, b].
- The mean square integral of X(t) on an interval [a, b], if it exists, is unique. This property follows immediately from the mean square convergence uniqueness property.
- If X(t) is mean square continuous on [a, b] then

$$\left\| \int_a^b X(t) dt \right\| \leq \int_a^b \|X(t)\| dt \leq M(b-a) \qquad \textit{Eq. B 17}$$

where

$$M = max_{t\epsilon[a,b]} \|X(t)\| \qquad \textit{Eq. B 18}$$

- If X(t) is mean square continuous on [a, t] then
-

$$Y(u) = \int_a^t X(s) ds \qquad \textit{Eq. B 19}$$

Is mean square continuous on T; it is also mean square differentiable on T with

$$\dot{Y(t)} = X(t) \qquad \textit{Eq. B 20}$$

We finally say that if $X_1, X_2 \dots X_m$ are square convergence random variable then $Cov[X_i, X_j] = E[X_i X_j] - E[X_i]E[X_j]$ and if a_i is a real number for each i. then:

$$VAR[\sum_{i=1}^{m} a_i X_i] = \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j COV[X_i, X_j] \qquad \textit{Eq. B 21}$$

Remembering some follow important properties of the mean and of the variance for two stochastic random variables X,Y:

$$E[X + Y] = E[X] + E[Y] \qquad \textit{Eq. B 22}$$

$$E[X \cdot Y] = E[X] \cdot E[Y] \qquad \textit{Eq. B 23}$$

$$VAR[X + Y] = VAR[X] + VAR[Y] \qquad \textit{Eq. B 24}$$

$$VAR[X \cdot Y] = VAR[X] \cdot VAR[Y] + VAR[Y] \cdot E[X]^2 + VAR[X] \cdot E[Y]^2 \qquad \textit{Eq. B 25}$$

APPENDIX C – EULER FINITE DIFFERENCE METHOD

We solve the 1-d diffusion equation for points on a grid using the finite difference method where we discretize in z and t for $0 \leq z \leq L$ and $0 \leq t \leq T\_max$.

We discretize in time with time step: $\Delta t = T\_max/m$ and in space with grid spacing: $\Delta z = L/(n+1)$ and let $t_j = j\Delta t$ and $z_i = i\Delta$

$$D\frac{C_{i+1}^{j} - 2C_i^j + C_{i-1}^j}{\Delta z^2} \qquad \textit{Eq. C 1}$$

And explicit:

$$C_i^{j+1} = C_i^j + \frac{\Delta t \cdot D}{\Delta z^2} z \qquad \textit{Eq. C 2}$$

We can explicit our equation so as:

$$\frac{C_i^{j+1} - C_i^j}{\Delta t} = \left(C_{i+1}^j - 2C_i^j + C_{i-1}^j\right) \qquad \textit{Eq. C 3}$$

Simple forward Euler method for solving the 1-d diffusion equation is subject to a numerical instability under certain conditions. (Fitzpatrick, 2006). Consider the time evolution of a single Fourier mode of wave-number k:

$$C(z,t) = C(t) * \widehat{e^{\imath kz}} \qquad \textit{Eq. C 4}$$

Substitution of the above expression into our finite difference scheme yields:

$$\widehat{C_\iota^{J+1}e^{\imath kz}} = \widehat{C_\iota^J e^{\imath kz}}\left[1 + r * (e^{-ik\delta z} - 2 + e^{ik\delta z})\right] \qquad \textit{Eq. C 5}$$

Or

$$\widehat{C_\iota^{J+1}} = AA * \widehat{C_\iota^J} \qquad \textit{Eq. C 6}$$

Where

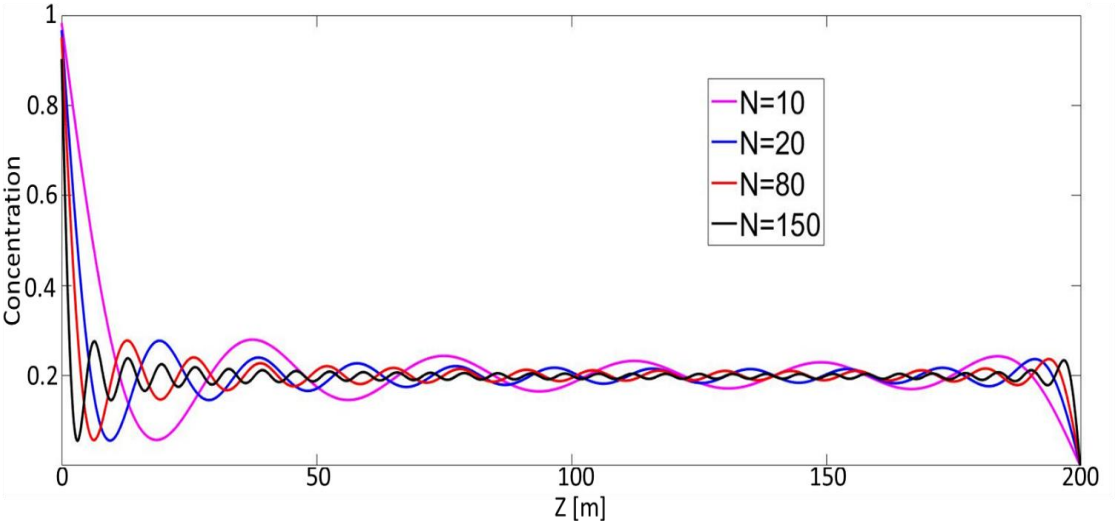
$$AA = 1 - 2r * (1 - cos\, k\delta z) = 1 - 4r * sin^2\left(\frac{k\delta z}{2}\right) \qquad \textit{Eq. C 7}$$

Thus, the amplitude of the Fourier mode is amplified by a factor AA at each time-step. In order for the differencing scheme to be stable, the modulus of this amplification factor must be less than unity for all possible values of "k". Now, the largest possible value of $\sin^2(\frac{k\delta z}{2})$ is unity; hence, the wave-length corresponding to this value is that of the most unstable Fourier mode. In fact, the most unstable mode possesses a wave-length which is half the grid-spacing: $\frac{\delta z}{2}$. It follows that from the equation of AA, our stability condition is:

$$r = \frac{\Delta t \cdot D}{\Delta z^2} < 0.5 \qquad \textit{Eq. C 8}$$

APPENDIX D – SOLUTION FOR SHORT DIFFUSION TIME

For the deterministic solution, we see that to obtain a good result a t=0 we have to increase the truncation of the series to N over 80; correspondingly the computed time increase from 0.30 minute to 5 minutes for 1000 Montecarlo simulations.

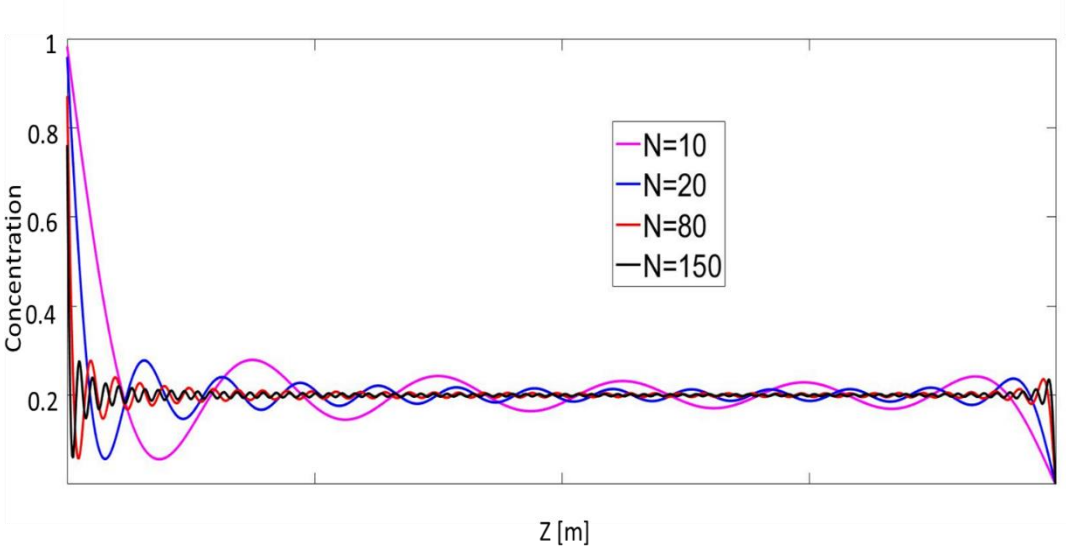


Comparison between the concentration deterministic solution profile at t=10000 years for different numbers of N in the solution series.

While for the analytical solution, the computed time is not limiting and can be seen for t=0:

$$E[C_N(z,t)] = c_t + (-c_t)\frac{z}{L} + \sum_{n=1}^{N} sin\frac{n\pi z}{L}\left[\frac{2}{L}\left(\frac{b+a}{2}\right)\left(\frac{L^2}{n\pi}(1 - (-1)^n)\right) - \frac{2}{n\pi}[c_t + (-1)^n(c_t)]\right] \qquad \textit{Eq. D 1}$$

If we truncated the solution for different N we obtain the follow results display the temporal evolution of the concentration computed with diverse values N of series truncation:



Comparison between the concentration analytical solution profile at t=10000 years for different truncated numbers of N in the solution series.

The considerations above permit to say that the initial uncertainly conditions influences the uncertainly result less and less in time and that to obtain a good result for short time we need more computational power. However after some time (c.a. 10000 years) the three solutions coincide and then we can use a lower N.

REFERENCES

Alain Prinzhofer, E. P. (1997). Isotopically light methane in natural gas: bacterial imprint or diffusive fractionation? *Chemical Geology 142*, 193-200.

Al-Bazali an J.Zhang, M. C. (9-12/10/2005). *Measurement of the Sealing Capacity of Shale Caprocks.* Dallas : SPE.

Barker, C. (1990). The role of source rock studies in petroleum exploration, in K.S. Johnson and B.J. Cardott. *Symposium:Oklahoma Geological Survey Circular 93* (pp. 3-20). Source Rocks in the Southern Midcontinent.

Barker, C., & Horsfield, B. (1982). Mechanical Versus Thermal Cause of Abnormally High Pore Pressures in Shales: DISCUSSION. *AAPG Bulletin*, 99-100.

Cameron, R., & Martin, W. (1947). The ortoghonal development of nonlinear functionals in series of Fourier-Hermite functionals. *Ann Math* , 385-392.

Crank, J. (1975). *The Mathematics of Diffusion.* Oxford: Clarendon Press.

Dell'Oca, A., Riva, M., & Guadagnini, A. (2017). Moment-Based metrics for global sensitivity analysis of hydrologiacal Systems. *Hydrology and Earth System Sciences, 21*, 1-16.

Durand , B. (1980). Sedimentary organic matter and kerogen: definition and quantitative importance of kerogen,. *ed., Kerogen: Techniq*, 13-14.

Encyclopædia Britannica, i. (2012, 03 19). *Encyclopædia Britannica*. Retrieved from Encyclopædia Britannica: https://www.britannica.com/science/gas-reservoir

Fitzpatrick, R. (2006). *http://farside.ph.utexas.edu.* Retrieved from http://farside.ph.utexas.edu/teaching/329/lectures/node79.html

Fuel Chemistry Division Public Education , & Outreach Committee. (n.d.). Retrieved from Fuel Chemistry Division: https://www.ems.psu.edu/~pisupati/ACSOutreach/Natural_Gas.html

Geology.com. (n.d.). *Geology.com_Geoscience News and Information*. Retrieved from geology.com: http://geology.com/rocks/shale.shtml

Ghanem, R., & Spanos, P. (1991). *Stochastic finite elements - spectral approach.* Berlin: Springer.

Hantschel, T., & Kauerauf, A. (2009). *Fundamentals of Basin and Petroleum Systems Modeling.* Springer.

Hirschfelder, Bird, R., & Curtiss, C. (1954). *Molecular Theory of Gases and Liquids.* New York: Wiley.

Howard , M., & Karlin, S. (1998). *An Introduction to Stochastic Modeling.* San Diego (USA): Accademic Press.

J. Hyne, N. (2001). Nontechnical Guide to Petroleum Geology, Exploration, Drilling, and Production. PennWell Books.

J.C.Cortes, P. S. (2005). Analytic-Numerical Approximating processes of Diffusion Equation with Data Uncertainty. *Computer and MAthematics with applications,49*, 1255-1266.

Karniadakis, G., & Xiu, D. (2002). Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics,187*, 137-167.

Keliu Wua, ⇑. Z. (2016). A model for multiple transport mechanisms through nanopores of shale gas reservoirs with real gas effect–adsorption-mechanic coupling. *International Journal of Heat and Mass Transfer 93*, 408-426.

L. Buryakovsky, N. E. (2005). *Geology and Geochemistry of Oil and Gas.* Elsevier Science.

Le Maitre, O., & Knio, O. (2010). *Spectral methods for uncertainty quantification. In: Scientific Computation.* New York: Springer.

Leonid F. Khilyuk, G. V. (2000). Mechanisms of Gas Migration . In G. V. Leonid F. Khilyuk, *Gas Migragtion* (pp. 248-263). Houston, Texas: Gulf Publishing company.

Maxwell, J. (1890). *The Scientific Papers of James Clerk Maxwell.* Cambridge, UK.: University Press,.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., . . . Tarantola, S. (2008). *Global Sensitivity Analysis* . Chichester,West Sussex PO19 8SQ, England: John Wiley & Sons Ltd,.

Sobol, I. (1993). Sensitivity Estimates for Nonlinear Mathematical Models. *MMCE,Vol1,NO4*.

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

Song, J., & Zhang, D. (2013). Comprehensive review of caprock — sealing mechanisms for geologic carbon sequestration. *Environ. Sci Technol. 47*, 9-22.

Soong, T. (1973). *Random Differential Equations in Science and Engineering.* New York: Academic Press.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansion. *Reliability Engineering and System Safety 93*, 964-979.

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

Sundquist, E., Burruss, R., Faulkner, S., & and all. . (n.d.). Carbon Sequestration to Mitigate Climate Change. *USGS* .

Wang, J., & Peng, Y. (2014). Numerical modeling for the combined effects of two-phase flow,deformation, gas diffusion and CO2 sorption on caprock sealing efficiency. *Journal of Geochemical Exploration 144*, 154-167.

Wang, J., Liu, J., & Kabir, A. (2013). Combined effects of directional compaction, non darcy flow and anistropic swelling on coal seam gas extraction. *Int. J. Coal Geol. 109-110*, 1-14.

Welty, J., Wicks, C., Rorrer, G., & Wilson, R. (2007). *Fundamentals of Momentum, Heat and Mass Transfer*. Wiley: VCH.

Wiener , N. (1938). The Homogeneous Chaos. *AM J MAth 60*, 897-936.

Yanfeng He, J. C. (2017). Research on shale gas transportation and apparent permeability in nanopores . *Journal of Natural Gas Science and Engineering (38)*, 450-457.

Zhenhao Duan *, S. M. (2006). A thermodynamic model for calculating methane solubility, density and gas phase composition of methane-bearing aqueous fluids from 273 to 523 K and from 1 to 2000 bar. *Geochimica et Cosmochimica Acta 70*, 3369-3386.